

Ấn bản Thư viện AI tiếng Việt

Trí tuệ nhân tạo dịch ngôn ngữ của loài vật

Câu chuyện về AI học cách
lắng nghe tiếng cá voi, cá
heo, chim và ong

Luật sư Kim Kyung-jin

Trí tuệ nhân tạo dịch ngôn ngữ của loài vật

Câu chuyện về AI học cách lắng nghe tiếng cá voi, cá heo, chim và ong

Luật sư Kim Kyung-jin

Câu chuyện bắt đầu từ những nhà nghiên cứu đã sống cùng cá heo ở Bahamas trong 40 năm, một nhà ngôn ngữ học tìm thấy những mẫu giống nguyên âm trong tiếng lách cách của cá nhà táng, và các nhà khoa học đã trao đổi âm thanh với cá voi lưng gù Twain trong 20 phút. Cuốn sách gồm mười hai chương. Sách giải thích cách trí tuệ nhân tạo vẽ âm thanh động vật như bản đồ để tìm quy luật. Sách giới thiệu công cụ mới đọc tín hiệu của chim, ong mật và khỉ. Sách nói về thái độ tôn trọng thế giới cảm giác của động vật. Sách nêu nguy cơ dùng âm thanh giả để lừa động vật và các quy tắc để ngăn việc đó. Sách cũng trình bày cuộc thảo luận về việc trao tư cách pháp lý cho cá voi. Tôi viết bằng câu ngắn và dễ hiểu để học sinh cuối cấp tiểu học cũng có thể tự đọc. Mỗi tiểu chương đều có tài liệu đã được kiểm chứng. Đây là bản đã được sửa sau khi hai người kiểm tra bằng trí tuệ nhân tạo rà soát.

Kho lưu trữ tài liệu nghiên cứu công khai

Cuốn sách này được sắp xếp và biên soạn bằng trí tuệ nhân tạo. Nó tổng hợp tài liệu bằng các ngôn ngữ chính trên thế giới. NotebookLM không giới hạn ngôn ngữ câu hỏi, dù tài liệu được viết bằng ngôn ngữ nào. Người đọc có thể hỏi kho lưu trữ tài liệu nghiên cứu đã dùng để chuẩn bị cuốn sách này bằng chính ngôn ngữ của mình. Điều này vẫn đúng dù một số tài liệu bằng tiếng Hàn, tiếng Anh hoặc ngôn ngữ khác. Nếu mở phần cài đặt trong NotebookLM và chọn ngôn ngữ đầu ra mong muốn, câu trả lời và kết quả được tạo sẽ hiện bằng ngôn ngữ đó.

Kho NotebookLM công khai đã dùng khi chuẩn bị cuốn sách này (Trí tuệ nhân tạo dịch ngôn ngữ của loài vật):

<https://notebook.google.com/notebook/bb152971-e02c-4008-b56e-4076e1180e53>

Bản gốc tiếng Hàn của cuốn sách này có trong AI Library của kimkj.com:

<https://kimkj.com/ai-library/?mod=document&uid=12655>

Mục lục

Nhấn vào tên từng chương để chuyển đến nội dung tương ứng. Nếu muốn đọc từ đầu, hãy chọn phần mở đầu.

Mở đầu: Mở ra tiếng nói của cộng đồng Trái Đất · 4

Chương 1. Wild Dolphin Project và thế giới của cá heo đốm · 10

Chương 2. Bảng chữ cái âm vị của cá nhà táng và Project CETI · 17

Chương 3. Gặp gỡ cá voi lưng gù Twain · 23

Chương 4. Dàn nhạc của biển sâu · 29

Chương 5. Ngôn ngữ của mặt đất · 35

Chương 6. Bản giao hưởng trên bầu trời và điệu múa vô hình · 41

Chương 7. Từ não động vật đến tấm vải deep learning · 47

Chương 8. Thế giới môi trường của động vật và ranh giới ký hiệu học · 53

Chương 9. Ngã rẽ giữa lừa dối và bảo tồn · 59

Chương 10. Nhân vị phi con người và quyền được máy móc hỗ trợ · 65

Lời kết: Internet giữa các loài và Trái Đất cùng chung sống · 71

Mở đầu: Mở ra tiếng nói của cộng đồng Trái Đất

0.1 Đôi mắt mới mà kính thiên văn Hubble đã mở ra

Tháng 12 năm 1995, các nhà khoa học quay kính thiên văn không gian lên bầu trời. Tên của kính thiên văn đó là Hubble Space Telescope. Nơi kính nhắm tới là điểm tối nhất và trống rỗng nhất trên bầu trời. Nhiều học giả nghĩ ở đó không có gì. Kính thiên văn gom ánh sáng trong bóng tối ấy suốt nhiều ngày. Rồi một kết quả bất ngờ xuất hiện. Trong bóng tối có hàng nghìn thiên hà đang ẩn mình. Thiên hà là một nhóm khổng lồ do nhiều ngôi sao hợp lại. Nơi đó không trống rỗng. Nó đầy ắp. Việc này để lại một bài học lớn cho khoa học. Khi dùng công cụ mới để soi chiếu, ngay cả nơi tưởng như trống rỗng cũng có rất nhiều thứ [1].

Ngày nay, trí tuệ nhân tạo đang làm một việc giống như vậy. Lần này, nơi nó nhắm tới không phải bầu trời. Đó là thế giới âm thanh mà các sinh vật trao đổi với nhau. Công cụ mới này được gọi là âm sinh học số (Digital Bioacoustics). Đây là ngành học ghi âm tiếng của sinh vật rồi nghiên cứu bằng máy tính. Từ lâu, khoa học phương Tây từng xem nhẹ tiếng của động vật. Họ coi đó chỉ là tiếng ồn do bản năng tạo ra. Nhưng khi trí tuệ nhân tạo xem xét các âm thanh ấy, cách nghĩ đã thay đổi. Rừng và biển là sân khấu đối thoại đầy thông tin [2].

Mắt và tai của con người chỉ cảm nhận một phần rất nhỏ của thế giới. Ta dễ lầm tưởng rằng điều ta không cảm nhận được thì không tồn tại. Có một loài chim nhỏ tên là dark-eyed junco. Tiếng của loài chim này nghe bằng tai người thì khá đơn điệu. Nhưng cảm giác thời gian của tai chim nhanh hơn con người. Với chim, âm thanh đó là một giai điệu tinh vi được dệt thành hai lớp. Voi trò chuyện đi rất xa bằng âm thanh rất thấp. Âm thanh này nằm trong khoảng 14 đến 35 hertz (Hz). Phần thấp trong đó là âm thanh con người không nghe được. Âm thanh này lan xa đến vài kilômét. Nếu điều kiện tốt, nó còn đi xa hơn. Dơi cãi nhau bằng âm thanh cao mà con người không nghe được. Ngay cả ấu trùng san hô không có tai cũng tìm nơi để đi bằng âm thanh [3].

Âm thanh con người nghe được có bức tường ở phía trên và phía dưới. Âm quá cao nên không nghe được gọi là siêu âm (ultrasonic). “Siêu” có nghĩa là vượt qua. Đó là âm thanh vượt qua bức tường phía trên của tai người. Âm quá thấp nên không nghe được gọi là hạ âm (infrasonic). Đó là âm thanh vượt qua bức tường phía dưới của tai người. Tiếng dơi thuộc loại siêu âm. Phần thấp trong tiếng voi thuộc loại hạ âm. Con người đã bỏ lỡ hai loại âm thanh này trong thời gian dài. Khi thiết bị ghi âm và kỹ thuật phân tích tốt hơn, ta có thể nghiên cứu từng loại âm thanh ấy. Không chỉ động vật có tai mới nghe âm thanh. Có nghiên cứu cho thấy cả thực vật không có tai cũng phản ứng với âm thanh.

Các nhà khoa học cũng bắt được những âm thanh như vậy dưới nước. Micro đặt dưới nước được gọi là hydrophone. Deep learning phân tích các âm thanh thu được như thế. Deep learning là kỹ thuật tự tìm quy luật từ rất nhiều dữ liệu. Trí tuệ nhân tạo đang phá vỡ chiếc hộp giác quan của con người như vậy. Cánh cửa đi vào cuộc đối thoại lâu đời của Trái Đất đang mở ra.

Tài liệu căn cứ liên quan

- [1] Bakker, K. (2022). *The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants*. Princeton University Press.
- [2] Earth Species Project. (2024-11-11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. <https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics>
- [3] Poole, J. H., Payne, K., Langbauer, W. R., & Moss, C. J. (1988). The social contexts of some very low frequency calls of African elephants. *Behavioral Ecology and Sociobiology*, 22(6), 385-392. <https://doi.org/10.1007/BF00294975>

0.2 Trí tuệ nhân tạo biến lời nói thành bản đồ

Trí tuệ nhân tạo xử lý lời nói của con người như thế nào. Nó biến lời nói thành một loại bản đồ. Bản đồ này được gọi là không gian tiềm ẩn (Latent Space). Đó là bản đồ vô hình đặt các từ có nghĩa giống nhau ở gần nhau. Có thể dễ hình dung bản đồ này như bầu trời đêm đầy sao. Mỗi từ trở thành một ngôi sao. Những từ có nghĩa giống nhau tụ lại gần nhau. Những từ có nghĩa xa nhau thì nằm cách xa nhau [1].

Sức mạnh của bản đồ này không chỉ đến từ khoảng cách. Quan hệ giữa các từ cũng hiện ra bằng hướng và khoảng cách. Ví dụ, có một mũi tên nối “vua” với “đàn ông”. Cũng có một mũi tên nối “nữ hoàng” với “phụ nữ”, và nó giống mũi tên kia. Vì vậy, nếu lấy “nữ hoàng” trừ “phụ nữ” rồi cộng “đàn ông”, kết quả sẽ gần với “vua”. Trong một số bản đồ từ, phép tính như vậy khá đúng. Từ “chó” cũng không đứng một mình. Nó nối với các từ như “người”, “tiếng sủa”, “lông” rồi có vị trí của mình.

Máy tính không cảm nhận nghĩa bên trong của từ như con người. Nó cũng không biết cảm xúc như buồn hay vui. Việc máy tính làm chỉ là tính toán quan hệ. Nó xét xem các từ nối với nhau như thế nào. Việc này giống như giải một bài điền chỗ trống rất lớn. Giống như lần lượt tìm từ điền vào ô trống, trí tuệ nhân tạo lần lượt tìm vị trí của từng từ. Nhờ vậy, hình dạng tinh vi của lời nói được tạo ra [2].

Âm thanh cũng có thể được đổi thành bản đồ như thế. Trí tuệ nhân tạo cắt nhỏ âm thanh rồi đổi thành con số. Nó xếp các con số ấy như những ngôi sao trên bản đồ. Âm thanh giống nhau tụ lại gần nhau. Âm thanh khác nhau thì nằm xa nhau. Dù là lời nói của người hay tiếng của động vật, cách đặt lên bản đồ khá giống nhau. Tuy vậy, việc ý nghĩa có nối được qua các loài hay không vẫn đang được nghiên cứu [3].

Lời nói giống như cái bóng do thế giới tạo ra. Trí tuệ nhân tạo lần ngược theo cái bóng đó. Vì vậy, nó vẽ ra bản đồ cấu trúc ẩn của thế giới. Cần có bản đồ này thì câu chuyện tiếp theo mới có thể bắt đầu. Bởi nếu có thể vẽ lời nói của con người thành bản đồ, ta cũng có thể vẽ tiếng của động vật thành bản đồ.

Tài liệu căn cứ liên quan

[1] Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. Proceedings of NAACL-HLT 2013, 746-751.

[2] Bakker, K. (2022). The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants. Princeton University Press.

[3] Earth Species Project. (2026). What we do: Technology. Earth Species Project.
<https://www.earthspecies.org/what-we-do/technology>

0.3 Dịch không cần từ điển và Earth Species Project

Vào khoảng năm 2017, một sự kiện lớn xảy ra trong dịch máy. Khi đó hoàn toàn không có từ điển nối hai ngôn ngữ. Ngày xưa cần có Rosetta Stone. Rosetta Stone là tấm bia cổ khắc cùng một văn bản bằng nhiều ngôn ngữ. Nhờ tấm bia này, người ta giải được chữ Ai Cập cổ. Nhưng trí tuệ nhân tạo đã ghép các từ của hai ngôn ngữ với nhau mà không cần loại từ điển như vậy. Nó làm được bằng machine learning, tức kỹ thuật máy tự học [1].

Cách làm là đặt chồng bản đồ của hai ngôn ngữ lên nhau. Trí tuệ nhân tạo tạo riêng một bản đồ sao của tiếng Anh. Nó cũng tạo riêng một bản đồ sao của tiếng Tây Ban Nha. Sau đó, nó cẩn thận đặt chồng hai bản đồ. Nhiều học giả cho rằng hai bản đồ sẽ không khớp. Vì mỗi ngôn ngữ có cách nhìn thế giới khác nhau. Nhưng hai bản đồ lại chồng lên nhau một cách mềm mại. Ngôi sao có nghĩa là “chó” gần như rơi vào cùng một vị trí. Cách kéo giãn và thu nhỏ hai bản đồ có hình dạng khác nhau để đặt khớp lên nhau được gọi là sắp thẳng hàng đa tạp (manifold alignment).

Việc này không chỉ xảy ra với các ngôn ngữ châu Âu. Những ngôn ngữ có ngữ pháp rất khác nhau cũng cho thấy hình dạng tương tự. Ngay cả Esperanto, một ngôn ngữ do con người tạo ra, cũng như vậy.

Nhiều ngôn ngữ của con người dường như cùng mang một hình dạng lớn. Có thể vì con người, với tư cách một loài, cùng cảm nhận một Trái Đất [2].

Aza Raskin của Earth Species Project đã đặt ra một câu hỏi từ đây. Nếu có thể chồng các ngôn ngữ của con người như vậy, liệu có thể dùng cách đó với tiếng của động vật không. Tổ chức này là một viện nghiên cứu phi lợi nhuận được thành lập năm 2017. Họ thu thập tiếng của beluga, cá voi sát thủ và đười ươi. Với mỗi loại tiếng, họ tạo một bản đồ sao riêng. Rồi họ cẩn thận đặt bản đồ đó lên bản đồ lời nói của con người. Khi đó, con người có thể nhìn tiếng nói của sự sống láng giềng bằng đôi mắt mới. Công cụ gần đây của họ là NatureLM-audio. Đây là trí tuệ nhân tạo nghe tiếng động vật rồi đoán loài và kiểu tiếng kêu. Năm 2025, bản demo cho mọi người thử dùng đã được công bố. Khi tải tệp âm thanh lên và viết câu hỏi, câu trả lời sẽ hiện ra. Năm 2026, chức năng đó đã tốt hơn. Tháng 7 năm 2026, một kho dữ liệu tên là alp-data, nơi nhiều viện nghiên cứu cùng dùng dữ liệu, cũng được mở. Dòng chảy này nối với ý tưởng rằng các trí tuệ nhân tạo khác nhau cuối cùng sẽ đi tới những bản đồ có hình dạng giống nhau. Ý tưởng này được gọi là giả thuyết biểu trưng kiểu Plato (Platonic Representation Hypothesis) [3][4][5].

Tài liệu căn cứ liên quan

- [1] Lample, G., Conneau, A., Denoyer, L., & Ranzato, M. (2018). Unsupervised machine translation using monolingual corpora only. International Conference on Learning Representations (ICLR) 2018. arXiv:1711.00043.
- [2] Robinson, D., Miron, M., Hagiwara, M., et al. (2024/2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. arXiv:2411.07186.
- [3] Earth Species Project. (2025-09). Hands-on with NatureLM-audio: A no-code demo. Earth Species Project. <https://earthspecies.org/blog/hands-on-with-naturelm-audio-a-no-code-demo>
- [4] Earth Species Project. (2026-07-22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. <https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>
- [5] Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). Position: The Platonic Representation Hypothesis. Proceedings of the 41st International Conference on Machine Learning (ICML). arXiv:2405.07987.

0.4 Tìm quy luật ẩn trong âm thanh

Tiếng của động vật có phải chỉ là tiếng ồn vô nghĩa không. Hay nó cũng có quy luật như lời nói của con người. Các nhà khoa học đo thống kê của âm thanh để tìm hiểu điều này. Một quy luật được biết rộng rãi là lời càng dùng thường xuyên thì càng ngắn. Quy luật này được gọi là luật rút ngắn của Zipf (Zipf's law of abbreviation). Nó được đặt theo tên nhà ngôn ngữ học George Kingsley Zipf. Con người làm ngắn các từ hay dùng để tiết kiệm sức [1].

Nhà ngôn ngữ học Stuart Semple và nhóm nghiên cứu của Ramon Ferrer-i-Cancho đã xem xét quy luật này ở nhiều loài động vật. Họ thu thập và so sánh tiếng của nhiều loài thú, trong đó có khỉ. Máy tính chia nhỏ âm thanh rồi đếm. Kết quả rất nhất quán. Tiếng càng được phát ra thường xuyên thì càng ngắn. Trong tiếng động vật cũng có quy luật giống lời nói của con người. Một quy luật khác cũng xuất hiện. Âm thanh càng dài thì các mảnh bên trong nó càng ngắn. Quy luật này có tên là luật Menzerath-Altmann (Menzerath-Altmann law).

Cũng có một giá trị đo mức khó đoán của âm thanh tiếp theo. Nó đo xem sau một âm thanh thì âm thanh nào sẽ đến, và việc đoán đó khó đến đâu. Giá trị này được gọi là entropy Shannon (Shannon entropy). Nó được đặt theo tên Claude Shannon, người tạo ra lý thuyết thông tin. Nếu âm thanh chỉ là tiếng hét vô nghĩa, giá trị này sẽ nghiêng hẳn về một phía. Nếu đó là lời nói có quy luật, nó sẽ có giá trị nằm ở giữa [4].

Nhóm nghiên cứu Whale-SETI đã xem xét tiếng cá voi bằng giá trị này. Brenda McCowan, Fred Sharpe và Laurence Doyle cùng tham gia. Họ nghiên cứu bài hát của cá voi lưng gù (humpback whale). Bài hát

của cá voi nằm vừa phải giữa quy luật và tự do. Trong bài hát cũng có cấu trúc: âm thường gặp thì ngắn, âm hiếm gặp thì dài. Làm như vậy có thể gửi cùng một ý nghĩa bằng ít sức hơn. Trong máy tính cũng có cách nén để giảm kích thước bằng cách ghi thứ thường dùng ngắn hơn và thứ hiếm dùng dài hơn. Cách này có tên là mã hóa Huffman (Huffman coding) [2].

Nhóm nghiên cứu cũng làm thí nghiệm trao đổi âm thanh với cá voi sống. Họ gặp một con cá voi lưng gù tên là Twain ở vùng biển Alaska. Nhóm nghiên cứu phát tiếng chào của cá voi đã ghi âm trước đó bằng loa dưới nước. Trí tuệ nhân tạo không tạo ra cuộc trò chuyện theo thời gian thực. Tiếng chào này được gọi là whup call. Cá voi trả lời theo lượt như con người. Nó đợi đến chỗ âm thanh kết thúc rồi lấp vào khoảng trống. Sự trao đổi này kéo dài 20 phút. Có thể nói chúng trao đổi âm thanh giống như bạn bè lần lượt nói chuyện trong giờ ăn trưa [3].

Tài liệu căn cứ liên quan

- [1] Semple, S., Ferrer-i-Cancho, R., & Gustison, M. L. (2022). Linguistic laws in biology. *Trends in Ecology & Evolution*, 37(1), 53-66. <https://doi.org/10.1016/j.tree.2021.08.012>
- [2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "conversing" with an Alaskan humpback whale. *PeerJ*, 11, e16349. <https://doi.org/10.7717/peerj.16349>
- [3] Earth.com. (2024). AI helps humans have a 20-minute 'conversation' with a humpback whale named Twain. Earth.com. <https://www.earth.com/animals/whale-seti-ai-helps-humans-have-20-minute-conversation-with-humpback-whale-twain/>
- [4] Bakker, K. (2022). *The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants*. Princeton University Press.

0.5 Nguyên âm tìm thấy trong tiếng click của cá nhà táng

Nơi phương pháp này tỏa sáng nhất là biển sâu. Nhân vật chính là cá nhà táng. Cá nhà táng có chiều dài cơ thể tới 20 mét. Não của nó nặng khoảng 9 kilôgam. Con số này lớn hơn hơn 6 lần trọng lượng não người. Loài cá voi này săn mồi ở biển sâu không có ánh sáng. Vì vậy, nó nhìn thế giới bằng âm thanh thay cho mắt. Nó phát ra những tiếng click ngắn để xem xét xung quanh. Cụm tiếng click này được gọi là coda [1].

Trong thời gian dài, các học giả xem coda là tín hiệu không có ý nghĩa. Họ nghĩ chỉ có vài âm báo cố định. Năm 2020, Project CETI đã thay đổi cách nghĩ này. Đây là nhóm gồm các nhà khoa học trí tuệ nhân tạo và nhà nghiên cứu cá voi từ nhiều đại học. Họ thu thập bản ghi âm ở vùng biển Dominica. Có hơn 8.700 coda được thu thập trong hơn 10 năm. Con số chính xác là 8.719. Trí tuệ nhân tạo đã phân tích các âm thanh này. Nó tìm ra nhiều tổ hợp hơn hẳn 21 kiểu đã được biết trước đó. Nó cũng phát hiện cá voi thay đổi tốc độ và khoảng cách giữa các âm thanh. Đôi khi cá voi gắn thêm tiếng click ở cuối như dấu chấm của câu. Nó cũng thay đổi độ dài của coda từng chút một. Sự thay đổi nhịp rất nhỏ này được gọi là rubato. Nó giống như người hát kéo chậm hoặc đẩy nhanh nhịp một chút. Đây là cách ghép các mảnh âm thanh nhỏ để tạo nhiều ý nghĩa. Có thể nói phương thức cũng có trong lời nói của con người này đã được xác nhận rõ ở động vật. Cá nhà táng cũng dùng coda khác nhau tùy theo tình huống nghe. Khả năng chọn lời theo đối tượng như vậy được gọi là thiết kế theo người nghe (audience design).

Năm 2025, một nghiên cứu mới báo tin về một phát hiện lớn hơn. Đây là nghiên cứu do Đại học Berkeley ở Mỹ và Project CETI cùng thực hiện. Giáo sư Gašper Beguš dẫn dắt nghiên cứu này. Nhóm nghiên cứu thu hẹp khoảng trống giữa các tiếng tách tách của cá nhà táng. Rồi họ phát lại âm thanh nhanh hơn ban đầu vài lần. Khi đó, những âm rời nhau nối lại thành một chuỗi. Âm đó nghe giống nguyên âm của con người. Nhóm nghiên cứu phân biệt được âm "a" và âm "i". Họ cũng tìm thấy nguyên âm đôi chảy từ một nguyên âm sang nguyên âm khác. Nhóm nghiên cứu còn dùng công cụ để nhìn vào bên trong trí tuệ nhân tạo. Họ xem trí tuệ nhân tạo phản ứng với phần nào khi phán đoán tiếng cá voi. Kết quả cho thấy

các đỉnh cộng hưởng của âm thanh rất quan trọng. Đỉnh cộng hưởng này được gọi là formant. Đó là phần âm nổi bật tùy theo hình dạng khoang vang trong miệng hoặc trong đầu. Con người cũng dùng giá trị này khi phân biệt nguyên âm. Năm 2026, một nghiên cứu tiếp theo dựa trên kết quả này đã xuất hiện. Nghiên cứu đó so sánh cấu trúc tiếng cá voi với âm vị học của con người và xem xét sâu hơn [2][5][3].

Tháng 3 năm 2026, Project CETI công bố một nghiên cứu khác. Đó là nghiên cứu ghi lại cảnh cá nhà táng sinh con ở vùng biển Dominica. Nghiên cứu ghi hơn sáu giờ âm thanh cùng với video quay từ trên không. Những con cá voi khác trong đàn đã giúp việc sinh con. Đây là lần đầu tiên sự hợp tác như vậy được làm rõ bằng số liệu ở động vật không phải linh trưởng. Nghiên cứu này lại cho thấy xã hội của cá nhà táng được nối với nhau chặt chẽ bằng âm thanh [4].

Cá nhà táng không có dây thanh như con người. Thay vào đó, chúng dùng cơ quan lớn ở phía trước đầu và mô trong mũi. Chúng làm rung nơi này để tạo độ vang của âm thanh. Một loài sống đi theo con đường hoàn toàn khác con người đã tự tạo ra nguyên âm giống con người.

Tài liệu căn cứ liên quan

- [1] Sharma, P., Gero, S., Rus, D., Goldwasser, S., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [2] Beguš, G., Sprouse, R. L., Leban, A., Silva, M., & Gero, S. (2025). Vowel- and diphthong-like spectral patterns in sperm whale codas. *Open Mind*, 9. <https://doi.org/10.1162/OPMI.a.252>
- [3] Beguš, G., Sprouse, R. L., Leban, A., & Gero, S. (2026). The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B*, 293. <https://doi.org/10.1098/rspb.2025.2994>
- [4] Aluma, A., Gero, S., et al. (2026). Description of a collaborative sperm whale birth and shifts in coda vocal styles during key events. *Scientific Reports*, 16, 9206. <https://doi.org/10.1038/s41598-025-27438-3>
- [5] University of California. (2025-11-13). UC Berkeley and Project CETI study shows sperm whales communicate in ways similar to humans. University of California. <https://www.universityofcalifornia.edu/news/uc-berkeley-and-project-ceti-study-shows-sperm-whales-communicate-ways-similar-humans>

0.6 Lắng nghe, không phải thống trị

Trong nghiên cứu trí tuệ nhân tạo có một bài học cũ. Đó là bài học cay đắng (The Bitter Lesson) do nhà khoa học máy tính Richard Sutton nói đến. Nó có nghĩa là dữ liệu nhiều và tính toán lớn thường thắng thiết kế thông minh. Công cụ giải mã tiếng động vật cũng đã lớn lên theo cách này. Một trong số đó là DolphinGemma. Google, Viện Công nghệ Georgia và Wild Dolphin Project đã cùng tạo ra nó. Wild Dolphin Project là nghiên cứu về cá heo đã kéo dài hơn 40 năm. DolphinGemma là một trí tuệ nhân tạo nhỏ. Bên trong nó có khoảng 400 triệu con số để phán đoán âm thanh. Vì nhỏ nên nó chạy được cả trên điện thoại Pixel cầm tay. DolphinGemma chia âm thanh thành các mảnh nhỏ bằng một phương pháp gọi là SoundStream. Nó nhìn các mảnh đó và đoán trước âm tiếp theo. Cách này giống trí tuệ nhân tạo viết tiếp lời người đến từ kế tiếp. Công cụ này hiện vẫn đang được thử nghiệm khi đi theo cá heo trên biển. Google muốn mở công cụ này để ai cũng có thể dùng [3][1].

Nhưng công nghệ nhanh này cũng có mặt đáng sợ. Aza Raskin nhiều lần nói về điểm này. Trí tuệ nhân tạo có thể bắt chước âm thanh trước khi hiểu ý nghĩa. Nó có thể tạo ra tiếng cá voi y hệt và tiếp tục đối thoại. Như vậy là đồ âm thanh giả xuống biển khi chưa hiểu nghĩa. Cá nhà táng chia sẻ văn hóa bằng âm thanh và hành động. Ngày xưa, khi tàu săn cá voi xuất hiện, cá nhà táng đã học từ nhau cách tránh nguy hiểm. Chỉ hai năm rưỡi sau khi tàu đến, số cá voi bị bắt giảm 58 phần trăm. Người ta cho rằng đó là kết quả của việc cá voi báo nguy cho nhau. Sự giao tiếp của chúng nổi bật hơn ta nghĩ. Bài hát của cá voi lưng gù cũng là một văn hóa lan xa qua biển. Âm thanh giả do con người tạo ra có thể làm rối sự nổi kết này. Tạo âm thanh trước khi hiểu hết ý nghĩa là việc nguy hiểm [2][4][5].

Vì vậy, hướng đi của hành trình này rất quan trọng. Công nghệ này không được trở thành công cụ điều khiển động vật theo ý con người. Nó phải là việc thừa nhận và lắng nghe tâm trí của động vật đúng như vốn có. Điều này được gọi là hiểu biết giữa các loài (Interspecies Understanding). Con người thay đổi không phải khi tuân lời của người khác. Con người thay đổi khi hạ thấp giọng và đón nhận đối phương.

Ngày xưa, con người tin rằng Trái Đất là trung tâm của vũ trụ. Kính viễn vọng cho biết suy nghĩ đó là sai. Giờ đây, bản đồ âm thanh lại một lần nữa đánh thức chúng ta. Con người không phải trung tâm của thế giới sự sống. Con người chỉ là một phần trong vòng tròn lớn, nơi mỗi loài đều có tiếng nói riêng. Nhận ra điều này là bước đầu tiên để mở ra tiếng nói của Trái Đất.

Tài liệu căn cứ liên quan

- [1] Google. (2025). DolphinGemma: How Google AI is helping decode dolphin communication. Google Blog. <https://blog.google/innovation-and-ai/products/dolphingemma/>
- [2] Robinson, D., Miron, M., Hagiwara, M., et al. (2024/2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. arXiv:2411.07186.
- [3] Sutton, R. S. (2019-03-13). The bitter lesson. incompleteideas.net. <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>
- [4] Whitehead, H., Smith, T. D., & Rendell, L. (2021). Adaptation of sperm whales to open-boat whalers: Rapid social learning on a large scale? Biology Letters, 17(3), 20210030. <https://doi.org/10.1098/rsbl.2021.0030>
- [5] Bakker, K. (2022). The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants. Princeton University Press.

Chương 1. Wild Dolphin Project và thế giới của cá heo đốm

1.1 Quan sát lâu dài ở Bahamas màu ngọc lục bảo

Trong biển có trí thông minh mà con người vẫn chưa biết hết. Cá heo là một trong số đó. Muốn hiểu tâm trí của cá heo thì phải làm gì. Có một cách là nhốt cá heo trong bể rồi quan sát. Nhưng có một người đã chọn con đường khác. Đó là tiến sĩ Denise Herzing. Tiến sĩ Herzing đã tự đi đến vùng biển nơi cá heo sinh sống [1].

Tiến sĩ Herzing đặt ra một nguyên tắc. Đó là nguyên tắc "trong thế giới của chúng, theo cách của chúng". Nghĩa là không bắt cá heo làm theo ý con người. Nghĩa là quan sát mà không làm cá heo bị thương. Cách này được gọi là quan sát không xâm lấn. Không xâm lấn là cách không chọc vào cơ thể và không bắt giữ.

Nếu nhốt cá heo trong bể, ta khó thấy đáng vẻ thật của chúng. Vì cá heo bị nhốt sẽ sống theo cách khác. Vì vậy, cách đi ra biển để quan sát rất có giá trị.

Tiến sĩ Herzing dẫn dắt 'Wild Dolphin Project'. Nghiên cứu này đã kéo dài gần 40 năm. 40 năm là khoảng thời gian dài, gần gấp bốn lần tuổi của một học sinh tiểu học. Nơi nghiên cứu là vùng biển Bahamas. Nước ở đây trong và sáng như nước hồ bơi. Vì vậy, nơi này rất tốt để ghi lại cả cử động và âm thanh của cá heo [2][3].

Nghiên cứu này bắt đầu vào năm 1985. Đây là nghiên cứu lâu đời nhất thế giới trong số các nghiên cứu theo dõi cá heo dưới nước. Mỗi năm, nhóm nghiên cứu đều tìm đến cùng một đàn, không bỏ năm nào. Nhờ vậy, họ có thể quan sát trọn vòng đời của một cá thể, từ khi sinh ra, lớn lên, rồi sinh con. Những ghi chép tích lũy như vậy về sau trở thành nền tảng để dạy trí tuệ nhân tạo [4].

Ở vùng biển này có cá heo đốm Đại Tây Dương. Tên khoa học là *Stenella frontalis*. Tên khoa học là tên chính thức mà các nhà khoa học trên thế giới cùng dùng. Nơi nghiên cứu là vùng biển có bãi cát nông. Nước nông nên ánh nắng chiếu tới tận đáy. Vì vậy, có thể nhìn rõ cử động của cá heo cả từ phía trên.

Loài cá heo này nhỏ hơn cá heo mũi chai. Nhưng chúng là động vật thông minh. Có một chỉ số đo xem não lớn bao nhiêu so với cơ thể. Chỉ số này được gọi là chỉ số não hóa. Đó là giá trị so sánh não lớn bao nhiêu so với thân hình. Nhóm cá heo có chỉ số này khá cao. Trong các loài thú, chúng thuộc nhóm cao nhất sau con người [5].

Nhóm nghiên cứu đã phân biệt từng con cá heo. Vì vết thương trên vây của mỗi con đều khác nhau. Hoa văn đốm trên cơ thể cũng khác nhau. Cá heo đốm càng lớn thì đốm trên cơ thể càng nhiều. Khi còn nhỏ, chúng không có đốm, nhưng càng già thì đốm càng tăng. Vì vậy, chỉ nhìn đốm cũng có thể ước đoán tuổi. Họ đã nhớ từng con cá heo như nhớ từng khuôn mặt bạn cùng lớp. Nhờ vậy, họ có thể theo dõi ba thế hệ nối nhau: bà, mẹ và con.

Việc quan sát trong thời gian dài có giá trị vì một lý do. Thói quen của cá heo được truyền từ mẹ sang con. Con học bên cạnh mẹ cách tìm thức ăn. Việc hành vi đã học được truyền từ mẹ sang con gọi là truyền thừa văn hóa. Nếu chỉ nhìn trong thời gian ngắn, ta không thể biết sự nối tiếp này. Chỉ khi quan sát một đàn thật lâu, điều đó mới hiện ra.

Xã hội cá heo phức tạp như xã hội con người. Các con đực kết bạn suốt đời và đi cùng nhau. Các con cái chăm con trong thời gian dài và dạy phép tắc. Muốn tạo nên một xã hội phức tạp như vậy, chúng phải có cách giao tiếp với nhau. Đây là lý do ta cần giải mã âm thanh của cá heo.

Tài liệu căn cứ liên quan

- [1] Georgia Tech Research. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. Georgia Tech Research News. <https://research.gatech.edu/dolphingemma-how-google-ai-helping-decode-dolphin-communication>
- [2] Herzing, D. L., Pack, A. A., Delfour, F., Starner, T., Mason, C., Gilliland, S., Ramsey, C., & Kohlsdorf, D. (2024). Imitation of computer-generated sounds by wild Atlantic spotted dolphins (*Stenella frontalis*). *Animal Behavior and Cognition*, 11(2), 136–166. <https://doi.org/10.26451/abc.11.02.02.2024>
- [3] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. <https://blog.google/innovation-and-ai/products/dolphingemma/>
- [4] Millman, R. (2025-04-15). Google launches AI that talks to dolphins. Newsweek. <https://www.newsweek.com/dolphingemma-google-ai-dolphins-2059955>
- [5] Marino, L. (1998). A comparison of encephalization between odontocete cetaceans and anthropoid primates. *Brain, Behavior and Evolution*, 51(4), 230–238. <https://doi.org/10.1159/000006540>

1.2 Biến khắc nghiệt và hệ thống CHAT

Việc ghi âm thanh của cá heo không dễ. Vì trong âm thanh của cá heo có nhiều âm cao mà con người không nghe được. Tai người có thể nghe tới khoảng 20 kilohertz. Kilohertz là đơn vị đo độ cao thấp của âm thanh. Cá heo còn phát ra âm cao gần 150 kilohertz. Âm đó cao hơn rất nhiều so với âm con người nghe được [1].

Cá heo dùng nhiều loại âm thanh. Có ba loại tiêu biểu. Một loại là tiếng huýt sáo lên xuống. Một loại khác là tiếng click lách tách. Loại còn lại là âm burst pulse bật ra ngắn. Burst pulse là âm bật ra liên tiếp thành nhiều mảnh nhỏ. Cá heo thường dùng khi tranh giành hoặc nói chuyện ở gần nhau. Tiếng click va vào vật rồi dội lại. Cá heo biết phía trước có gì nhờ tiếng vang đó. Điều này được gọi là định vị bằng tiếng vang. Định vị bằng tiếng vang là khả năng biết vị trí của vật bằng âm thanh.

Có một vấn đề. Nếu chỉ đặt một chiếc micro xuống nước thì không có ích. Vì không thể biết trong nhiều con, con nào đã phát ra âm thanh. Vì vậy, tiến sĩ Herzing đã tìm đến một kỹ sư. Đó là tiến sĩ Thad Starner ở Georgia Tech. Tiến sĩ Starner là người tạo ra công nghệ gắn máy tính nhỏ lên cơ thể.

Tiến sĩ Starner nghĩ việc này sẽ xong nhanh. Nhưng biến là một đối tượng khó. Chỉ riêng việc làm cho máy ổn định đã mất 4 năm. Kẻ thù lớn nhất là nước mặn. Nước mặn chui vào khe hở và làm cháy máy [2].

Chiếc hộp nhôm đựng máy cũng gây rắc rối. Nó hút ánh nắng nóng và làm bên trong nóng lên. Khi đó, máy tính dừng lại. Nhôm cũng chặn tín hiệu không dây. Nhóm nghiên cứu đổi mặt sau của hộp thành nhựa trong suốt. Dây điện và mạch được bọc kín hoàn toàn bằng nhựa đặc biệt. Khi đó máy mới chạy tốt dưới nước.

Việc để máy lâu trong nước cũng không dễ. Vì muối ăn mòn kim loại từng chút một. Nhóm nghiên cứu rửa và kiểm tra máy mỗi lần dùng. Chỉ một khe nhỏ cũng có thể làm máy hỏng.

Cũng có nhiều việc nguy hiểm. Nếu pin lớn mà thợ lặn mang chạm vào nước, nó có thể nổ. Dòng điện yếu rò ra từ máy cũng có thể gọi cá mập đến. Vì cá mập có giác quan cảm nhận tín hiệu điện yếu. Thực tế, trong lúc thử nghiệm đã có lần cá mập xuất hiện.

Cá mập cảm nhận được điện yếu nhờ một cơ quan cảm giác. Cơ quan này được gọi là cơ quan Lorenzini. Nó hiện ra như những lỗ nhỏ quanh mũi cá mập. Nhờ vậy, cá mập cảm nhận cả dòng điện yếu do con mồi phát ra. Vì thế, nhóm nghiên cứu lo rằng máy dưới nước có thể kích thích giác quan này.

Thiết bị ra đời sau những vất vả đó là hệ thống CHAT. Cái tên CHAT có ý nghĩa riêng. Dịch sang tiếng Hàn, đó là thiết bị đo từ xa tăng cường thính giác cho động vật dạng cá voi. Nói dễ hiểu, đó là máy tính

đeo trên người để gửi, nhận và nhận biết âm thanh dưới nước. Động vật dạng cá voi là cách gọi chung cá heo và cá voi [3].

CHAT là máy tính dưới nước để nghe và ghi âm thanh của cá heo. Máy này ghi âm rất dày. Nó chia và nhận âm thanh hơn 190.000 lần trong 1 giây. Nghĩa là nó nghe dày hơn tai người rất nhiều. Phải nghe dày như vậy mới không bỏ lỡ cả âm cao. Vì thế, bộ nhớ đầy rất nhanh. Do đó, nó được dùng cùng thiết bị lưu trữ lớn.

Thợ lặn đeo máy này trên lưng và đeo một bàn phím nhỏ ở cổ tay. Âm thanh được nghe bằng tai nghe dẫn truyền qua xương. Dẫn truyền qua xương là cách truyền âm bằng cách làm rung xương sọ thay cho tai. Khi thợ lặn bấm bàn phím, âm đã hẹn trước được phát xuống nước. Âm này không phải âm cá heo vốn phát ra. Đó là âm do con người mới tạo. Họ cố ý chọn âm mới để không bị lẫn với âm cá heo. Máy của thợ lặn khác nhận ra âm này. Nếu khớp với âm đã định trước, máy sẽ cho biết tên đồ vật gắn với âm đó.

Khi gửi và nhận âm thanh, máy báo bằng giọng nói. Khi thợ lặn gửi âm, họ nghe giọng nam. Ở phía nhận âm, giọng nữ đọc ý nghĩa của âm đó. Hai giọng được tách ra để không nhầm ai đã nói.

Tài liệu căn cứ liên quan

- [1] Georgia Tech Research. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. Georgia Tech Research News. <https://research.gatech.edu/dolphingemma-how-google-ai-helping-decode-dolphin-communication>
- [2] Herzing, D. L., Pack, A. A., Delfour, F., Starner, T., Mason, C., Gilliland, S., Ramsey, C., & Kohlsdorf, D. (2024). Imitation of computer-generated sounds by wild Atlantic spotted dolphins (*Stenella frontalis*). *Animal Behavior and Cognition*, 11(2), 136–166. <https://doi.org/10.26451/abc.11.02.02.2024>
- [3] Kohlsdorf, D., Gilliland, S., Presti, P., Starner, T., & Herzing, D. (2013). An underwater wearable computer for two way human-dolphin communication experimentation. *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC)*, 147–148. <https://doi.org/10.1145/2493988.2494346>

1.3 Trò chơi giao tiếp học qua việc chơi cùng nhau

Dù đã làm xong máy, vấn đề vẫn còn. Đó là dạy âm thanh cho cá heo bằng cách nào. Họ quyết định không dùng cách huấn luyện ép buộc. Thay vào đó, họ chọn cách để cá heo tự học [1].

Cách tiến sĩ Herzing chọn có tên gọi. Đó là phương pháp 'mô hình hóa cạnh tranh xã hội'. Đây là cách cho thấy hai người hỏi đáp với nhau. Tiến sĩ Irene Pepperberg đã dùng cách này khi dạy vẹt. Khi cho động vật thấy hai người thi đua, chúng học theo ở bên cạnh. Cách làm như sau. Hai người trình diễn trò chơi trước mặt cá heo. Một người cầm đồ vật, người kia phát âm của đồ vật đó. Người phát âm nhận đồ vật rồi cùng chơi.

Trò chơi thật diễn ra như sau. Một thợ lặn cầm chiếc khăn vàng trong tay. Một thợ lặn khác bấm bàn phím ở cổ tay để phát âm của khăn. Khi đó, thợ lặn cầm khăn đưa khăn cho thợ lặn đã phát âm. Hai người kéo chiếc khăn dưới nước và chơi vui vẻ.

Đồ chơi dùng để chơi với cá heo có nhiều loại. Không chỉ khăn vàng, họ còn chơi bằng rong mơ trôi trên biển. Rong mơ là loài rong biển màu nâu nổi trên mặt biển. Họ cũng chơi bằng cách kéo một sợi dây dài. Cá heo còn thích chơi trò cưỡi con sóng trước mũi thuyền. Mỗi đồ chơi được gắn với một âm khác nhau để gửi và nhận.

Cá heo quan sát trò chơi này bên cạnh. Chúng nhận ra quy tắc rằng khi phát âm thì đồ chơi được chuyển qua lại. Chẳng bao lâu, cá heo bắt đầu phát lại những âm đó. Chúng bắt chước y nguyên độ lên xuống của tiếng huýt sáo do con người tạo ra. Việc phát lại âm thanh như vậy gọi là bắt chước âm thanh. Các

nhà khoa học gọi đó là bắt chước âm học. Việc quan sát trò chơi như vậy về sau trở thành nền tảng cho nghiên cứu trí tuệ nhân tạo [2][3].

Cá heo đã nhiều lần tự bắt chước âm này. Có lúc chúng nghe một lần rồi bắt chước ngay. Có lúc chúng bắt chước sau một lát. Điều đó cho thấy chúng có thể linh hoạt bắt chước âm mới. Nhưng có một điểm cần nêu rõ. Vẫn chưa xác nhận được liệu chúng có hiểu và dùng âm đó để chỉ đồ vật hay không. Việc bắt chước âm thanh đã thành công. Nhưng chưa có bằng chứng chắc chắn rằng cá heo dùng âm như tên đồ vật.

Vì sao cá heo tự phát âm. Không phải chỉ để kiếm thức ăn. Chúng muốn chơi cùng con người. Khi trò chơi vui, cá heo chủ động lại gần. Niềm vui đã dẫn dắt giao tiếp.

Thí nghiệm này cho thấy một sự thật quan trọng. Đó là cá heo có thể học và phát lại âm mới. Điều đó cũng cho thấy không cần ép chúng học tiếng người. Có thể nói cá heo đã chủ động lại gần và chia sẻ âm thanh với con người.

Tài liệu căn cứ liên quan

- [1] Georgia Tech Research. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. Georgia Tech Research News.
<https://research.gatech.edu/dolphingemma-how-google-ai-helping-decode-dolphin-communication>
- [2] Herzing, D. L., Pack, A. A., Delfour, F., Starner, T., Mason, C., Gilliland, S., Ramsey, C., & Kohlsdorf, D. (2024). Imitation of computer-generated sounds by wild Atlantic spotted dolphins (*Stenella frontalis*). *Animal Behavior and Cognition*, 11(2), 136–166. <https://doi.org/10.26451/abc.11.02.02.2024>
- [3] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword.
<https://blog.google/innovation-and-ai/products/dolphingemma/>

1.4 Bản đồ âm thanh do DolphinGemma vẽ

Các quy tắc âm thanh do con người đặt ra có giới hạn. Vì âm thanh thật mà cá heo trao đổi trong đàn phức tạp hơn nhiều. Muốn giải mã âm thanh phức tạp này, cần một công cụ mạnh. Công cụ đó là trí tuệ nhân tạo [1].

Năm 2025 có một tin lớn. Google DeepMind và Wild Dolphin Project cùng công bố một trí tuệ nhân tạo. Google DeepMind là tổ chức nghiên cứu trí tuệ nhân tạo của Google. Tên của trí tuệ nhân tạo này là 'DolphinGemma'. DolphinGemma dựa trên một trí tuệ nhân tạo nhỏ tên là Gemma do Google tạo ra. Họ dạy nó bằng âm thanh và video cá heo được thu thập trong 40 năm [2].

Thiết bị trước đây là máy tính ba lô nặng đeo trên lưng. Thợ lặn phải mang chiếc máy nặng đó xuống biển. Bây giờ, một chiếc điện thoại thông minh bằng lòng bàn tay làm việc ấy. Thiết bị đã nhỏ đi rất nhiều. Có lý do khiến DolphinGemma được làm nhỏ. Trí tuệ nhân tạo này có khoảng 400 triệu giá trị điều chỉnh dùng khi học. Những giá trị điều chỉnh này được gọi là tham số. Với một trí tuệ nhân tạo lớn, con số này thuộc loại nhỏ. Vì vậy, nó chạy được cả trên điện thoại thông minh trong tay. Nhóm nghiên cứu đưa trí tuệ nhân tạo này vào điện thoại Google Pixel. Nhờ đó, họ có thể dùng ngay trên biển mà không cần đặt máy tính nặng trên thuyền [4].

Tên DolphinGemma cũng có ý nghĩa. Nó ghép Gemma với Dolphin, nghĩa là cá heo. Nghĩa là trí tuệ nhân tạo này được tạo ra bằng cách đặt âm thanh cá heo lên trên trí tuệ nhân tạo Gemma.

Việc khó nhất là biến âm thanh thành số. Âm thanh dưới nước chảy liên tục không ngắt. Trí tuệ nhân tạo thích dữ liệu được chia thành các mảnh rời. Vì vậy, nhóm nghiên cứu cắt âm thanh thành mảnh nhỏ. Giống như chia chữ viết thành phụ âm và nguyên âm, họ chia âm thanh thành nhiều phần nhỏ. Những mảnh âm thanh này được gọi là token âm học. Nghĩa là các mảnh được chia nhỏ từ âm thanh.

Cũng có một công cụ dùng khi chia âm thanh. Đó là công nghệ chia nhỏ âm thanh của Google tên SoundStream. Công nghệ này chia âm thanh dưới nước thành nhiều mảnh nhỏ. DolphinGemma học bằng cách nối các mảnh đó lại.

Dựa vào các mảnh âm thanh này, DolphinGemma luyện đoán trước âm tiếp theo. Nó nhìn âm phía trước rồi đoán âm sẽ đến phía sau. Kết quả là DolphinGemma đã đoán khá tốt âm tiếp theo. Điều này giống như con người hiểu quy tắc của lời nói thì có thể đoán lời tiếp theo.

Đoán đúng âm tiếp theo có nghĩa là hiểu quy tắc. Trí tuệ nhân tạo đã tìm thấy quy tắc như vậy trong âm thanh của cá heo.

DolphinGemma cũng tự gom các âm thanh giống nhau lại. Trí tuệ nhân tạo biến âm thanh thành các điểm rồi xếp ra như một tấm bản đồ. Âm thanh giống nhau được đặt gần nhau. Âm thanh khác nhau được đặt xa nhau. Nhóm nghiên cứu đã thu thập âm thanh trong nhiều tình huống trong thời gian dài. Ví dụ, có tiếng huýt sáo chữ ký khi mẹ và con gọi nhau. Tiếng huýt sáo chữ ký là âm thanh khác nhau ở mỗi con cá heo, giống như tên gọi. Cũng có burst pulse phát ra khi đánh nhau. Cũng có tiếng click phát ra khi tìm bạn tình hoặc đuổi theo cá mập. Ghép các âm thanh này với tình huống có một điểm tốt. Nó giúp trí tuệ nhân tạo tìm manh mối của âm thanh nhanh hơn.

Tiến sĩ Hejing giải thích việc này như sau. Bà nói rằng nếu con người tự xem hết dữ liệu này bằng tay thì sẽ mất 150 năm. Bà nói thêm rằng trí tuệ nhân tạo làm việc đó nhanh hơn nhiều. 150 năm là thời gian dài hơn sống hai đời người [3].

Tài liệu căn cứ liên quan

- [1] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. <https://blog.google/innovation-and-ai/products/dolphingemma/>
- [2] Google DeepMind. (2025). DolphinGemma. Google DeepMind. <https://deepmind.google/models/gemma/dolphingemma/>
- [3] Scientific American. (2025). DolphinGemma could enable AI communication with dolphins [Podcast]. Scientific American. <https://www.scientificamerican.com/podcast/episode/dolphingemma-could-enable-ai-communication-with-dolphins/>
- [4] Lanz, J. A. (2025-04-14). Google launches AI to decode dolphin talk, runs on Pixel phones. Decrypt. <https://decrypt.co/314766/google-launches-ai-decode-dolphin-talk>

1.5 Cuộc trò chuyện thật bắt đầu từ biển

Vào mùa hè năm 2025, một cảnh mới mở ra. Wild Dolphin Project bắt đầu dùng DolphinGemma ngay tại biển. Trước đó, nó chỉ học bên trong máy tính. Bây giờ, thợ lặn mang trí tuệ nhân tạo đó xuống nước. Nhóm nghiên cứu dùng điện thoại Google Pixel mới ra [1][9].

Cách dùng ở biển là như sau. Thợ lặn mang theo điện thoại Pixel đặt trong hộp chống nước. Trí tuệ nhân tạo kiểm tra âm thanh ngay dưới nước. Rồi nó tìm các mẫu âm thanh lặp lại và báo cho nhóm nghiên cứu. Họ không cần quay lại thuyền để xem dữ liệu. Họ kiểm tra ngay dưới nước.

Nghiên cứu trên biển lần này có một điểm đặc biệt. DolphinGemma tự tạo ra âm thanh mới giống âm thanh của cá heo. Đó là âm thanh khó tạo bằng chương trình mà con người từng dùng. Nhóm nghiên cứu dự định dùng âm thanh này trong nghiên cứu trao đổi với cá heo sau này. Họ muốn thử nói chuyện với cá heo bằng âm thanh mới [2].

Công nghệ này được dùng rất cẩn thận. Vì nếu con người phát âm thanh bừa bãi, xã hội cá heo có thể bị rối loạn. Đàn cá heo đã kết nối với nhau bằng âm thanh trong thời gian rất dài. Vì vậy, nhóm nghiên cứu cố không làm xáo trộn âm thanh vốn có của cá heo. Âm thanh dạy cho cá heo cũng được chọn là tiếng huýt sáo mới không có trong tự nhiên. Mục đích là để không nhầm với âm thanh có sẵn. Nhóm nghiên

cứu cũng cố không làm cản dòng trao đổi âm thanh của cá heo. Họ quan sát những khoảng lặng khi cá heo ngừng phát tiếng. Đó là phép lịch sự giống như chờ đến lượt trong hàng ăn.

Không chỉ cá heo được nghiên cứu như vậy. Cũng có một tổ chức tên là Earth Species Project. Tổ chức này dùng trí tuệ nhân tạo để xem xét âm thanh của nhiều loài động vật. Đối tượng là các loài khác nhau như chim, cá voi và voi. Vì mỗi loài động vật phát âm thanh vì lý do khác nhau. Khi so sánh nhiều loài, quy luật của âm thanh hiện rõ hơn. Năm 2026, họ cũng mở một kho mới để dùng chung dữ liệu âm thanh động vật. Tên của nó là alp-data. Tên này lấy từ các chữ cái đầu trong tiếng Anh, có nghĩa là xử lý ngôn ngữ động vật. Nó được tạo ra để nhiều nhà nghiên cứu chia sẻ dữ liệu với nhau [3][4].

Dùng chung dữ liệu có một điểm tốt. Các nhà nghiên cứu ở nhiều nước gom âm thanh động vật vào một nơi. Dữ liệu càng nhiều, trí tuệ nhân tạo học càng tốt. Kho này trở thành nền tảng chung cho nhiều nhà nghiên cứu âm thanh động vật.

Một trí tuệ nhân tạo cũng có thể xử lý nhiều loài động vật cùng lúc. Vì cách xử lý âm thanh khá giống nhau giữa các loài. Cách học từ một loài được dùng cho loài khác. Vì vậy, nhiều nghiên cứu giúp đỡ lẫn nhau.

Nghiên cứu giải mã cá nhà táng cũng đang tiến lên. Một tổ chức tên là Project CETI làm việc này. Cá nhà táng phát ra nhiều tiếng click ghép lại. Cụm này được gọi là coda. Nhóm nghiên cứu đã tìm ra quy luật âm thanh bên trong coda. Quy luật này được gọi là bảng chữ cái âm thanh của cá nhà táng. Từ bảng chữ cái là cách nói ví von. Nó không phải chữ viết của con người. Nó ví quy luật các tiếng click tụ lại với bảng chữ cái [5].

Năm 2026, một sự thật mới được phát hiện. Nhóm nghiên cứu tìm thấy trong coda của cá nhà táng những mẫu giống nguyên âm trong lời nói của con người. Trong lời nói của con người có nguyên âm như a và i. Coda của cá nhà táng cũng được chia thành hai loại theo tính chất âm thanh. Độ dài của coda và thứ tự sắp xếp cũng có quy luật. Cá nhà táng cũng có khung âm thanh riêng của mình [6][7].

Nghiên cứu này sẽ tiếp tục một cách cẩn thận. Biết tiếng nói của động vật không có nghĩa là biết tất cả. Vì con người và động vật sống trong những thế giới khác nhau. Dù vậy, việc lắng nghe vẫn có ý nghĩa. Khi hiểu nhau, ta dễ sống cùng nhau hơn [8].

Tất cả các nghiên cứu này chỉ về một hướng. Đó không phải là ép động vật học tiếng người. Đó là lắng nghe âm thanh của động vật đúng như nó vốn có. Một số người nói rằng nghiên cứu này cũng có thể giúp ích cho quyền của động vật. Vì nếu biết động vật đang nghĩ gì, con người có thể bảo vệ chúng tốt hơn. Cuộc trò chuyện bắt đầu từ biển này đang trở thành cây cầu mới nối con người với tự nhiên.

Tài liệu căn cứ liên quan

- [1] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. <https://blog.google/innovation-and-ai/products/dolphingemma/>
- [2] Scientific American. (2025). DolphinGemma could enable AI communication with dolphins [Podcast]. Scientific American. <https://www.scientificamerican.com/podcast/episode/dolphingemma-could-enable-ai-communication-with-dolphins/>
- [3] Earth Species Project. (2026-04-02). Animal language processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>
- [4] Earth Species Project. (2026-07-22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. <https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>
- [5] Project CETI. (2024). Sperm whale phonetic alphabet proposed for the first time. Project CETI Blog. <https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time>

- [6] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B*, 293(2069), 20252994. <https://doi.org/10.1098/rspb.2025.2994>
- [7] National Geographic. (2026). We're one step closer to understanding the sperm whale 'alphabet'. National Geographic. <https://www.nationalgeographic.com/animals/article/sperm-whale-vowel-patterns-communication-discovery>
- [8] Boing Boing. (2026-06-11). Decoding whale language could unlock new legal rights for them. Boing Boing. <https://boingboing.net/2026/06/11/decoding-whale-language-could-unlock-new-legal-rights-for-them.html>
- [9] Schwartz, E. H. (2025-04-15). Google's new AI model could someday let you understand and talk to dolphins. TechRadar. <https://www.techradar.com/computing/artificial-intelligence/googles-new-ai-model-could-someday-let-you-understand-and-talk-to-dolphins>

Chương 2. Bảng chữ cái âm vị của cá nhà táng và Project CETI

2.1 Người giải mật mã gặp người khổng lồ dưới biển sâu

Loài vật có bộ não lớn nhất trên Trái Đất không phải là con người. Đó là cá nhà táng. Tên khoa học của cá nhà táng là *Physeter macrocephalus*. Đây là loài cá voi rất lớn sống ở biển sâu. Khi dài, cơ thể nó có thể tới 20 mét. Chiều dài ấy giống như nối hai chiếc xe buýt lớn lại với nhau. Cân nặng của nó có thể tới 50 tấn. Gần bằng cân nặng của mười con voi cộng lại [1][2].

Não của loài cá voi này nặng khoảng 8 đến 9 kilôgam. Não người nặng khoảng 1,4 kilôgam. Vì vậy, não cá nhà táng lớn gấp khoảng 6 lần não người. Bộ não lớn này được dùng cho đời sống theo nhóm và xử lý cảm xúc. Cá nhà táng dành 80 phần trăm cả đời trong vùng biển sâu và tối. Ánh sáng không tới được nơi sâu ấy. Vì vậy, cá nhà táng nhìn thế giới bằng âm thanh thay cho mắt.

Cá nhà táng phát ra âm thanh rồi nghe tiếng vang quay lại. Âm thanh va vào vật thể rồi trở về như tiếng vọng. Nó nghe tiếng vọng đó để biết vị trí và khoảng cách của vật thể. Cách này gọi là định vị bằng tiếng vang (echolocation). Nghĩa là dùng âm thanh để xem xét xung quanh. Tiếng tách dùng lúc này giúp nó tìm đường và đo con mồi. Nhưng khi nói chuyện, cá nhà táng dùng âm thanh khác. Nó gom nhiều tiếng tách ngắn theo nhịp đã định. Nhóm âm thanh này gọi là coda. coda là mảnh âm thanh mà cá voi dùng khi nói chuyện với nhau [3][5].

Trong một thời gian dài, các học giả không xem coda là quan trọng. Họ nghĩ đó chỉ là tín hiệu ngắn. Nhưng vào năm 2017, một ý tưởng mới nảy ra tại Viện Radcliffe của Đại học Harvard ở Hoa Kỳ. Tiến sĩ David Gruber, người nghiên cứu sinh vật biển, đã ghi âm tiếng cá voi. Có một người tình cờ nghe được âm thanh đó. Đó là giáo sư Shafi Goldwasser, nhà nghiên cứu mật mã của MIT ở Hoa Kỳ. Mật mã là tín hiệu có quy tắc để che giấu bí mật. Giáo sư Goldwasser cảm thấy có quy tắc trong tiếng cá voi. Nó trông giống mật mã do con người tạo ra. Vì vậy, bà đề nghị dùng máy tính để giải các quy tắc đó [4].

Ý tưởng này nối với một học giả đã theo dõi cá voi lâu năm. Đó là Tiến sĩ Shane Gero. Tiến sĩ Gero đã theo các gia đình cá voi ngoài khơi Dominica ở vùng biển Caribbean hơn 13 năm. Và vào năm 2020, Project CETI (Cetacean Translation Initiative) ra đời. Theo tiếng Hàn, tên này có nghĩa là kế hoạch dịch tiếng nói của cá voi. Các học giả từ nhiều nước đã tụ họp. Họ lập một căn cứ nghiên cứu ở vùng biển Dominica. Họ gắn micro dưới nước xuống tới độ sâu 1.000 mét. Trên thân cá voi, họ gắn thiết bị nhỏ để đo âm thanh và vị trí. Các kỹ sư của Project CETI đã làm ra thiết bị này [6][7][8].

Thiết bị nhỏ này rất đặc biệt. Bên trong nó có ba micro. Vì có nhiều micro, nó lọc được âm thanh ngay cả trong biển ồn ào. Dù nhiều cá voi cùng kêu, nó vẫn chọn riêng tiếng của một con. Thiết bị này không chỉ ghi âm. Nó còn đo cả độ sâu và chuyển động của cá voi. Nó cũng ghi nhiệt độ nước và vị trí. Vì vậy, ta có thể biết âm thanh nào được phát ra khi nào và ở đâu. Nhóm nghiên cứu dùng máy tính để phân tích các coda thu được như vậy. Số coda được phân tích lên tới 8.719. Nhóm âm thanh này trở thành chìa khóa để giải tiếng nói của cá voi [9].

Tài liệu căn cứ liên quan

[1] Whitehead, H. (2003). *Sperm Whales: Social Evolution in the Ocean*. University of Chicago Press.

[2] Project CETI. (2024). About Sperm Whales. Project CETI. <https://www.projectceti.org>

[3] Madsen, P. T., Wahlberg, M., & Møhl, B. (2002). Male sperm whale acoustics in a high-latitude habitat: implications for echolocation and communication. *Behavioral Ecology and Sociobiology*, 53(1), 31-41. <https://doi.org/10.1007/s00265-002-0548-1>

- [4] Radcliffe Institute. (2024). The CETI Moment. Harvard Radcliffe Institute. <https://www.radcliffe.harvard.edu/radcliffe-25/ceti-moment>
- [5] Andrews, R. (2024-05-07). Scientists Discover a 'Phonetic Alphabet' Used by Sperm Whales. Smithsonian Magazine. <https://www.smithsonianmag.com/smart-news/scientists-discover-a-phonetic-alphabet-used-by-sperm-whales-movin-g-one-step-closer-to-decoding-their-chatter-180984326/>
- [6] Gero, S., Whitehead, H., & Rendell, L. (2016). Individual, unit and vocal clan level identity cues in sperm whale codas. Royal Society Open Science, 3(1), 150372. <https://doi.org/10.1098/rsos.150372>
- [7] Project CETI. (2020). Cetacean Translation Initiative. Project CETI. <https://www.projectceti.org>
- [8] National Geographic Society. (2024). Project CETI. National Geographic Society. <https://www.nationalgeographic.org/explore/programs/project-ceti>
- [9] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>

2.2 Âm thanh vượt qua đèn giao thông: bốn yếu tố của coda

Sách giáo khoa trước đây nhìn âm thanh của cá nhà táng theo cách hẹp. Sách viết rằng cá voi vùng Caribbean chỉ dùng 21 loại coda. Người ta cho rằng một coda chỉ truyền một ý nghĩa. Họ xem nó giống đèn giao thông. Đèn đỏ là dừng lại, đèn xanh là đi. Nhưng khi trí tuệ nhân tạo của Project CETI xem xét âm thanh, cách nghĩ này sụp đổ [1].

Trí tuệ nhân tạo là chương trình máy tính do con người tạo ra. Nó tự tìm các quy tắc ẩn trong rất nhiều dữ liệu. Trí tuệ nhân tạo này tìm ra 156 loại coda trong cuộc trò chuyện của cá voi. Con số này nhiều hơn hơn 7 lần so với 21 loại từng biết. Hơn nữa, trong một coda cũng có nhiều mảnh nhỏ ẩn bên trong. Cá voi thay đổi bốn yếu tố để tạo ra âm thanh khác nhau.

Thứ nhất là tempo. Đó là cá voi phát tiếng tách nhanh đến mức nào. Con người cũng nói nhanh hơn khi vội. Cá voi cũng đổi tốc độ âm thanh tùy tình huống. Thứ hai là rhythm. Đó là khoảng cách giữa các tiếng tách. Tín hiệu thay đổi tùy âm thanh dày hay thưa. Thứ ba là ornamentation. Đó là thêm một tiếng tách vào cuối nhóm âm thanh. Nó làm cảm giác thay đổi nhẹ, giống dấu hỏi ở cuối câu. Thứ tư là rubato. Đó là thay đổi rất nhỏ độ dài âm thanh, giống như kéo hoặc đẩy nhịp một chút trong giờ âm nhạc. Cá voi điều chỉnh độ dài này theo dòng trò chuyện [3][4].

Bốn yếu tố này kết hợp để tạo ra 156 loại coda. Các coda đó nối tiếp theo thời gian và tạo thành tín hiệu dài hơn. Một mảnh âm thanh nhỏ không có nghĩa riêng. Nhưng khi ghép các mảnh ấy theo nhiều cách, vô số tín hiệu xuất hiện. Lời nói của con người cũng được tạo ra như vậy. Phụ âm và nguyên âm, vốn là âm thanh không có nghĩa riêng, tụ lại thành từ và câu. Nhóm nghiên cứu gọi các mảnh âm thanh nhỏ tạo nên coda là bảng chữ cái âm vị, so sánh với phụ âm và nguyên âm của con người. Trong ngôn ngữ học, đặc điểm này được gọi là tính hai tầng của mẫu hình (duality of patterning). Lâu nay con người tin rằng đặc điểm này chỉ có trong tiếng nói con người. Nhưng nay người ta phát hiện rằng tiếng nói của cá voi cũng có thể có cấu trúc tương tự [2].

Cá voi dường như cũng đổi âm thanh tùy người nghe. Âm thanh nó phát cho cá con khác với âm thanh phát cho bạn cùng nhóm. Các nhà nghiên cứu đoán rằng cá voi chọn âm thanh sau khi xét đến con đang nghe. Điều này vẫn chưa được xác nhận chắc chắn. Cá voi cũng chờ cho âm thanh của đối phương kết thúc. Khi có khoảng trống, nó mới thêm âm thanh của mình vào. Điều này giống như chờ bạn nói xong trong lớp rồi mới nói. Việc trao đổi âm thanh theo lượt như vậy gọi là luân phiên phát âm (turn-taking) [6].

Mỗi đàn cá voi dùng coda hơi khác nhau. Điều này giống như giọng nói của con người khác nhau theo vùng. Các học giả gọi những nhóm như vậy là thị tộc (clan). Thị tộc là nhóm lớn cùng chia sẻ coda. Cá voi trong cùng một thị tộc phát âm gần giống nhau. Cá voi con học coda của đàn khi lớn lên. Điều này giống

như con người học lời nói của cha mẹ. Vì vậy, coda của cá voi không phải do sinh ra đã có, mà là do học được [5].

Tài liệu căn cứ liên quan

- [1] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [2] Project CETI. (2024). Sperm Whale Phonetic Alphabet Proposed for the First Time. Project CETI Blog. <https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time>
- [3] Andrews, R. (2024-05-07). Scientists Discover a 'Phonetic Alphabet' Used by Sperm Whales. *Smithsonian Magazine*. <https://www.smithsonianmag.com/smart-news/scientists-discover-a-phonetic-alphabet-used-by-sperm-whales-movin-g-one-step-closer-to-decoding-their-chatter-180984326/>
- [4] Al Jazeera. (2024-05-08). Scientists discover sperm whale 'phonetic alphabet'. Al Jazeera. <https://www.aljazeera.com/news/2024/5/8/scientists-discover-sperm-whale-phonetic-alphabet>
- [5] Gero, S., Whitehead, H., & Rendell, L. (2016). Individual, unit and vocal clan level identity cues in sperm whale codas. *Royal Society Open Science*, 3(1), 150372. <https://doi.org/10.1098/rsos.150372>
- [6] Schulz, T. M., Whitehead, H., Gero, S., & Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Animal Behaviour*, 76(6), 1977-1988. <https://doi.org/10.1016/j.anbehav.2008.07.032>

2.3 Nguyên âm tìm thấy dưới biển sâu: phát hiện của Giáo sư Beguš

Trong một thời gian dài, nghiên cứu tiếng cá voi chỉ đo một điều. Đó là khoảng thời gian giữa các tiếng tách. Người ta chưa thể xem xét độ cao thấp của âm thanh. Có một người đã vượt qua bức tường này. Đó là Giáo sư Gašper Beguš của UC Berkeley ở Hoa Kỳ. Giáo sư Beguš chú ý đến độ cao thấp của âm thanh. Âm thanh được tạo từ nhiều độ cao thấp trộn với nhau. Sự pha trộn này gọi là phổ tần số (spectrum). Khi chiếu ánh sáng qua kính, nó trải ra thành màu cầu vồng. Âm thanh cũng được trải ra theo từng độ cao thấp như vậy [1][3].

Giáo sư Beguš quan sát lúc trí tuệ nhân tạo bắt chước và tạo ra tiếng cá voi. Trí tuệ nhân tạo không chỉ xem khoảng thời gian là quan trọng. Nó cũng coi độ cao thấp của âm thanh là quan trọng. Vì vậy, giáo sư kéo dài tiếng tách của cá voi chậm hơn hàng chục lần. Ông đổi nó thành tốc độ phù hợp với tai người. Khi đó, một điều mới xuất hiện trong tiếng tách. Đó là âm thanh giống nguyên âm của con người [2][4].

Khi con người phát nguyên âm, một số độ cao âm vang lên mạnh. Độ cao âm vang mạnh như vậy gọi là formant. Nó giống như khi thổi hơi vào chai rỗng, âm u vang lên. Giáo sư Beguš tách ra hai loại nguyên âm trong tiếng cá voi. Một loại là a-coda. Nó cho thấy cấu trúc độ cao thấp giống nguyên âm a của con người. Loại kia là i-coda. Nó cho thấy cấu trúc độ cao thấp giống nguyên âm i của con người.

Cá voi không trộn mờ hai nguyên âm. Nó phát âm a và i tách biệt rõ. Cũng có âm thanh trượt từ một nguyên âm sang nguyên âm khác. Nếu so với con người, đó là nguyên âm đôi. Đó là âm có hai tiếng nối nhau, như ai. Với tai người, nó chỉ nghe như tiếng gõ vào gỗ. Nhưng khi kéo chậm tốc độ, nó biến thành cuộc trò chuyện qua lại giữa a và i.

Đằng sau phát hiện này có một trí tuệ nhân tạo đặc biệt. Đó là trí tuệ nhân tạo mà con người có thể xem vì sao nó đưa ra câu trả lời như vậy. Khi trí tuệ nhân tạo này xử lý độ cao thấp của tiếng cá voi, một phần bên trong nó phản ứng rất mạnh với một số âm thanh nhất định. Nó giống như một công tắc chỉ bật lên trước một âm thanh cụ thể. Có một điều con người đã bỏ qua trong thời gian dài. Đó là trong tiếng cá voi có ẩn độ cao thấp giống nguyên âm. Trí tuệ nhân tạo đã nhận ra điều đó trước.

Trí tuệ nhân tạo học bằng cách tự bắt chước và tạo ra tiếng cá voi. Nó tự học dù con người không cho đáp án đúng. Cách tự học như vậy gọi là học không giám sát. Trí tuệ nhân tạo vừa tạo âm thanh vừa tự chọn phần nào là quan trọng. Điều nó chọn chính là độ cao thấp của âm thanh. Tháng 4 năm 2026,

nghiên cứu này được đăng trên tạp chí học thuật của Hội Hoàng gia Anh. Nội dung nói rằng a-coda và i-coda tuân theo quy tắc giống nguyên âm của con người.

Nguyên âm của cá voi còn có điểm giống lời nói con người. Hãy nhìn tiếng tách gấn ở cuối nhóm âm thanh. Âm thanh này trở nên giống nguyên âm đứng cạnh nó. Khi con người nói, âm trước và âm sau cũng nhuộm màu lẫn nhau. Như ta học trong giờ tiếng Hàn, từ gukmul được phát âm thành gungmul. Đó là vì âm trước và âm sau trở nên giống nhau. Điều tương tự cũng xảy ra trong tiếng cá voi.

Tài liệu căn cứ liên quan

- [1] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B*, 293(2069), 20252994. <https://doi.org/10.1098/rspb.2025.2994>
- [2] Beguš, G., Sprouse, R. L., Leban, A., Silva, M., & Gero, S. (2025). Vowel- and Diphthong-Like Spectral Patterns in Sperm Whale Codas. *Open Mind*, 9, 1849-1874. <https://doi.org/10.1162/OPMI.a.252>
- [3] University of California, Berkeley. (2024). Gašper Beguš, Speech and Computation Lab. UC Berkeley Department of Linguistics. <https://linguistics.berkeley.edu>
- [4] Zimmer, K. (2026-04-01). Sperm whale clicks follow similar rules to human speech. *Phys.org*. <https://phys.org/news/2026-04-sperm-whale-clicks-similar-human.html>

2.4 Cơ thể tạo nguyên âm trong đường mũi

Con người và cá nhà táng đã đi những con đường khác nhau. Nghĩa là chúng tiến hóa riêng trong thời gian rất dài. Vậy mà cả hai đều phát ra nguyên âm như a và i. Việc này gọi là tiến hóa hội tụ (convergent evolution). Đó là hiện tượng các sinh vật khác nhau tìm cùng một đáp án trước cùng một vấn đề. Giống như chim và dơi tiến hóa riêng, nhưng cả hai đều có cánh. Vì quy luật vật lý nơi chúng sống giống nhau, đáp án cũng trở nên giống nhau [1][2].

Con người dùng dây thanh trong cổ để phát nguyên âm. Dây thanh là các nếp gấp tạo âm thanh đầu tiên trong cổ. Con người hạ thanh quản xuống để phát nguyên âm. Thanh quản là hộp cộng hưởng ở giữa cổ, và dây thanh nằm trong đó. Con người cũng tiến hóa để môi và lưỡi cử động thật tinh tế. Nhưng cá nhà táng không phát âm bằng cách rung dây thanh trong cổ như con người. Thay vào đó, cá voi phát triển cơ quan âm thanh riêng trong đường mũi bên trong đầu. Vì nó phải phát âm trong khi chịu áp lực nước lớn ở biển sâu.

Cơ quan âm thanh của cá nhà táng gồm hai phần đi thành cặp. Thứ nhất là môi âm thanh (phonic lips). Đó là cấu trúc có nếp gấp ở bên trong đường mũi. Nó làm việc giống dây thanh của con người. Khi không khí đi qua, nó rung và tạo âm thanh đầu tiên. Điều này giống như bóp miệng quả bóng bay rồi để hơi thoát ra, tạo tiếng rít. Thứ hai là túi dầu và túi khí. Chúng là các cơ quan lấp đầy phần trán lớn của cá voi. Âm thanh đầu tiên đi qua các túi này và va chạm nhiều lần. Trong lúc đó, một số độ cao âm vang lên mạnh. Điều này giống như miệng người chỉnh lại âm thanh [3].

Hai phần này cùng tạo ra âm thanh. Nơi phát ra âm thanh đầu tiên gọi là nguồn (source). Đường đi chỉnh sửa âm thanh gọi là bộ lọc (filter). Nơi dây đàn guitar được gảy để phát âm là nguồn. Thùng đàn làm âm thanh vang đẹp là bộ lọc. Khi hai phần gặp nhau, sắc thái âm thanh được quyết định. Nguyên lý này gọi là lý thuyết nguồn-bộ lọc (source-filter theory). Cá nhà táng đã tự tạo ra nguyên lý này dù không dùng dây thanh [4].

Cơ quan âm thanh này còn có một mặt ẩn khác. Đó chính là cơ quan giúp cá voi phát nguyên âm mềm. Nhưng cùng cơ quan ấy cũng tạo ra âm thanh rất mạnh. Độ mạnh của nó được đo khoảng 230 decibel. Nó thuộc nhóm âm thanh mạnh nhất trong các âm thanh của động vật đo dưới nước. Tuy nhiên, decibel dưới nước có chuẩn khác với decibel trong không khí. Vì vậy, không thể so sánh trực tiếp hai giá trị này. Chuyện âm thanh này làm con người bị thương vẫn chưa được xác nhận. Cá voi không nhắm sức mạnh

lớn ấy vào nhau. Thay vào đó, chúng bao bọc nhau bằng giọng nói mềm. Nhờ vậy, chúng đã giữ tình bạn của biển trong thời gian rất dài [5].

Tài liệu căn cứ liên quan

- [1] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B*, 293(2069), 20252994. <https://doi.org/10.1098/rspb.2025.2994>
- [2] Madsen, P. T., Wahlberg, M., & Møhl, B. (2002). Male sperm whale acoustics in a high-latitude habitat: implications for echolocation and communication. *Behavioral Ecology and Sociobiology*, 53(1), 31-41. <https://doi.org/10.1007/s00265-002-0548-1>
- [3] Cranford, T. W., Amundin, M., & Norris, K. S. (1996). Functional morphology and homology in the odontocete nasal complex. *Journal of Morphology*, 228(3), 223-285. [https://doi.org/10.1002/\(SICI\)1097-4687\(199606\)228:3<223::AID-JMOR1>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4687(199606)228:3<223::AID-JMOR1>3.0.CO;2-3)
- [4] Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton.
- [5] Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., & Lund, A. (2003). The monopulsed nature of sperm whale clicks. *Journal of the Acoustical Society of America*, 114(2), 1143-1154. <https://doi.org/10.1121/1.1586258>

2.5 Thư viện biển 30 triệu năm và quyền của cá voi

Nghiên cứu của Project CETI và Giáo sư Beguš để lại một điều lớn để suy nghĩ. Cá nhà táng không phải là nhóm chỉ kêu bừa theo bản năng. Chúng là những nhóm được nối với nhau bằng âm thanh ở khắp biển. Chúng là những tồn tại chia sẻ văn hóa riêng và tình gắn bó gia đình [1].

Một nghiên cứu cho thấy mối gắn bó như vậy đã ra đời vào tháng 3 năm 2026. Project CETI đã ghi lại chi tiết một lần sinh con của cá nhà táng. Khi cá con chào đời, những con cá voi cái khác cũng có mặt. Chúng ở bên cạnh và giúp cá voi mẹ. Ngay cả những con cá voi không cùng họ hàng cũng góp sức. Đây là một trường hợp hiếm. Nó cho thấy bằng số liệu rằng một loài vật không phải linh trưởng cũng giúp nhau khi sinh con. Khi đó, âm thanh của cá voi cũng thay đổi. Ở mỗi khoảnh khắc quan trọng, sắc thái tiếng kêu lại đổi khác. Trong đó có cả âm thanh giống nguyên âm. Nghiên cứu này một lần nữa cho thấy cá voi là loài biết giúp nhau [2].

Cá voi ở vùng biển Caribe thường dùng một âm thanh. Các nhà khoa học gọi đó là coda 1 cộng 1 cộng 3. Nó gồm hai tiếng tách, một khoảng dừng ngắn, rồi ba tiếng nhanh. Nhóm nghiên cứu thận trọng giải nghĩa âm thanh này. Họ cho rằng nó có thể có nghĩa là tôi đang ở đây. Âm thanh này là một phương ngữ văn hóa mà đàn cá voi Caribe cùng dùng. Mỗi nhóm ở từng vùng lại dùng phương ngữ khác nhau. Trong khi đó, dòng cá nhà táng đã sống thích nghi với biển trong hàng chục triệu năm. Nghĩa là chúng đã nhận ra nhau bằng âm thanh trong biển tối suốt một thời gian rất dài [3].

Âm thanh của cá voi cũng là một mạng lưới báo nguy. Vào những năm 1800, con người đi tàu để săn cá voi. Khi đó, một số cá voi sống sót đã truyền tin. Chúng dùng âm thanh để báo cho nhau cách tránh tàu. Kết quả là tỉ lệ săn cá voi thành công giảm mạnh. Chỉ sau 2 năm rưỡi, tỉ lệ đó giảm khoảng 58 phần trăm. Đây là ghi chép cho thấy âm thanh của cá voi là một mạng lưới thông tin rộng lớn [4].

Ngôn ngữ cá voi mà trí tuệ nhân tạo làm sáng tỏ đặt ra một câu hỏi mới. Đó là con người phải đối xử với cá voi như thế nào khi cùng sống trên Trái Đất. Vì vậy, phong trào trao tư cách pháp nhân cho cá voi đang lớn dần. Tư cách pháp nhân là tư cách được pháp luật bảo vệ như con người. Điều đó có nghĩa là cá voi không bị xem là đồ vật, mà là đối tượng có quyền. Tháng 3 năm 2024, các lãnh đạo bản địa của New Zealand và các đảo quốc Thái Bình Dương đã công bố một tuyên bố. Đó là tuyên bố công nhận cá voi là pháp nhân. Tên của tuyên bố là He Whakaputanga Moana. Tuy nhiên, tuyên bố này không phải là quy tắc có hiệu lực bắt buộc như luật. Phong trào này đã hợp tác với Project CETI. Chương trình MOTH của Trường Luật Đại học New York ở Hoa Kỳ cũng cùng tham gia. MOTH là tên chỉ sự sống vượt ra ngoài con người [5][6].

Trong khi đó, tiếng ồn do con người tạo ra cản trở âm thanh của cá voi. Đó là tiếng chân vịt tàu và tiếng ồn khi khảo sát biển. Những tiếng ồn này che lấp âm thanh mà cá voi dùng để nói với nhau. Khi đó, cá voi non khó học ngôn ngữ của đàn. Văn hóa không thể truyền sang thế hệ sau. Earth Species Project cũng góp sức cho nghiên cứu này. Nơi này đã tạo ra trí tuệ nhân tạo xử lý cùng lúc âm thanh của nhiều loài vật. Tên của nó là NatureLM-audio [7][8].

Việc chúng ta cần làm rất rõ. Ta không được tùy tiện phát âm thanh nhân tạo về phía cá voi và làm lung lay văn hóa của chúng. Thay vào đó, ta phải lắng nghe những âm thanh mà chúng đã tích lũy trong biển qua thời gian dài. Đó là thư viện của biển, được nối tiếp trong hàng chục triệu năm. Trước hết, ta cần lắng lẽ lắng tai theo giai điệu ấy. Đó là con đường để con người và cá voi cùng sống bên nhau.

Tài liệu căn cứ liên quan

- [1] Whitehead, H. (2003). *Sperm Whales: Social Evolution in the Ocean*. University of Chicago Press.
- [2] EurekAlert!. (2026-03). Studies documenting rare sperm whale birth and ancient cooperative care released. EurekAlert! / Project CETI. <https://www.eurekalert.org/news-releases/1120823>
- [3] National Geographic. (2026). Whales could one day be heard in court and in their own words. National Geographic. <https://www.nationalgeographic.com/animals/article/whale-communication-legal-personhood>
- [4] Whitehead, H., Smith, T. D., & Rendell, L. (2021). Adaptation of sperm whales to open-boat whalers: rapid social learning on a large scale? *Biology Letters*, 17(3), 20210030. <https://doi.org/10.1098/rsbl.2021.0030>
- [5] Rodríguez-Garavito, C. (2026). Animal Communication, AI, and the Future of Nonhuman Rights. NYU Law MOTH Program. <https://www.law.nyu.edu/news/ideas/cesar-rodriquez-garavito-animal-communication-ai>
- [6] Te Ao Māori News. (2024-03-28). Māori king and other indigenous Pacific leaders sign up to granting whales legal personhood. Te Ao Māori News. <https://www.teaonews.co.nz/2024/03/28/maori-king-and-other-indigenous-pacific-leaders-sign-up-to-granting-whales-legal-personhood/>
- [7] Erbe, C., Marley, S. A., Schoeman, R. P., Smith, J. N., Trigg, L. E., & Embling, C. B. (2019). The Effects of Ship Noise on Marine Mammals—A Review. *Frontiers in Marine Science*, 6, 606. <https://doi.org/10.3389/fmars.2019.00606>
- [8] Earth Species Project. (2025). Using AI to Understand Animal Communication. Earth Species Project. <https://www.earthspecies.org/blog/using-ai-to-understand-animal-communication>

Chương 3. Gặp gỡ cá voi lưng gù Twain

3.1 Buổi tập của Trái Đất để hướng tới trí tuệ ngoài hành tinh

Hãy tưởng tượng một tín hiệu bay đến từ rất xa trong vũ trụ. Tín hiệu đó chỉ là tiếng nhiễu, hay là lời nói do ai đó gửi tới? Tiếng nhiễu là âm thanh lộn xộn, không có ý nghĩa. Chúng ta không có cuốn từ điển nào để giải nghĩa lời đó. Có nghĩa là ta không có chiếc chìa khóa nào để biết ý nghĩa.

Có những người đã quay mắt xuống biển, không nhìn lên trời, để trả lời câu hỏi này. Có một viện nghiên cứu tìm tín hiệu trí tuệ đến từ vũ trụ. Tên của viện là Viện Tìm kiếm Sinh vật Thông minh Ngoài Trái Đất (SETI Institute). Viện này đã hợp tác với các nhà nghiên cứu của Đại học California, cơ sở Davis, Hoa Kỳ (UC Davis). Các học giả đã nghiên cứu tiếng cá voi lâu năm cũng góp sức. Họ cùng lập một nhóm nghiên cứu tên là 'Whale-SETI' [1].

Người dẫn dắt nghiên cứu là Giáo sư Brenda McCowan. Tiến sĩ Laurance Doyle của SETI Institute cũng tham gia. Tiến sĩ Fred Sharpe, chuyên gia về cá voi lưng gù, cũng vào nhóm. Họ cùng chia sẻ một ý nghĩ. Trước khi gặp một sinh vật xa lạ trong vũ trụ, ta hãy tìm hiểu người hàng xóm trên Trái Đất trước.

Người hàng xóm đó là cá voi lưng gù (*Megaptera novaeangliae*), một loài cá voi tấm sừng hàm. Cá voi lưng gù là con vật biển lớn biết hát những bài hát dài. Chủ yếu con đực hát những bài này. Những con đực trong cùng nhóm học bài hát của nhau. Như vậy, bài hát lan ra và thay đổi từng chút một. Mỗi năm, cá voi lưng gù băng qua biển rộng và liên lạc với nhau bằng âm thanh [2].

Kế hoạch của nhóm nghiên cứu như sau. Họ thả một chiếc thuyền nhỏ ngoài khơi Alaska. Họ tắt hẳn động cơ thuyền. Họ thả micro và loa xuống nước. Rồi họ phát một âm thanh cá voi một lần. Có một cách là phát lại âm thanh đã ghi và xem con vật phản ứng ra sao. Cách này gọi là thí nghiệm phát lại (playback) [3].

Âm thanh được phát là 'tiếng gọi tiếp xúc'. Đó là âm thanh cá voi phát ra để báo vị trí của mình trong đàn. Âm thanh này gọi là 'whup call'. Đây không phải là âm thanh báo nguy. Đây là âm thanh yên hòa. Nếu chuyển sang lời người, nó được xem là gần giống câu "tôi đang ở đây".

Âm thanh này là giọng cá voi thật được ghi ngay hôm trước ở vùng biển đó. Trong đàn ấy có một cá voi lưng gù cái 38 tuổi. Tên nó là 'Twain'. Twain là con cá voi mà nhóm nghiên cứu đã theo dõi từ lâu.

Vì sao nhóm nghiên cứu chọn cá voi? Cá voi sống lâu như con người. Chúng sống cùng nhau theo đàn. Chúng liên lạc với nhau bằng âm thanh. Chúng cũng truyền bài hát đã học cho cá voi khác. Những điểm này giống con người. Vì vậy, nhóm nghiên cứu cho rằng cá voi là đối tượng phù hợp để luyện tập hiểu một trí tuệ xa lạ.

Việc giải tín hiệu vũ trụ có một khó khăn lớn. Bên gửi tín hiệu và bên nhận tín hiệu không có quy ước trước. Nếu có một cuốn sách ghi song song ngôn ngữ của hai nước, việc dịch sẽ dễ. Nhưng tín hiệu ngoài hành tinh không có cuốn sách như vậy. Vì thế, ta phải tìm ra quy tắc từ đầu.

Trên thuyền trước thí nghiệm rất yên tĩnh. Nhóm nghiên cứu nín thở và nhìn mặt nước. Không ai biết Twain có đáp lại hay không.

Tài liệu căn cứ liên quan

[1] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. <https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/>

[2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. *PeerJ*, 11, e16349. <https://doi.org/10.7717/peerj.16349>

[3] Ralls, E. (2026-04-29). AI helps humans have a 20-minute 'conversation' with a humpback whale named Twain. Earth.com. <https://www.earth.com/animals/whale-seti-ai-helps-humans-have-20-minute-conversation-with-humpback-whale-twain/>

3.2 20 phút với Twain

Âm thanh đầu tiên vang ra từ loa. Không lâu sau, Twain xuất hiện. Twain lặng lẽ tiến lại gần thuyền. Nó bơi một vòng lớn quanh thuyền. Rồi nó phát ra âm thanh của mình về phía loa.

Twain không phát âm thanh một lần rồi biến mất. Khi con người phát âm thanh, Twain đáp lại bằng âm thanh. Khi con người dừng, Twain cũng chờ. Khi âm thanh lại vang lên, nó lại đáp. Việc trao đổi này kéo dài 20 phút. 20 phút gần bằng độ dài của một phim hoạt hình ngắn.

Trong 20 phút, nhóm nghiên cứu phát whup call 38 lần. Twain phát whup call 36 lần. Trong đó, 33 lần là câu đáp khớp ngay với âm thanh con người phát. 3 lần còn lại là âm thanh phát ra sau khi con người đã dừng. Đây là lần đầu tiên con người và cá voi hoang dã trao đổi âm thanh lâu như vậy.

Twain còn khớp cả khoảng cách giữa các âm thanh. Khi con người để khoảng cách dài giữa các âm, Twain cũng để dài. Khi khoảng cách ngắn, Twain cũng đáp ngắn. Nó đã làm theo cách của đối phương.

Tiến sĩ McCowan, người dẫn dắt nghiên cứu, nói như sau. Bà nói việc này có vẻ là một cuộc trao đổi âm thanh có ý nghĩa giữa người và cá voi. Tuy vậy, con người không dịch được ý nghĩa trong lời cá voi. Nhóm nghiên cứu hiểu đây là việc hai bên chia sẻ âm thanh theo lượt.

Twain không chỉ phát âm thanh vui vẻ trong suốt 20 phút. Lúc đầu, nó tham gia tốt vào việc trao đổi âm thanh. Nhưng khi nhóm nghiên cứu cứ phát đi phát lại cùng một whup call, phản ứng của nó thay đổi. Twain trông ngày càng bồn chồn. Nó thở thường xuyên và cũng đập mặt nước. Rồi cuối cùng nó rời khỏi nơi đó. Nó đã cho thấy ba giai đoạn: tham gia, bồn chồn, rời đi.

Cuộc gặp này có giới hạn rõ ràng. Đối tượng thí nghiệm chỉ là một con Twain. Thí nghiệm cũng chỉ là một cơ hội duy nhất tình cờ xuất hiện. Vì vậy, không thể chỉ dựa vào kết quả này mà khẳng định cá voi trò chuyện. Nhóm nghiên cứu cũng nói điểm này một cách thận trọng.

Dù vậy, cuộc gặp này cho thấy sức mạnh của thí nghiệm phát lại. Đó là cách phát âm thanh đã ghi và quan sát câu đáp. Bằng cách này, ta có thể nhìn gần phản ứng của động vật hoang dã. Nhóm nghiên cứu gọi đây là một công cụ nghiên cứu mới.

Tài liệu căn cứ liên quan

- [1] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. <https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/>
- [2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. *PeerJ*, 11, e16349. <https://doi.org/10.7717/peerj.16349>
- [3] Frediani, J. (2024). "Conversing" with a humpback whale named Twain. The Safina Center. <https://www.safinacenter.org/blog/conversing-with-a-humpback-whale-named-twain>

3.3 Cuộc trò chuyện biết giữ lượt

Khi nói chuyện với nhau, con người giữ quy tắc. Ta chờ cho đến khi người kia nói xong. Khi người kia dừng, lúc đó ta bắt đầu nói. Quy tắc nói luân phiên như vậy gọi là luân phiên phát âm (turn-taking). Nó giống cảnh xếp hàng nhận bữa ăn và giữ đúng lượt.

Có một khoảnh khắc người kia nói xong và im lặng một chút. Khoảng trống ngắn đó gọi là khoảng nghỉ hội thoại (conversational pause). Cần có khoảng nghỉ này để lời nói của hai bên không chồng lên nhau. Cá voi lưng gù trong tự nhiên cũng giữ quy tắc này.

Trong 20 phút trao đổi, Twain không phá quy tắc này. Nó không chen vào khi âm thanh của con người đang phát. Nó chờ đến khi âm thanh kết thúc hẳn. Nó chọn đúng khoảnh khắc trở nên yên lặng. Rồi nó đưa âm thanh của mình vào chỗ trống đó.

Nhóm nghiên cứu đã đo thời gian Twain đáp lại. Họ xem sau khi âm thanh kết thúc mấy giây thì câu đáp xuất hiện. Việc này không dùng trí tuệ nhân tạo. Họ dùng âm thanh đã ghi và thống kê để xem xét từng mốc thời gian. Thời gian đó khớp tốt với quy tắc trò chuyện của con người.

Kết quả này cho biết một điều. Twain không kêu lên vì khó chịu với âm thanh máy móc. Twain phân biệt lượt của đối phương và lượt của mình rồi mới phát âm thanh. Nhóm nghiên cứu cho rằng Twain đã tham gia vào việc trao đổi âm thanh.

Con người học quy tắc này từ khi còn là em bé. Em bé nhìn vào mắt mẹ và trao đổi âm thanh. Trước khi học nói, em bé đã học thứ tự này trước. Đây là quy tắc làm nền cho đối thoại.

Khoảng nghỉ hội thoại là một thời gian rất ngắn. Đôi khi nó còn ngắn hơn một cái chớp mắt. Twain không bỏ lỡ khoảng trống ngắn này. Nó đưa âm thanh vào đúng khoảnh khắc.

Sự chính xác như vậy rất khó nếu không luyện tập. Twain vẫn luôn trao đổi âm thanh như thế trong tự nhiên. Nó dùng nguyên cách quen thuộc ấy với âm thanh từ phía con người. Dù đối phương không phải cá voi thật, nó vẫn giữ quy tắc.

Tài liệu căn cứ liên quan

- [1] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. *PeerJ*, 11, e16349. <https://doi.org/10.7717/peerj.16349>
- [2] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. <https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/>
- [3] Ralls, E. (2024-01-06). Scientists have a 20-minute 'conversation' with a humpback whale named Twain. *Earth.com*. <https://www.earth.com/news/scientists-have-20-minute-conversation-with-a-humpback-whale-named-twain/>

3.4 Ngữ pháp cá voi được toán học làm sáng tỏ

Tiến sĩ Doyle đã thử giải bài hát cá voi bằng các con số. Vì ông nghĩ bài hát cá voi không chỉ là âm thanh vô nghĩa. Vì vậy, ông dùng một công cụ gọi là lý thuyết thông tin (information theory). Lý thuyết thông tin là cách đo bằng số xem trong âm thanh hay chữ viết có bao nhiêu quy tắc.

Nhóm nghiên cứu đã thử áp dụng định luật Zipf (Zipf's law) vào bài hát cá voi. Định luật Zipf cho biết một quy tắc ẩn trong lời nói. Trong lời nói của con người, số lượng từ dùng thường xuyên và từ ít dùng tạo thành một tỉ lệ nhất định. Ví dụ, trong tiếng Hàn, các từ '은, 는, 이, 가' xuất hiện rất thường xuyên. Từ khó thì thỉnh thoảng mới xuất hiện một lần. Giữa nhiều và ít luôn có một tỉ lệ gần giống nhau.

Năm 2025, có một nghiên cứu tìm thấy quy tắc này trong bài hát cá voi. Nhóm nghiên cứu của Inbal Arnon đã quan sát lâu dài bài hát cá voi lưng gù. Đó là những bài hát được thu trong 8 năm ở Nam Thái Bình Dương. Những bài hát ấy khớp tốt với định luật Zipf. Đây là lần đầu tiên quy tắc này được tìm thấy rõ ràng ngoài lời nói của con người.

Kết quả này có nghĩa gì? Nó có thể là dấu hiệu cho thấy trong bài hát cá voi có quy tắc. Nhưng chỉ riêng điều này chưa chứng minh được bài hát cá voi có ý nghĩa hay ngữ pháp. Nhóm nghiên cứu cũng không nói bài hát cá voi giống lời nói của con người. Họ cho rằng nó có thể gần với âm nhạc của con người hơn.

Nhóm nghiên cứu cũng xem xét entropy Shannon (Shannon entropy). Entropy Shannon là cách đo bằng số xem khó đoán âm thanh tiếp theo đến mức nào. Nếu âm thanh hoàn toàn lộn xộn, giá trị này cao. Nếu chỉ lặp lại cùng một âm thanh, giá trị này gần bằng 0. Bài hát cá voi nằm giữa hai cực đó. Nó giữ quy tắc, nhưng không phải lúc nào cũng giống hệt nhau.

Cấu trúc như vậy cũng xuất hiện trong nghiên cứu về loài cá voi khác. Ở đây, ta tạm lấy câu chuyện về loài cá voi khác làm ví dụ. Có một nhóm nghiên cứu tên là Project CETI. Họ quan sát cá nhà táng. Cá nhà táng là loài khác với cá voi lưng gù. Cá nhà táng phát ra các tiếng lách cách gom thành cụm ngắn. Cụm âm thanh này gọi là coda.

Năm 2024, nhóm nghiên cứu Project CETI đã làm rõ cấu trúc của coda. Cá nhà táng tạo âm thanh bằng cách trộn bốn yếu tố: nhịp, tốc độ, rubato và âm trang trí. Rubato là việc kéo dài hoặc rút ngắn nhịp một chút. Âm trang trí là việc gắn thêm một tiếng lách cách vào âm thanh. Chúng kết hợp bốn yếu tố này theo nhiều cách để tạo ra rất nhiều âm thanh. Người ta từng nghĩ chỉ con người mới dùng kiểu kết hợp này, nhưng cá nhà táng cũng làm như vậy [4].

Tài liệu căn cứ liên quan

- [1] Doyle, L. R., & McCowan, B. (2011). Information theory, animal communication, and the search for extraterrestrial intelligence. *Acta Astronautica*, 68(3-4), 406-417.
<https://www.sciencedirect.com/science/article/abs/pii/S0094576509005682>
- [2] Arnon, I., et al. (2025). Whale song shows language-like statistical structure. *Science*, 387.
<https://doi.org/10.1126/science.adq7055>
- [3] Sharma, P., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [4] Project CETI. (2024-05-07). Sperm whale phonetic alphabet proposed for the first time. Project CETI.
<https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time>

3.5 Lá thư gửi vào vũ trụ

Nghiên cứu Whale-SETI có một mục tiêu cuối cùng. Nó không dừng ở việc tìm hiểu cá voi. Mục tiêu là tạo ra công cụ để một ngày nào đó nhận ra tín hiệu xa lạ đến từ vũ trụ. Vì ta cần một thước đo để phân biệt tiếng nhiễu và tín hiệu có ý nghĩa.

Có những phương pháp giải bài hát cá voi bằng con số. Đó là các phương pháp như định luật Zipf hay entropy Shannon. Sau này, ta có thể thử dùng các phương pháp này với tín hiệu vũ trụ. Tuy vậy, chúng chưa được dùng trực tiếp cho tín hiệu vũ trụ. Hiện nay, các nhà nghiên cứu đang xem chúng là những phương pháp đáng thử trong tương lai.

Nhiều học giả đã chỉ ra sự kiêu ngạo của con người. Đó là việc ta thậm chí không hiểu âm thanh của cá voi sống cùng Trái Đất. Họ hỏi rằng nếu vậy, ta sẽ hiểu tín hiệu từ phía bên kia vũ trụ bằng cách nào. Câu hỏi này làm con người khiêm tốn hơn.

Ngày xưa, con người tin rằng Trái Đất là trung tâm của vũ trụ. Nhưng niềm tin ấy đã được chứng minh là sai. Niềm tin rằng chỉ con người mới chia sẻ ý nghĩa cũng có thể giống như vậy. Có thể dưới biển cũng có những người hàng xóm đang chia sẻ âm thanh.

Trong thời gian dài, con người chỉ xem bài hát cá voi là tiếng nhiễu. Con người cũng đã săn bắt rất nhiều cá voi. Tàu thuyền và máy móc làm biển trở nên ồn ào. Dù vậy, cá voi không ngừng trao đổi âm thanh. Từ thời rất xa xưa, chúng đã gửi âm thanh cho nhau.

Vì vậy, 20 phút với Twain là một buổi tập. Đó là nơi ta tập trước phương pháp sẽ dùng khi gặp một trí tuệ xa lạ. Trước khi gửi lời vào vũ trụ, ta cần lắng nghe âm thanh bên cạnh. Khả năng lắng nghe tốt là điểm bắt đầu của phiên dịch.

Tài liệu căn cứ liên quan

- [1] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. <https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/>
- [2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. *PeerJ*, 11, e16349. <https://doi.org/10.7717/peerj.16349>
- [3] Ralls, E. (2024-01-06). Scientists have a 20-minute 'conversation' with a humpback whale named Twain. *Earth.com*. <https://www.earth.com/news/scientists-have-20-minute-conversation-with-a-humpback-whale-named-twain/>

3.6 Cánh cửa mới do trí tuệ nhân tạo mở ra

Sau cuộc gặp với Twain, nghiên cứu đang phát triển nhanh. Vì trí tuệ nhân tạo giúp ta quan sát âm thanh cá voi sâu hơn. Những việc một mình con người phải mất hàng chục năm mới làm được, trí tuệ nhân tạo đang giúp làm nhanh hơn.

Có một tổ chức tên là Earth Species Project. Họ tạo ra trí tuệ nhân tạo để hiểu âm thanh động vật. Tên của trí tuệ nhân tạo đó là NatureLM-audio. Công cụ này chỉ nghe âm thanh rồi phân biệt đó là con vật nào. Nó cũng phân loại các loại âm thanh. Những việc này thì nó đã làm tốt.

Trí tuệ nhân tạo này cũng có năng lực đang ở giai đoạn thử nghiệm. Đó là đếm xem trong bản ghi có mấy con vật. Nó cũng có việc nhận ra âm thanh của loài vật chưa từng học. Những việc này vẫn chưa hoàn hảo. Nó chỉ đoán đúng hơn một chút so với đoán bừa.

Trong trí tuệ nhân tạo có loại trí tuệ nhân tạo lớn làm nền tảng. Loại trí tuệ nhân tạo lớn này gọi là mô hình nền tảng (foundation model). Đó là trí tuệ nhân tạo dùng một mô hình cho nhiều việc khác nhau. Perch do Google tạo ra là một ví dụ như vậy.

Câu chuyện trí tuệ nhân tạo giúp nghiên cứu cá voi có thể chia theo thứ tự thời gian. Trước hết là chuyện năm 2018. Google và Cơ quan Khí quyển và Đại dương Quốc gia Hoa Kỳ (NOAA) đã hợp tác. Cơ quan Khí quyển và Đại dương Quốc gia là một cơ quan của Hoa Kỳ nghiên cứu biển và thời tiết. Hai bên cùng xem xét âm thanh dưới biển được thu trong 14 năm. Lượng dữ liệu đó hơn 187,000 giờ. Nếu con người nghe không nghỉ, cũng mất khoảng 21 năm. Trí tuệ nhân tạo đã tìm ra bài hát của cá voi lưng gù trong khối dữ liệu lớn này.

Tiếp theo là chuyện năm 2025. Google đã chỉnh sửa Perch và đưa ra Perch 2.0. Perch ban đầu là trí tuệ nhân tạo học từ tiếng chim. Nhưng nó cũng rất hợp với âm thanh dưới biển. Nó đạt kết quả tốt trong việc nhận ra âm thanh của cá voi. Trí tuệ nhân tạo học về chim đã giúp ích cho nghiên cứu cá voi.

Sức mạnh lớn của trí tuệ nhân tạo nằm ở tốc độ. Âm thanh mà con người phải nghe trong nhiều tháng, trí tuệ nhân tạo có thể rà trong vài giờ. Nhờ vậy, nghiên cứu đã nhanh hơn. Nó cũng tìm được những âm thanh nhỏ mà con người dễ bỏ lỡ. Nhờ đó, những sự thật mới xuất hiện trong dữ liệu cũ.

Ở đây có một điểm cần cẩn thận. Trí tuệ nhân tạo nhận ra âm thanh của cá voi không có nghĩa là đã giải mã hết tiếng nói của cá voi. Nó đã tìm thấy quy luật trong âm thanh, nhưng chưa biết hết ý nghĩa của chúng. Hiện vẫn chưa có chứng cứ cho thấy ý nghĩa của tiếng cá voi đã được giải thích đầy đủ. Cuốn từ điển giữa con người và cá voi mới chỉ vừa lật qua trang đầu.

Earth Species Project không chỉ nghiên cứu cá voi. Dự án này cũng nghiên cứu tiếng quạ. Dự án này cũng nghiên cứu tiếng cá voi trắng. Dự án xử lý âm thanh của

nhiều loài động vật bằng một công cụ. Trong tương lai, họ còn có kế hoạch phát âm thanh do con người tạo ra cho động vật nghe. Mục đích là xem động vật đáp lại âm thanh đó như thế nào. Việc trao đổi âm thanh với Twain là bước đầu tiên.

Gần đây cũng có nghiên cứu làm rõ cấu trúc bài hát của cá voi. Đó là nhóm nghiên cứu của Julia Zeh tại Syracuse University ở Hoa Kỳ năm 2026. Họ gắn một thiết bị ghi nhỏ lên cơ thể cá voi lưng gù được đang hát. Đó là thiết bị vừa ghi âm vừa ghi chuyển động, được gắn tạm thời bằng giác hút. Nghiên cứu này có chủ đề khác với việc trao đổi âm thanh của Twain. Nội dung là chuyển động lên xuống trong nước của cá voi có liên quan với bài hát. Bài hát khi lặn xuống sâu khác với bài hát khi bơi lên. Điều đó có nghĩa là bài hát được sắp xếp theo chuyển động của cơ thể [5].

Vì vậy, nghiên cứu vẫn đang tiếp tục. Con người đang điền từng trang vào cuốn từ điển giữa người và cá voi. 20 phút của Twain đã mở ra câu đầu tiên của cuốn từ điển đó.

Tài liệu căn cứ liên quan

- [1] Earth Species Project. (2024-11-11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. <https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics>
- [2] Robinson, D., Miron, M., Hagiwara, M., Weck, B., Keen, S., Alizadeh, M., Narula, G., Geist, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. <https://arxiv.org/abs/2411.07186>
- [3] Google Research. (2018-10). Acoustic detection of humpback whales using a convolutional neural network. Google Research Blog. <https://research.google/blog/acoustic-detection-of-humpback-whales-using-a-convolutional-neural-network/>
- [4] Google Research. (2026-02-09). How AI trained on birds is surfacing underwater mysteries. Google Research Blog. <https://research.google/blog/how-ai-trained-on-birds-is-surfacing-underwater-mysteries/>
- [5] Syracuse University. (2026-08-31). How a whale's dive shapes its song. Syracuse University News. <https://news.syr.edu/2026/08/31/how-a-whales-dive-shapes-its-song/>

Chương 4. Dàn nhạc của biển sâu

4.1 Tách giọng nói ra khỏi biển ồn ào

Biển không phải là nơi yên lặng. Dưới nước là một không gian rộng đầy âm thanh. Tiếng hát của cá voi, tiếng huýt sáo của cá heo và tiếng sóng luôn trộn vào nhau. Dưới nước, âm thanh lan nhanh hơn nhiều so với trong không khí. Nó nhanh hơn khoảng 4,5 lần. Vì vậy cá voi đọc thể giới bằng tai nhiều hơn bằng mắt. Khi tìm nhau ngoài biển xa, chúng dùng âm thanh. Khi đuổi theo mồi và tránh nguy hiểm, chúng cũng dùng âm thanh [1][2].

Dưới nước, âm thanh đi khoảng 1.500 mét trong 1 giây. Khoảng cách này xa hơn nhiều so với âm thanh trên đất liền. Vì vậy cá voi có thể nghe cả tiếng của bạn ở rất xa.

Vấn đề là tiếng ồn do con người tạo ra. Tiếng chân vịt của tàu lớn làm rung động dưới nước. Tiếng máy dò bằng sóng âm dưới biển, gọi là sonar, cũng rất lớn. Còn có tiếng nổ để tìm tài nguyên dưới đáy biển. Những tiếng ồn này lan xa hàng nghìn kilômét. Ngày nay, tiếng tàu là tiếng ồn của con người lan rộng nhất trong biển. Khi tiếng ồn quá lớn, tiếng cá voi bị vùi trong đó. Các nhà khoa học gọi hiện tượng này là âm thanh bị che lấp. Với cá voi sống nhờ âm thanh, điều này là một mối đe dọa lớn.

Tiếng ồn cũng làm cơ thể cá voi bị đau. Năm 2001, sau một sự cố lớn ở Hoa Kỳ, số tàu tạm thời giảm mạnh. Khi ấy biển trở nên yên hơn. Nhóm của Rosalind Rolland đã quan sát cơ hội này. Họ thu phân của cá voi trơn ở vịnh Fundy, Canada để nghiên cứu. Khi biển yên hơn, chất gây căng thẳng trong cơ thể cá voi giảm xuống. Điều này giống như tim con người đập nhanh khi ở cạnh công trường ồn ào. Đó là bằng chứng đầu tiên cho thấy tiếng ồn làm cá voi kiệt sức [3].

Rất khó tách riêng giọng của một con cá voi ở nơi nhiều âm thanh lẫn vào nhau. Các nhà khoa học gọi khó khăn này là “vấn đề bữa tiệc cocktail”. Nó giống việc nghe riêng lời của một người bạn trong căn phòng đông người. Cũng giống như dù nhà ăn ồn ào, ta vẫn nghe được bạn ngồi cạnh nói. Máy móc trước đây không làm tốt việc này. Khi tiếng kim loại của tàu chòng lên tiếng cá voi, máy thường nhầm lẫn.

Trí tuệ nhân tạo đã giải bài toán này theo cách mới. Đó là công nghệ tách riêng âm thanh ta muốn khỏi các âm thanh lẫn lộn. Trước hết, máy đổi âm thanh thành một bản đồ số. Trên bản đồ này, tiếng tàu và tiếng cá voi nằm ở những vị trí khác nhau. Giống như khi các viên bi nhiều màu bị trộn lẫn, ta lại xếp chúng theo từng màu. Máy chia các vị trí đó ra và gom lại chỉ tiếng cá voi. Nhóm nghiên cứu Earth Species Project đã tạo ra cách xóa tiếng ồn ngay cả khi không có âm thanh đáp án sạch. Nhờ vậy, con người có thể nghe lại giọng cá voi rõ hơn [4].

Tài liệu căn cứ liên quan

- [1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.
- [2] NOAA Fisheries. (2025). A whale's world of sound. National Oceanic and Atmospheric Administration. <https://www.fisheries.noaa.gov/feature-story/whales-world-sound>
- [3] Rolland, R. M., Parks, S. E., Hunt, K. E., et al. (2012). Evidence that ship noise increases stress in right whales. Proceedings of the Royal Society B, 279(1737), 2363-2368. <https://doi.org/10.1098/rspb.2011.2429>
- [4] Miron, M., Keen, S., Hoffman, B., Hagiwara, M., et al. (2024). Biodenoising: Animal vocalization denoising without access to clean data. Earth Species Project. arXiv:2410.03427. <https://arxiv.org/abs/2410.03427>

4.2 Phương ngữ của gia đình cá voi sát thủ

Khi tiếng ồn được gạt bỏ, các nhà khoa học bắt đầu đọc những bí mật trong giọng cá voi. Bí mật đó là “phương ngữ” khác nhau của từng gia đình. Ở bờ biển phía tây Canada có một trạm nghiên cứu cá voi sát thủ. Tên của nó là OrcaLab. Tại đây, Helena Symonds và tiến sĩ Paul Spong đã ghi âm tiếng cá voi sát thủ hoang dã trong nhiều chục năm. Họ không bắt giữ hay quấy rầy cá voi. Họ chỉ lặng lẽ thu âm bằng micro dưới nước [1].

Cá voi sát thủ là loài sống quanh mẹ. Con non không rời đàn của mẹ suốt đời. Vì vậy mỗi gia đình có âm thanh riêng được truyền qua nhiều đời. Các nhà khoa học gọi đó là phương ngữ mà gia đình bên mẹ cùng dùng. Nó giống như cách nói của lớp chúng ta và lớp bên cạnh hơi khác nhau. Nhờ phương ngữ này, cá voi sát thủ nhận ra gia đình mình giữa biển rộng. Nó cũng giúp ngăn cá voi họ hàng gần giao phối với nhau.

OrcaLab đã thu âm trong một thời gian rất dài. Riêng số bản ghi đã gần 20.000 giờ. Con người khó nghe hết từng ấy âm thanh. Vì vậy các nhà khoa học giao việc này cho trí tuệ nhân tạo. Chương trình do các nhà nghiên cứu Đức và Canada tạo ra có tên là ORCA-SPOT. Chương trình này tự động tìm riêng tiếng cá voi sát thủ trong các bản ghi dài. Khi so với đáp án do con người gán nhãn, nó đúng 93 lần trong 100 lần. Việc có thể mất vài năm, máy hoàn thành chỉ trong tám ngày [5].

Khi trí tuệ nhân tạo chia âm thanh, nó chú ý mạnh vào phần quan trọng. Điều này giống như con người gạch dưới từ quan trọng trong một bài dài. Nhờ vậy máy nhận ra cá khác biệt rất nhỏ trong ngữ điệu. Cá voi sát thủ con học phương ngữ bằng cách nghe tiếng của mẹ và đàn. Điều này giống trẻ em học nói theo cha mẹ. Các nhà khoa học cũng tìm thấy điều mới khi xem bản đồ âm thanh do trí tuệ nhân tạo vẽ. Đó là âm thanh của một gia đình đã thay đổi từng chút qua các thế hệ. Nó giống như lời nói được dùng lâu cũng đổi dần theo thời đại. Điều này có nghĩa tiếng cá voi không chỉ là bản năng bẩm sinh. Đó là bằng chứng cho thấy nó là văn hóa được học, được truyền lại và thay đổi như tiếng nói của con người.

Biết phương ngữ cũng giúp bảo vệ cá voi. Vì ta có thể biết gia đình nào sống ở đâu nhờ âm thanh. Ta có thể xem gia đình nào đang ở vùng biển có nhiều tàu qua lại. Nếu đếm gia đình bằng âm thanh, ta cũng sớm biết số lượng có giảm hay không. Những nghiên cứu như vậy vẫn tiếp tục. Năm 2026, một chương trình mới tên là orcaAI ra đời. Nó tự động tìm và phân loại tiếng cá voi sát thủ. Chương trình này học tiếng cá voi sát thủ ở vùng biển Iceland, không phải Canada. Cá voi sát thủ ở đó sống bằng cách săn cá trích. Dù con người không nghe trong thời gian dài, máy vẫn thay họ kiểm tra các bản ghi dài [6].

Tài liệu căn cứ liên quan

- [1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.
- [5] Bergler, C., Schröter, H., Cheng, R. X., et al. (2019). ORCA-SPOT: An automatic killer whale sound detection toolkit using deep learning. Scientific Reports, 9, 10997. <https://doi.org/10.1038/s41598-019-47335-w>
- [6] Bonhoeffer, S., Selbmann, A., et al. (2026). orcaAI: A machine learning tool to detect and classify acoustic signals of killer whales in audio recordings. Marine Mammal Science, 42(1), e70083. <https://doi.org/10.1111/mms.70083>

4.3 Khôi phục tiếng cá voi trắng dưới lớp băng

Nếu cá voi sát thủ nuôi giữ phương ngữ ở vùng ven biển trong, thì cũng có loài cá voi sống trong nước lạnh và đục. Đó là cá voi trắng Beluga Whale sống ở Bắc Cực và sông St. Lawrence River của Canada. Vì thân màu trắng, nó được gọi là thiên thần của Bắc Cực. Vì giỏi phát ra âm thanh cao và trong như chim, nó còn có biệt danh là chim hoàng yến của biển. Cá voi trắng ở sông St. Lawrence River hiện đang giảm mạnh. Nguyên nhân là tiếng ồn nhà máy và ô nhiễm nước tích tụ. Dù đã lập khu bảo vệ, số lượng vẫn không tăng tốt, nên nhiều người lo lắng [1][7].

Nhà nghiên cứu Jaclyn Aubin của University of Windsor ở Canada đã tham gia việc này. Bà quyết định giải mã tiếng cá voi trắng. Việc này không dễ. Cá voi trắng là loài sống theo đàn và rất thích giao tiếp. Hàng chục con tụ lại và phát ra âm thanh. Chúng phát cùng lúc tiếng huýt sáo và tiếng lách cách nhanh. Âm thanh lẫn vào nhau, nên con người khó chia đâu là điểm bắt đầu và đâu là điểm kết thúc.

Nhóm nghiên cứu giao âm thanh này cho trí tuệ nhân tạo. Trước hết, một máy tìm âm thanh đã cắt riêng tiếng cá voi trắng khỏi bản ghi dài. Sau đó, một máy khác chia các tiếng đó theo từng loại. Lúc này họ cũng dùng chương trình AVES do Earth Species Project tạo ra. AVES là mô hình nền học âm thanh động vật. Mô hình nền là trí tuệ nhân tạo học sẵn rất nhiều dữ liệu và được dùng cho nhiều việc. AVES tự học cấu trúc âm thanh mà không cần con người cho đáp án [8].

Khi so tiếng do trí tuệ nhân tạo chia với tiếng do con người chia, hai kết quả gần như giống nhau. Nghĩa là máy có thể phân loại tiếng cá voi trắng giỏi như chuyên gia mà không cần con người can thiệp. Cách này có lợi. Máy kiểm tra nhanh những bản ghi dài mà con người phải nghe trong nhiều ngày. Ta có thể ghi lại đều đặn khi nào và ở đâu cá voi trắng trong sông cất tiếng. Những ghi chép này trở thành nền tảng để bảo vệ cá voi trắng đang giảm số lượng.

Tiếng băng vỡ và va vào nhau là trở ngại lớn. Nhóm nghiên cứu đã dùng thêm trí tuệ nhân tạo chuyên khử tiếng ồn để xóa tiếng ồn này. Máy này khôi phục riêng giọng cá voi trắng trong các âm thanh ồn ào. Nhờ vậy, ta có thể nghe lại tiếng cá voi trắng suýt bị mất. Như thế, trí tuệ nhân tạo đã bảo vệ giọng nói bị vùi dưới lớp băng.

Tài liệu căn cứ liên quan

- [1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.
- [7] Cotillard, T., Aubin, J., Vergara, V., et al. (2024). Automatic detection and classification of beluga whale calls in the St. Lawrence estuary. The Journal of the Acoustical Society of America, 156(6), 3723-3740.
<https://doi.org/10.1121/10.0030472>
- [8] Hagiwara, M. (2023). AVES: Animal vocalization encoder based on self-supervision. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing.
<https://doi.org/10.1109/ICASSP49357.2023.10095642>

4.4 AI học từ tiếng chim lắng nghe biển

Vì sao đang nói chuyện biển lại đột nhiên nói đến chim? Có lý do. Vì trí tuệ nhân tạo học từ tiếng chim đã giúp ích lớn trong việc giải mã tiếng động vật biển. Khởi đầu đến từ Google. Đó là dự án “Perch” của nhóm nghiên cứu do nhà khoa học máy tính tiến sĩ Tom Denton dẫn đầu. Ước mơ của nhóm này rất lớn. Họ muốn sắp xếp mọi âm thanh tự nhiên trên Trái Đất thành một hệ thống [10].

Nhóm Perch dạy trí tuệ nhân tạo bằng tiếng chim. Họ cho máy học tiếng hót của hơn 10.000 loài chim trên thế giới. Tổng thời gian học lên tới hàng trăm nghìn giờ. Trí tuệ nhân tạo được tạo ra như vậy đối âm thanh thành tọa độ số để sắp xếp. Có thể hiểu là nó vẽ một bản đồ rộng, trong đó mỗi âm thanh có một vị trí.

Nhưng một điều bất ngờ đã xảy ra. Trí tuệ nhân tạo học tiếng chim cũng phân loại tốt tiếng động vật dưới nước. Quy tắc học từ một loài lại dùng được cho tiếng của loài khác. Chim kêu trong không khí, còn cá voi kêu trong nước. Nơi sống và âm thanh đều khác nhau. Vậy mà bản đồ học từ tiếng chim vẫn phù hợp với tiếng cá voi.

Nhóm nghiên cứu xác nhận sức mạnh này bằng tiếng ếch. Dữ liệu được dùng khi ấy là AnuraSet. Đây là dữ liệu tập hợp tiếng của 42 loài ếch Brazil. Nó được ghi âm không phải ở Amazon, mà ở các khu rừng gọi là Cerrado và Atlantic Forest. Dữ liệu này do nhóm nghiên cứu Cañas tạo ra, không phải Google.

Google mượn dữ liệu công khai này để thử trí tuệ nhân tạo. Dù chỉ cho xem vài tiếng ếch, trí tuệ nhân tạo vẫn nhận đúng hơn 95%. Đó là nhờ cách học việc mới chỉ từ vài ví dụ. Sức mạnh này không dừng ở ếch. Nó cũng dùng được với tiếng lách cách của cá voi và tiếng huýt sáo của cá heo [9].

Nghiên cứu này tiếp tục phát triển. Tháng 8 năm 2025, Google công bố phiên bản mới của Perch, gọi là Perch 2.0. Trí tuệ nhân tạo này học từ tiếng của nhiều loài động vật trên cạn, không chỉ chim. Số loài đã học vượt quá 14.000. Nó học tiếng chim, động vật có vú, ếch và cả côn trùng. Nhưng nó hầu như chưa học âm thanh dưới biển. Dù vậy, tháng 2 năm 2026, Google cho biết trí tuệ nhân tạo này đạt kết quả tốt trong nghiên cứu tiếng cá voi. Trong nhiều bài thử về tiếng cá voi, trí tuệ nhân tạo này đứng nhất hoặc đứng thứ hai. Google cũng tạo ra DolphinGemma, trí tuệ nhân tạo dành cho cá heo. Chương trình này nhỏ nên có thể chạy trong điện thoại thông minh. Nó học từ âm thanh mà Wild Dolphin Project đã thu thập trong thời gian dài. Nó chưa giải được nghĩa lời nói của cá heo. Nó là công cụ nghiên cứu để tìm cấu trúc và quy tắc của âm thanh [11].

DolphinGemma không chỉ phân loại âm thanh. Nó còn bắt chước tiếng cá heo và tự tạo ra âm thanh mới. Các thí nghiệm phát âm thanh này cho cá heo nghe trên biển cũng đang tiếp tục. Âm thanh mà DolphinGemma học được thu thập trong thời gian dài. Wild Dolphin Project đã thu âm từ năm 1985. Đó là tiếng của cá heo đốm Đại Tây Dương sống ở vùng biển Bahamas. Họ ghi lại cả hành vi như giao phối hoặc tranh chấp cùng với âm thanh. Google đã mở chương trình này để ai cũng có thể dùng.

Tài liệu căn cứ liên quan

[9] Cañas, J. S., Toledo, L. F., et al. (2023). AnuraSet: A dataset for benchmarking Neotropical anuran calls identification in passive acoustic monitoring. *PeerJ Computer Science*, 9, e1575. <https://doi.org/10.7717/peerj-cs.1575>

[10] Google Research. (2026, February 9). How AI trained on birds is surfacing underwater mysteries. Google Research Blog. <https://research.google/blog/how-ai-trained-on-birds-is-surfacing-underwater-mysteries/>

[11] Google. (2025, April 14). DolphinGemma: How Google AI is helping decode dolphin communication. Google Blog. <https://blog.google/innovation-and-ai/products/dolphingemma/>

4.5 AI nhỏ bảo vệ cá heo hồng trong nước đục

Trí tuệ nhân tạo dưới nước làm việc không chỉ ở biển rộng, mà cả ở sông. Đó là trong làn nước đục của Amazon, nơi mắt không thể nhìn thấy gì. Ở đây có loài cá heo sông thân màu hồng tên là Boto. Mamirauá Institute của Brazil và Viện Nghiên cứu Amazon Quốc gia (INPA) dẫn đầu việc bảo vệ Boto [12].

Boto hiện đang giảm nhanh. Chúng mắc vào lưới do con người bỏ dưới sông. Chúng bị nhiễm thủy ngân chảy ra từ việc khai thác vàng. Chúng cũng chịu tiếng ồn do phát triển gây ra. Sông Amazon có nhiều đất và lá cây, nên nước luôn đục. Vì vậy máy ảnh hay drone không thể tìm Boto. Chỉ âm thanh mới là chìa khóa để bảo vệ Boto [13].

Nhóm nghiên cứu đặt micro nghe âm thanh xuống dưới nước sông. Micro này ghi âm liên tục trong thời gian dài. Con người khó nghe hết các bản ghi dài như vậy. Vì thế trí tuệ nhân tạo kiểm tra thay. Trí tuệ nhân tạo được dùng là máy nhìn hình ảnh âm thanh để phán đoán. Nó lọc bỏ tiếng mưa gió và tiếng cành cây trong rừng. Sau đó nó tìm tiếng lách cách do Boto phát ra.

Trí tuệ nhân tạo này tìm ra tiếng Boto với độ chính xác 95%. Nó cũng phân biệt tiếng tàu đi qua với độ chính xác 92%. Dưới lớp nước đục tối hoàn toàn, máy giữ vị trí của Boto chỉ bằng âm thanh. Nó cũng phân biệt được tiếng tucuxi, một loài cá heo khác sống trong sông. Như vậy, trí tuệ nhân tạo tìm cá heo không nhìn thấy bằng âm thanh.

Bản ghi này được dùng ngay để bảo vệ boto. Nhóm nghiên cứu đã dùng âm thanh để vẽ ra những lối boto thường đi qua. Họ cũng biết được vào mùa nào boto tụ lại ở đâu. Khi đến mùa nước dâng, boto rời

sông lớn và đi vào khu rừng bị ngập nước. Khi nghe và xem xét âm thanh, chuyển động này hiện ra rất rõ. Bản đồ này giúp đổi đường đi của tàu thuyền và nơi đặt lưới. Âm thanh đã trở thành biển chỉ đường bảo vệ mạng sống của boto. Ngay cả ở vùng rừng sâu xa xôi, nơi không có máy tính lớn, các nghiên cứu vẫn tiếp tục. Họ tìm cách nghe âm thanh lâu dài và bảo vệ boto [2].

Tài liệu căn cứ liên quan

- [2] NOAA Fisheries. (2025). A whale's world of sound. National Oceanic and Atmospheric Administration. <https://www.fisheries.noaa.gov/feature-story/whales-world-sound>
- [12] Erbs, F., Gaona, M., van der Schaar, M., Zaugg, S., & André, M. (2023). Towards automated long-term acoustic monitoring of endangered river dolphins: A case study in the Brazilian Amazon floodplains. *Scientific Reports*, 13, 10801. <https://doi.org/10.1038/s41598-023-36518-1>
- [13] Kane, S. (2023, August 1). AI unlocks secrets of Amazon river dolphins' behavior, no tagging required. Mongabay. <https://news.mongabay.com/2023/08/ai-unlocks-secrets-of-amazon-river-dolphins-behavior-no-tagging-required/>

4.6 Lắng nghe biển cả

Nghiên cứu dưới nước bằng trí tuệ nhân tạo trong Chương 4 làm thay đổi một suy nghĩ cũ. Trong thời gian dài, con người xem các loài vật khác chỉ là sinh vật làm theo bản năng [1]. Nhưng các nghiên cứu gần đây cho thấy một hình ảnh khác [1]. Phương ngữ của cá voi sát thủ cho thấy văn hóa được truyền qua nhiều thế hệ [1]. Cá heo trắng trao đổi âm thanh với nhau trong sông. Điều đó là bằng chứng rằng chúng tạo thành một xã hội gắn với nhau bằng các mối quan hệ [1].

Những sự thật mới về tiếng nói của cá voi cũng tiếp tục xuất hiện. Đó là nghiên cứu của Project CETI và các nhóm ở nhiều đại học. Năm 2024, họ đã làm rõ cấu trúc âm thanh của cá nhà táng. Cá nhà táng nổi tiếng click để tạo thành tín hiệu. Nhóm âm thanh này được gọi là coda. Nhóm nghiên cứu tìm thấy trong coda có các quy tắc như tốc độ và nhịp điệu. Khi trộn các quy tắc này, rất nhiều âm thanh khác nhau xuất hiện. Điều đó có nghĩa là tín hiệu ấy có cấu trúc hơn ta tưởng nhiều. Nhóm nghiên cứu cũng công bố một “bảng chữ cái âm thanh” sắp xếp các mảnh âm thanh này. Năm 2026, họ tiến thêm một bước. Trong âm thanh của cá nhà táng, họ tìm thấy hai dạng âm giống nguyên âm trong tiếng người. Chúng được chia giống như âm “a” và âm “i” của con người. Nhưng điều này không có nghĩa là ta đã hiểu được ý nghĩa trong tiếng nói của cá nhà táng [14][15].

Trí tuệ nhân tạo giúp hỏi và đáp về âm thanh động vật bằng lời người cũng đang phát triển. Đó là NatureLM-audio do Earth Species Project tạo ra. Đây là mô hình nền tảng âm thanh và ngôn ngữ đầu tiên dành cho âm thanh động vật. Khi con người hỏi bằng lời bình thường, trí tuệ nhân tạo xem xét âm thanh rồi trả lời. Nó phân loại loài, tìm âm thanh, và còn đếm xem có bao nhiêu con đang kêu [16].

Công nghệ đọc tiếng nói của cá voi cũng đặt ra câu hỏi mới cho pháp luật. Tháng 10 năm 2025, các học giả luật và nhóm nghiên cứu Project CETI cùng công bố một bài luận. Họ xem cá voi là sinh vật biết chia sẻ văn hóa và giao tiếp với nhau. Vì vậy, họ đề xuất trao cho cá voi quyền không bị đối xử tàn ác. Nhóm “sự sống vượt ngoài con người (MOTH)” của Trường Luật Đại học New York dẫn dắt việc này. Đôi tai mới do trí tuệ nhân tạo mở ra đã dẫn tới câu hỏi về luật pháp và đạo đức [17].

Từ trước đến nay, chúng ta đã đổ rất nhiều tiếng ồn vào biển. Tiếng tàu lớn và tiếng nổ thăm dò đã cắt đứt con đường giao tiếp của cá voi. Giờ đây, trí tuệ nhân tạo đã trở thành chiếc ống nghe đọc âm thanh biển cả một cách cẩn thận. Công cụ này nói với con người một điều. Đó là hãy hạ thấp tiếng nói của mình và lắng nghe âm thanh tự nhiên như vốn có. Khi bỏ xuống ý muốn thống trị biển cả, chúng ta mới có thể nghe đúng bài ca lâu đời của cá voi. Dân nhạc dưới biển sâu vẫn chưa ngừng biểu diễn. Chúng ta càng lắng tai nghe, âm thanh ấy càng quay trở lại với chúng ta [2][1].

Tài liệu căn cứ liên quan

- [1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.
- [2] NOAA Fisheries. (2025). A whale's world of sound. National Oceanic and Atmospheric Administration. <https://www.fisheries.noaa.gov/feature-story/whales-world-sound>
- [14] Sharma, P., Gero, S., Payne, R., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [15] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293(2069). <https://doi.org/10.1098/rspb.2025.2994>
- [16] Robinson, D., Miron, M., Hagiwara, M., Pietquin, O., et al. (2024). NatureLM-audio: An audio-language foundation model for bioacoustics. arXiv:2411.07186. <https://arxiv.org/abs/2411.07186>
- [17] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025, October). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. Ecology Law Quarterly, 52(1). <https://www.ecologylawquarterly.org/print/what-if-we-understood-what-animals-are-saying-the-legal-impact-of-ai-assisted-studies-of-animal-communication/>

Chương 5. Ngôn ngữ của mặt đất

5.1 Âm thanh của tinh tinh có quy tắc

Tiếng nói của con người được tạo bằng cách nối vài âm lại với nhau theo nhiều cách. Từ một số ít âm, con người tạo ra vô số ý nghĩa. Trong thời gian dài, con người tin rằng tài năng này chỉ có ở con người. Nhưng các học giả quan sát âm thanh của động vật trong thời gian dài đã nhìn lại niềm tin này.

Sân khấu là Vườn quốc gia Taï ở Côte d'Ivoire, phía tây châu Phi. Ở đây có các đàn tinh tinh hoang dã sinh sống. Các học giả của Viện Nhân học Tiến hóa Max Planck ở Đức và Trung tâm Nghiên cứu Khoa học Quốc gia Pháp đã quan sát chúng trong thời gian dài. Tiến sĩ Cédric Girard-Buttoz và Tiến sĩ Roman Wittig dẫn đầu nghiên cứu này [1].

Nghiên cứu này không kết thúc trong một hai năm. Đây là kết quả quan sát cùng một đàn trong nhiều năm. Các học giả đã thu thập âm thanh của 46 con tinh tinh trưởng thành. Số âm thanh được ghi lại lên tới 4.826. Trước đây, người ta chỉ xem những âm thanh này là tiếng rên bộc lộ cảm xúc. Nhưng khi đếm từng âm thanh, họ thấy bên trong có trật tự.

Các học giả đã nghe lâu và phân loại rất nhiều âm thanh này. Họ chia cao độ và độ dài của âm thanh như dùng thước để đo. Kết quả là âm thanh cơ bản mà tinh tinh dùng được chia thành 12 loại. Đó là những âm như tiếng rên, tiếng sửa, tiếng la hét. Có thể xem chúng là những mảnh âm nhỏ nhất, giống phụ âm và nguyên âm trong tiếng người.

Tinh tinh không phát ra 12 âm này một cách tùy tiện. Chúng nối các âm theo một thứ tự nhất định. Những nhóm âm được tạo như vậy có hơn 390 loại. Có thứ tự xuất hiện thường xuyên. Có thứ tự hầu như không xuất hiện. Điều đó có nghĩa là có khuôn mẫu khi nối âm.

Ngày nay, khi xem xét nhiều âm thanh như vậy, trí tuệ nhân tạo gọi là học máy giúp các học giả. Học máy nhanh chóng tìm ra cấu trúc lặp lại trong vô số âm thanh. Việc mà một người làm một mình có thể mất nhiều năm, nó làm trong vài ngày. Người tìm ra quy tắc là học giả, còn trí tuệ nhân tạo là công cụ giúp việc đó.

Các nhóm âm tạo thành nhiều tầng. Một nhóm gồm hai âm cũng có thể nằm bên trong một nhóm gồm ba âm. Tiếng người cũng tạo thành các tầng như vậy. Điều này giống như từ hợp lại thành cụm từ, rồi cụm từ hợp lại thành câu.

Các học giả nêu một đặc điểm lớn của tiếng người. Đó là sức mạnh nối những mảnh âm nhỏ không có nghĩa để tạo ra nghĩa mới. Một mảnh âm riêng lẻ không có nghĩa. Chỉ khi được nối theo thứ tự, nó mới tạo ra nghĩa. Sức mạnh này được gọi là cấu trúc kép của âm thanh. Tuy vậy, 12 âm của tinh tinh không hoàn toàn giống phụ âm và nguyên âm của con người. Các học giả vẫn đang xem xét thận trọng [2].

Theo thời gian, nghiên cứu mới tiếp tục xuất hiện. Năm 2025, các học giả cho thấy tinh tinh ghép hai âm lại để đổi nghĩa. Khi xem xét các nhóm hai âm như vậy, họ thấy có 16 loại. Có nhóm giữ nguyên nghĩa của hai âm rồi cộng lại. Có nhóm lại đổi thành một nghĩa hoàn toàn mới.

Ví dụ, khi một âm khác được gắn sau một âm nào đó, nghĩa thay đổi. Nó có thể trở thành tín hiệu rủ nhau làm chỗ ngủ trên cây. Tinh tinh cũng nối âm báo nguy hiểm với âm gọi giúp đỡ để báo hãy coi chừng rắn. Khi chỉ có một âm và khi có hai âm, ý nghĩa khác nhau [3].

Những phát hiện này là dấu chân cho biết tiếng người đã sinh ra như thế nào. Tinh tinh dùng các nhóm âm này để báo nguy hiểm và phối hợp hành động với nhau. Điều đó có nghĩa là tài năng kết nối âm thanh không chỉ thuộc về con người. Quy tắc âm thanh của tinh tinh cho thấy dấu vết giao tiếp đã có từ xa xưa [4].

Tài liệu căn cứ liên quan

- [1] Girard-Buttoz, C., Wittig, R. M., et al. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications Biology*, 5, 410. <https://doi.org/10.1038/s42003-022-03350-8>
- [2] Girard-Buttoz, C., et al. (2025). Versatile use of chimpanzee call combinations promotes meaning expansion. *Science Advances*, 11(19), eadq2879. <https://doi.org/10.1126/sciadv.adq2879>
- [3] Leroux, M., et al. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, 14, 2225. <https://doi.org/10.1038/s41467-023-37816-y>
- [4] Sun, W. (2025, May 9). Origins of language: Wild chimps mirror linguistic structures in human language. *Phys.org*. <https://phys.org/news/2025-05-language-wild-chimps-mirror-linguistic.html>

5.2 Nhịp lời giống con người của khỉ đầu chó gelada

Nếu tình tình cho thấy quy tắc âm thanh trong rừng, thì một loài khỉ khác lại phát ra âm thanh giống con người hơn. Đó là khỉ đầu chó gelada (*Theropithecus gelada*), sống trên núi cao ở Ethiopia. Chúng sống thành đàn lớn trên những vách đá hiểm trở. Âm thanh chúng trao đổi nghe giống như con người đang trò chuyện rì rầm [1][2].

Âm thanh của gelada nghe thân mật với tai người. Nó giống như nhiều người đang lẩm bẩm bằng giọng thấp. Âm thanh ấy vang quanh vách đá suốt cả ngày. Vì vậy, các học giả đã bị cuốn hút bởi âm thanh này trong thời gian dài.

Giáo sư Thore Bergman của Đại học Michigan, Hoa Kỳ, đã xem xét âm thanh này trong thời gian dài. Khi gelada sống hòa thuận, chúng rung môi rất nhanh. Âm thanh này được gọi là rung môi. Chúng chồng thêm lên đó một âm trầm phát ra từ cổ họng. Âm này được gọi là âm wobble.

Giáo sư Bergman đã đo nhịp của âm thanh này. Khi con người nói, miệng mở và đóng khoảng ba đến tám lần trong 1 giây. Âm wobble của gelada rung sáu đến chín lần trong 1 giây. Hai nhịp này chồng lên nhau. Đơn vị đếm số lần lặp lại trong 1 giây được gọi là hertz (Hz).

Phát hiện này được công bố vào năm 2013. Con người và gelada sống ở những nơi rất xa nhau. Thế nhưng nhịp phát âm lại giống nhau. Giống như chúng đã hẹn trước với nhau.

Gelada không chỉ rung môi. Chúng chồng hai âm lên nhau để chứa nhiều ý nghĩa hơn. Các học giả đã so sánh âm thanh của gelada với âm thanh của khỉ đầu chó chacma. Đàn sống chung càng lớn thì loại âm thanh càng nhiều. Nếu có nhiều hàng xóm để trao đổi, thì cũng có nhiều lời để trao đổi.

Ngày nay, dữ liệu âm thanh như vậy đã tích lũy rất nhiều. Vì vậy, các học giả bắt đầu dùng trí tuệ nhân tạo làm công cụ. Trí tuệ nhân tạo nhanh chóng phân loại vô số âm thanh. Nó cho thấy rõ bằng con số những nhịp mà trước đây chỉ đoán bằng tai. Người phát hiện là học giả, còn trí tuệ nhân tạo là bàn tay giúp việc đó [3].

Ở đây nảy sinh một câu hỏi. Con người và gelada có tổ tiên khác nhau, vậy vì sao nhịp lại giống nhau? Việc những loài có hình dáng khác nhau tiến theo cùng một hướng được gọi là tiến hóa hội tụ. Nó giống như hai lớp khác nhau tự làm khẩu hiệu cổ vũ, nhưng lại có nhịp gần giống nhau. Khi sống theo cách giống nhau, các loài có thể đạt tới tài năng giống nhau.

Tài liệu căn cứ liên quan

- [1] Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, 23(7), R268-R269. <https://doi.org/10.1016/j.cub.2013.02.038>
- [2] Gustison, M. L., le Roux, A., & Bergman, T. J. (2012). Derived vocalizations of geladas (*Theropithecus gelada*) and the evolution of vocal complexity in primates. *Philosophical Transactions of the Royal Society B*, 367(1597), 1847-1859. <https://doi.org/10.1098/rstb.2011.0218>

5.3 Phương ngữ của chuột chũi trụi lông dưới lòng đất

Nếu trên mặt đất sáng, khi trao đổi âm thanh, thì dưới lòng đất tối cũng có một xã hội của âm thanh. Đó là trong hang đất không có một chút ánh sáng. Ở đây có chuột chũi trụi lông (*Heterocephalus glaber*) sinh sống. Loài này sống thành đàn giống kiến hoặc ong. Một con nữ hoàng sinh con, còn những con khác chia nhau làm việc. Xã hội chia vai trò như nữ hoàng và thợ như vậy được gọi là tính xã hội thật sự [1].

Nhóm nghiên cứu tại Trung tâm Max Delbrück ở Berlin, Đức, đã thu thập âm thanh của chúng. Tiến sĩ Alison Barker dẫn đầu nghiên cứu. Khi chuột chũi trụi lông gặp nhau, chúng kêu chirp chirp nhẹ nhàng. Nhóm nghiên cứu thu thập hơn 30.000 âm thanh như vậy và xem xét từng âm. Họ rút ra nhiều đặc điểm như cao độ và độ dài của âm thanh, giống như dùng thước để đo [2].

Điều được phát hiện là như sau. Mỗi đàn có thói quen âm thanh riêng. Nếu so với con người, nó giống phương ngữ khác nhau giữa các vùng. Điều này được gọi là phương ngữ đàn dưới lòng đất. Chuột chũi trụi lông không thể nhìn bằng mắt trong hang tối. Chúng chỉ dùng âm thanh để phân biệt phe mình và kẻ khác.

Trong hang lúc nào cũng tối đen. Vì không có ánh sáng, chúng không thể nhận ra khuôn mặt. Vì vậy âm thanh đóng vai trò như thẻ tên. Nếu dùng cùng phương ngữ, đó là cùng đàn. Nếu nghe phương ngữ lạ, chúng xem đó là đàn khác. Một âm thanh giống như người gác cửa bảo vệ cả đàn.

Phương ngữ này không phải bẩm sinh. Con non mới sinh liên tục bắt chước âm thanh của con trưởng thành. Nhờ vậy, chúng học thói quen âm thanh của đàn. Những con chuột mới nhập vào từ đàn khác cũng như vậy. Chúng bắt chước âm thanh của đàn mới và học phương ngữ đó. Phương ngữ không do huyết thống quyết định, mà do học tập quyết định.

Con đường học này giống cách em bé học nói. Giống như em bé nghe lời người lớn rồi dần bắt chước, chuột con cũng học âm thanh như vậy [3].

Điểm đáng chú ý nhất là sức mạnh của nữ hoàng. Khi nữ hoàng còn sống, phương ngữ của đàn được giữ thống nhất. Nhưng nếu nữ hoàng đột ngột chết, phương ngữ bị rối loạn. Quy tắc âm thanh đồng loạt lung lay. Có thể xem nữ hoàng là cột trụ buộc âm thanh của đàn thành một khối.

Trí tuệ nhân tạo đã giúp ích trong việc phân loại rất nhiều âm thanh này. Trí tuệ nhân tạo nhanh chóng nhận ra thói quen âm thanh khác nhau của từng đàn. Nhưng việc phát hiện rằng phương ngữ được học qua học tập và được nữ hoàng giữ gìn là kết quả của quan sát lâu dài. Trí tuệ nhân tạo là công cụ giúp cho quan sát đó.

Sự thật rằng nữ hoàng sụp đổ thì âm thanh cũng sụp đổ có ý nghĩa sâu. Để buộc âm thanh thành một khối, cần có một trung tâm. Điều này cũng giống quy tắc mà cả lớp chúng ta cùng đặt ra. Nếu trung tâm giữ quy tắc bị lung lay, quy tắc cũng lung lay.

Tài liệu căn cứ liên quan

[1] Barker, A. J., et al. (2021). Cultural transmission of vocal dialect in the naked mole-rat. *Science*, 371(6528), 503-507.
<https://doi.org/10.1126/science.abc6588>

[2] Max Delbrück Center for Molecular Medicine. (2021, January 28). Naked mole-rats 'chirp' in colony-specific, culturally transmitted dialects. *EurekAlert!* <https://www.eurekalert.org/news-releases/510214>

[3] Vaidyanathan, G. (2021, February). The queen sets the tone: Deciphering the dialects of naked mole-rats. *Mongabay News*. <https://news.mongabay.com/2021/02/the-queen-sets-the-tone-deciphering-the-dialects-of-naked-mole-rats/>

5.4 Tiếng báo động của meerkat và tiếng cười của linh cẩu

Nhiều loài vật trên mặt đất cũng chứa nhiều ý nghĩa trong âm thanh. Meerkat (*Suricata suricatta*) là một ví dụ tốt. Meerkat dùng âm thanh để báo kẻ thù đến từ trời hay từ đất. Chúng phát âm khác nhau khi đại bàng đến và khi rắn đến. Chúng còn đưa vào âm thanh mức độ nguy hiểm khẩn cấp đến đâu [1].

Sự thật này đã được Giáo sư Marta Manser của Thụy Sĩ phát hiện từ lâu. Giáo sư Manser đã quan sát meerkat trong thời gian dài tại sa mạc Kalahari ở Nam Phi. Bà nghe và phân loại từng tiếng báo động. Nguy hiểm càng khẩn cấp thì âm thanh càng thô và ồn hơn. Có âm thanh riêng để báo loại nguy hiểm. Có âm thanh riêng để báo mức độ khẩn cấp. Meerkat dùng một âm thanh để khiến cả đàn cùng di chuyển.

Ngày nay, trí tuệ nhân tạo cũng phân loại những tiếng báo động như vậy. Trí tuệ nhân tạo nhanh chóng nhận ra những nét âm nhỏ mà con người khó nghe thấy. Nhưng việc âm thanh chứa loại nguy hiểm và mức độ khẩn cấp đã được quan sát thực địa lâu dài phát hiện trước.

Linh cẩu đốm (*Crocuta crocuta*), loài làm chủ đêm trên đồng cỏ châu Phi, cũng như vậy. Linh cẩu phát ra âm thanh giống tiếng cười. Trong tiếng cười khúc khích này có nhiều thông tin. Nhóm nghiên cứu của Nicolas Mathevon ở Pháp đã xem xét âm thanh này. Nó nghe như tiếng ồn lộn xộn, nhưng không phải vậy. Bên trong có tuổi và thứ bậc trong đàn. Nó cũng chứa dấu hiệu cho biết từng cá thể khác nhau. Nó giống dấu vân tay bằng giọng nói của con người [2].

Nhóm nghiên cứu của Mathevon đã ghi âm tiếng của 17 con linh cẩu ở Berkeley, Hoa Kỳ. Sau đó họ rút ra hai nét của tiếng cười. Một nét là hoa văn âm thanh do hình dạng miệng và cổ tạo ra. Nét này được gọi là formant. Nét kia là cao độ của âm thanh. Nét này được gọi là pitch. Linh cẩu ở vị trí cao và linh cẩu ở vị trí thấp khác nhau ở hai nét này. Chỉ nghe âm thanh cũng có thể biết con nào mạnh hơn.

Chó (*Canis lupus familiaris*), loài gần gũi với con người, cũng như vậy. Nhóm nghiên cứu của Đại học Michigan, Hoa Kỳ, đã xem xét âm thanh của chó. Họ thu thập tiếng sủa, tiếng gầm gừ và tiếng rên ư ử. Sau đó họ thử dùng trí tuệ nhân tạo vốn phân tích tiếng người cho âm thanh của chó. Tên của trí tuệ nhân tạo này là Wav2Vec2. Kết quả là nó phân biệt được âm thanh vui đùa hay hung dữ. Nó cũng đoán được ở mức nào đó giống chó, giới tính và tuổi. Tỷ lệ đoán đúng đạt 70%. Công cụ xử lý tiếng người cũng có tác dụng với âm thanh của chó [3].

Ở đây nảy sinh một câu hỏi. Ở trên, ta nói nhịp của gelada và con người giống nhau là tiến hóa hội tụ. Tổ tiên khác nhau, nhưng vì sống theo cách giống nhau nên trở nên giống nhau. Nhưng nền tảng phát âm của con người và chó giống nhau vì một lý do khác. Con người và chó đều là động vật có vú. Từ rất xa xưa, chúng tách ra từ một tổ tiên chung. Vì vậy, cấu trúc cơ thể dùng cổ và dây thanh để phát ra âm thanh đã giống nhau từ đầu. Sự giống nhau có cùng tổ tiên và cùng gốc rễ như vậy được gọi là tương đồng. Nó giống như bàn tay và chân trước tách ra từ cùng loại xương. Sự giống nhau có hai con đường như vậy. Một là con đường sống riêng rồi trở nên giống nhau. Hai là con đường tách ra từ một gốc chung.

Tài liệu căn cứ liên quan

- [1] Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society B*, 268(1483), 2315-2324. <https://doi.org/10.1098/rspb.2001.1773>
- [2] Mathevon, N., et al. (2010). What the hyena's laugh tells: Sex, age, dominance and individual signature in the giggling call of *Crocuta crocuta*. *BMC Ecology*, 10, 9. <https://doi.org/10.1186/1472-6785-10-9>
- [3] Abzaliev, A., Pérez Espinosa, H., & Mihalcea, R. (2024). Towards dog bark decoding: Leveraging human speech processing for automated bark classification. *Proceedings of LREC-COLING 2024*. <https://aclanthology.org/2024.lrec-main.1432/>

5.5 Một trí tuệ nhân tạo học âm thanh của nhiều loài vật

Cho đến nay, các học giả đã xem xét âm thanh của từng loài vật riêng biệt. Tinh tinh thì được xem xét theo tinh tinh. Gelada thì được xem xét theo gelada. Nhưng ngày nay, một trí tuệ nhân tạo học chung âm thanh của nhiều loài vật [1].

Earth Species Project đứng ở phía trước việc này. Tổ chức này muốn giải mã âm thanh của động vật bằng trí tuệ nhân tạo. Họ đã tạo ra một mô hình tên là NatureLM-audio. Đây là trí tuệ nhân tạo nghe âm thanh động vật và trả lời câu hỏi của con người. Nó lần đầu ra mắt vào tháng 11 năm 2024, và được mở để ai cũng dùng vào năm 2025.

Mô hình này đã học âm thanh của hàng nghìn loài. Chỉ nghe âm thanh, nó đoán đó là loài vật nào. Nó cũng phân biệt con cái hay con đực, con trưởng thành hay con non. Nó còn đếm trong bản ghi có bao nhiêu con vật. Đôi khi nó tự cho thấy cả những tài năng chưa được dạy.

Tổ chức này nhận được sự giúp đỡ lớn vào tháng 10 năm 2024. Họ nhận 17 triệu đô la tiền nghiên cứu từ nhiều nhà tài trợ. Đây là số tiền lớn, hơn 23 tỷ won Hàn Quốc. Với số tiền này, họ tăng mạnh số nhà nghiên cứu. Tổ chức này cho rằng đến năm 2030, họ có thể làm sáng tỏ nhiều điều về âm thanh của động vật [2].

Vào mùa hè năm 2026, nhiều nhóm nghiên cứu cũng mở một kho dữ liệu để cùng dùng. Tên của kho này là dữ liệu xử lý ngôn ngữ động vật (alp-data). Đây là nơi gom các dữ liệu âm thanh động vật vốn nằm rải rác về một chỗ. Hiện nay, người ta cùng xem xét âm thanh của quạ, cá voi trắng và voi. Dữ liệu càng nhiều, trí tuệ nhân tạo học càng nhanh [5].

Cũng đã có một giải thưởng giúp thúc đẩy hướng nghiên cứu này. Đó là Giải Coller-Dolittle (Coller-Dolittle Prize). Giải này tiếp sức cho nghiên cứu giải mã âm thanh của động vật. Người đoạt giải năm 2026 là Tiến sĩ Julie Elie ở Berkeley, Hoa Kỳ. Tiến sĩ Elie đã giải mã tiếng kêu của một loài chim nhỏ tên là zebra finch. Bà tìm ra 11 âm cơ bản mà loài chim này dùng. Trong đó có những ý như đói, sợ hãi và chào hỏi. Tiến sĩ Elie cũng nhận được sự giúp đỡ của trí tuệ nhân tạo gọi là học máy [3].

Trí tuệ nhân tạo cũng tìm thấy trong âm thanh động vật một quy luật tiết kiệm giống như trong lời nói con người. Con người thường rút ngắn những từ hay dùng. Càng dùng thường xuyên, âm càng ngắn. Năm 2025, một học giả đã xem xét âm thanh của 16 loài cá voi. Sau đó, ông so sánh các âm này với 51 ngôn ngữ của con người ở nhiều nước. Trong bài hát của cá voi lưng gù và cá voi xanh cũng thấy sự tiết kiệm như vậy. Âm nào được hát thường xuyên hơn thì ngắn hơn. Sự khôn ngoan trong việc tiết kiệm âm thanh không chỉ thuộc về con người. Nhưng không phải loài cá voi nào cũng như vậy, nên các học giả vẫn đang nghiên cứu thêm [4].

Hãy nhớ lại quy luật âm thanh của tinh tinh, nhịp lời nói của gelada, và thổ ngữ của chuột chũi trụi lông. Âm thanh của những người láng giềng này cho chúng ta nhiều điều để suy nghĩ. Ngôn ngữ con người không phải là một phép màu rơi thẳng từ trời xuống. Nó nở ra trên một nền tảng rộng lớn, nơi nhiều sinh vật cùng đứng chung.

Trí tuệ nhân tạo giúp tìm ra các quy luật bên trong âm thanh. Việc một người làm một mình có thể mất nhiều năm, nó làm được trong vài ngày. Nhờ vậy, các học giả có thể xem xét nhiều âm thanh hơn. Tuy nhiên, trí tuệ nhân tạo không dịch nguyên lời nói của động vật. Nó chỉ giúp phân loại âm thanh và tìm quy luật.

Hiểu âm thanh của động vật cũng giúp bảo vệ chúng. Nếu biết âm nào có nghĩa là sợ hãi, ta có thể nhận ra khi động vật đang khổ sở. Ta cũng có thể đoán qua âm thanh điều gì đang làm chúng khó chịu. Khả năng đọc âm thanh trở thành khả năng chăm sóc những người láng giềng.

Trong suốt thời gian qua, chúng ta chỉ xem những âm thanh này là tiếng ồn trong bụi cây. Chúng ta cũng từng tùy tiện đào xới vùng đất nơi chúng sống. Nhưng chúng đã dùng âm thanh để nối tiếp lịch sử của bầy đàn trong thời gian dài. Việc chúng ta cần làm bây giờ không phải là cho chúng nghe âm thanh nhân tạo trước. Chúng ta phải bắt đầu bằng việc khiêm nhường lắng nghe tiếng nói của chúng.

Trí tuệ nhân tạo trở thành đôi tai tốt giúp việc lắng nghe đó. Nó nghe thay cả những chi tiết nhỏ mà con người từng bỏ lỡ. Nhưng nghe điều gì và vì sao nghe là việc con người quyết định. Ta phải cầm công cụ tốt và đi cùng một tấm lòng khiêm nhường biết lắng nghe. Khi đó, ta mới có thể sống lâu dài bên cạnh những người láng giềng của mặt đất.

Tài liệu căn cứ liên quan

- [1] Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (Earth Species Project). (2024). NatureLM-audio: An audio-language foundation model for bioacoustics. arXiv:2411.07186. <https://arxiv.org/abs/2411.07186>
- [2] Earth Species Project. (2024, October 17). Earth Species Project secures \$17M in grants to apply frontier AI to decoding animal communication. <https://earthspecies.org/2024/10/17/earth-species-project-secures-17m-in-grants-to-apply-frontier-ai-to-decoding-animal-communication/>
- [3] Jeremy Collier Foundation. (2026, July 3). Scientists win US\$100,000 in 2026 Collier Dolittle Prize. <https://www.jeremycollerfoundation.org/news-and-insights/press-releases/scientists-win-us100-000-in-2026-coller-dolittle-prize>
- [4] Youngblood, M. (2025). Language-like efficiency in whale communication. Science Advances, 11(6), eads6014. <https://doi.org/10.1126/sciadv.ads6014>
- [5] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. <https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>

Chương 6. Bản giao hưởng trên bầu trời và điệu múa vô hình

6.1 Cuộc trò chuyện giữa chim và trí tuệ nhân tạo

Cuộc trò chuyện của con người không phải là bản thảo viết sẵn. Chúng ta nghe lời nói và giọng điệu của người đối diện. Chúng ta còn để ý cả khoảnh khắc họ ngừng lại một chút. Rồi ta đáp lời theo dòng đó. Trong tiếng hót của chim cũng có sự qua lại như vậy.

Trước đây, khi nghiên cứu tiếng chim, các nhà khoa học thường phát bản ghi âm cho chim nghe. Cách này gọi là phát lại âm thanh, tức playback. Playback trước đây thường là phát một âm thanh ngắn đã định sẵn hoặc một bản ghi cố định. Cách đó khó chứa được dòng trò chuyện dài qua lại của chim. Vì vậy, trong thời gian dài, người ta cho rằng tiếng hót của chim là âm thanh đã được định sẵn từ khi sinh ra [1].

Để vượt qua bức tường này, hai nhóm nghiên cứu đã hợp tác. Đó là Earth Species Project, dự án muốn giải mã giao tiếp của động vật bằng trí tuệ nhân tạo, và McGill University ở Canada. Tiến sĩ Logan James và giáo sư Sarah Woolley dẫn dắt nghiên cứu. Họ thu thập âm thanh của chim sẻ vằn cái (zebra finch), một loài chim nhỏ cỡ chim sẻ. Âm thanh trao đổi qua lại dài hơn 1.000 giờ [2].

Nhóm nghiên cứu dùng các âm thanh này để tạo mô hình trí tuệ nhân tạo ZF-AIM. Khi một con chim hót một đoạn, mô hình này liền tạo đoạn tiếp theo để đáp lại. Nó giống như điện thoại thông minh gợi ý từ tiếp theo khi ta gõ chữ. Cách nghe âm thanh của bên kia rồi nối lời đáp ngay như vậy được gọi là trao đổi hội thoại.

Nhóm nghiên cứu thiết kế thí nghiệm rất cẩn thận. Họ đặt riêng chim sẻ vằn cái trong hộp cách âm để âm thanh không lọt ra ngoài. Con chim trong hộp trao đổi âm thanh qua micrô và loa. Khi chim phát ra âm thanh, ZF-AIM tạo âm thanh tiếp theo phù hợp rồi phát lại. Chim và trí tuệ nhân tạo đã trao đổi âm thanh theo thời gian thực.

Nhóm nghiên cứu so sánh ba tình huống. Đó là khi chim trò chuyện với chim sống, khi chỉ phát bản ghi âm, và khi trò chuyện với trí tuệ nhân tạo theo thời gian thực. Những đặc điểm tự nhiên chỉ xuất hiện khi chim trao đổi với chim sống hoặc với trí tuệ nhân tạo. Khi chỉ phát bản ghi âm, những đặc điểm đó không xuất hiện.

Kết quả đã lật lại suy nghĩ lâu nay. Sách giáo khoa viết rằng tiếng gọi của chim sẻ vằn cái do di truyền quy định và không thay đổi. Nhưng những con chim trao đổi âm thanh với trí tuệ nhân tạo thì khác. Chúng thay đổi tiếng của mình theo âm thanh của đối phương. Điều đó cho thấy cuộc trò chuyện của chim không phải là sự lặp lại của âm thanh cố định.

Vì sao nhóm nghiên cứu chọn chim sẻ vằn? Chim sẻ vằn không sống một mình. Nhiều con sống thành bầy và liên tục trao đổi âm thanh. Chúng gọi nhau, đáp lại nhau và di chuyển vị trí. Cuộc trò chuyện sôi nổi như vậy phù hợp với nghiên cứu. Trí tuệ nhân tạo đã bắt chước đúng sự trao đổi nhanh đó.

Nhóm nghiên cứu tìm thấy ba đặc điểm trong cuộc trò chuyện của chim. Thứ nhất, âm thanh của hai con chim dần giống nhau. Thứ hai, chúng thay đổi âm thanh rất nhanh. Thứ ba, chúng không đáp lại mọi âm thanh. Chúng chọn âm thanh mà mình sẽ đáp lại. Điều này giống như trong lớp học ồn ào, ta chỉ quay lại khi nghe tên mình.

Nhóm nghiên cứu thử tắt từng phần của mô hình. Khi tắt phần bắt nhịp, cuộc trò chuyện bị lệch. Khi tắt phần tạo hình dạng âm thanh, phần khác lại bị lệch. Hai phần này đảm nhận hai vai trò khác nhau.

Nhóm nghiên cứu cho rằng cách này cũng giúp hiểu cuộc trò chuyện của con người. Khi trò chuyện, con người cũng đoán lời tiếp theo. Trong sự trao đổi giữa chim và trí tuệ nhân tạo cũng thấy sự đoán trước như vậy. Quy tắc trao đổi âm thanh có thể đúng với nhiều loài động vật.

Nhóm nghiên cứu đặt hai bộ ZF-AIM đối diện nhau và cho chúng trò chuyện. Họ để trí tuệ nhân tạo tự trao đổi âm thanh với nhau, không có bàn tay con người can thiệp. Cuộc trò chuyện đó tái hiện các đặc điểm của cuộc trò chuyện thật giữa các con chim. Điều này có nghĩa là mô hình đã học đúng quy tắc.

Nghiên cứu này dẫn đến một mục tiêu lớn hơn. Đó là dùng âm thanh do trí tuệ nhân tạo tạo ra để nói với động vật. Hiện nay, chúng ta đã đến bước chim đáp lại âm thanh của trí tuệ nhân tạo. Bước tiếp theo là con người gửi âm thanh có chứa ý nghĩa. Khi ngày đó đến, con người và chim có thể thật sự trò chuyện với nhau [3].

Tài liệu căn cứ liên quan

- [1] Earth Species Project. (2023, December 3). Using AI to understand animal communication. Earth Species Project. <https://www.earthspecies.org/blog/using-ai-to-understand-animal-communication>
- [2] James, L. S., Hoffman, B., Woolley, S. C., et al. (2026). Leveraging AI-powered interactive playbacks to decipher rules of communication in zebra finches. bioRxiv. <https://doi.org/10.64898/2026.02.12.705387>
- [3] Earth Species Project. (2026, April 2). Animal Language Processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>

6.2 Tìm quy tắc ẩn trong bài hát

Trong khi chim sẽ vẫn trao đổi âm thanh với trí tuệ nhân tạo, việc học cũng diễn ra trong não chim. Giống con người, chim non bắt chước tiếng của chim trưởng thành khi còn nhỏ. Giai đoạn này gọi là thời kỳ nhạy cảm (sensitive period), tức thời điểm tốt để học âm thanh. Việc học này không chỉ do di truyền quyết định. Bầu không khí của bầy sống cùng cũng ảnh hưởng lớn [1].

Chim sẽ vẫn non lớn lên bằng cách nghe tiếng hót của chim trưởng thành. Lúc đầu, chúng bắt chước còn vụng về. Dần dần, tiếng hót của chúng gần giống tiếng hót của chim trưởng thành. Điều đó giống cách em bé học nói. Càng nghe nhiều trong bầy, chúng càng học tốt.

Trong đầu chim có một vùng dẫn dắt thứ tự bài hát. Vùng này gọi là HVC, có thể gọi là trung tâm chỉ huy bài hát. Vùng này sắp xếp khi nào chim phát ra âm thanh nào. Nó giống hình ảnh nhạc trưởng dẫn nhịp.

Nhóm của giáo sư Timothy Gentner tại UC San Diego ở Hoa Kỳ đã quan sát chim sáo đá châu Âu (European starling). Loài chim này nối hàng trăm mảnh âm thanh thành bài hát dài. Con người khó tìm ra quy tắc đó nếu chỉ nghe bằng tai. Nhóm nghiên cứu đưa các bài hát không có nhãn vào trí tuệ nhân tạo [2].

Trí tuệ nhân tạo gom các mảnh âm thanh giống nhau vào cùng một chỗ. Nó giống như trải bài hát ra thành một tấm bản đồ. Trên bản đồ đó, các âm thanh tạo thành từng nhóm. Cấu trúc mà con người không nhìn thấy trước đây đã hiện ra. Bằng cách này, nhóm nghiên cứu có thể so sánh âm thanh của nhiều loài động vật với nhau.

Nhóm nghiên cứu dùng cách này cho âm thanh của nhiều loài động vật. Họ đặt âm thanh của chim, chuột, khỉ, người và cá voi lên cùng một bản đồ. Dù khác loài, chúng vẫn có thể được so sánh bằng cùng một cách. Những đặc điểm giúp nhận ra âm thanh do loài nào tạo ra cũng hiện rõ.

Con người nghe tiếng chim hót rất lâu vẫn không tìm được hết quy tắc. Lý do là âm thanh nhanh và có nhiều mảnh. Trí tuệ nhân tạo có thể xem xét hàng nghìn giờ tiếng hót mà không mệt. Nó lần lượt chỉ ra những thứ tự mà con người bỏ sót.

Trong tiếng hót của chim hoàng yến (canary) cũng có quy tắc ẩn. Nhóm nghiên cứu của Boston University ở Hoa Kỳ phân tích bài hát dài của chim hoàng yến. Trước đây, người ta chỉ nhìn âm thanh ngay trước đó để đoán âm thanh tiếp theo. Cách đơn giản này gọi là mô hình Markov. Nhưng bài hát của chim hoàng yến không đơn giản như vậy [3].

Nhóm nghiên cứu phát hiện rằng phần đầu của bài hát quyết định âm thanh ở khá lâu về sau. Âm thanh được hót ở đầu bài đã dẫn dắt cả âm thanh ở đoạn xa phía sau. Tính chất này gọi là phụ thuộc dài hạn. Nó giống như từ đầu và từ cuối trong một câu dài đi thành một cặp. Dù có vẻ hót tùy ý, bên trong vẫn có cấu trúc.

Khi các nghiên cứu như vậy tích lũy, một bức tranh lớn hiện ra. Earth Species Project đã tạo ra trí tuệ nhân tạo học chung âm thanh của nhiều loài động vật. Mô hình này có tên là NatureLM-audio. Đây là mô hình học cùng lúc tiếng nói con người, âm nhạc và âm thanh tự nhiên. Mô hình này chỉ nghe âm thanh cũng nhận ra đó là động vật nào. Nó còn đoán được âm thanh của loài động vật mà nó chưa từng học. Khi con người viết câu hỏi bằng chữ, nó nghe âm thanh rồi trả lời. Nếu ta viết hỏi con vật tạo ra âm thanh này là loài nào, nó sẽ chọn loài. Nó cũng đếm xem trong bản ghi có bao nhiêu con đang kêu [4].

Năm 2026, Earth Species Project đặt tên cho dòng nghiên cứu này. Đó là nghiên cứu xử lý âm thanh động vật như dữ liệu, tức xử lý ngôn ngữ động vật (Animal Language Processing). Tên này được dùng lần đầu vào năm 2024. Nó lấy ý từ xử lý ngôn ngữ tự nhiên, lĩnh vực xử lý lời nói của con người. Việc giải quy tắc tiếng hót của chim cũng phát triển trong lĩnh vực này [5].

Tài liệu căn cứ liên quan

- [1] Earth Species Project. (2023, December 3). Using AI to understand animal communication. Earth Species Project. <https://www.earthspecies.org/blog/using-ai-to-understand-animal-communication>
- [2] Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLOS Computational Biology, 16(10), e1008228. <https://doi.org/10.1371/journal.pcbi.1008228>
- [3] Markowitz, J. E., Ivie, E., Kligler, L., & Gardner, T. J. (2013). Long-range order in canary song. PLOS Computational Biology, 9(5), e1003052. <https://doi.org/10.1371/journal.pcbi.1003052>
- [4] Earth Species Project. (2024, November 11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. <https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics>
- [5] Earth Species Project. (2026, April 2). Animal Language Processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>

6.3 Điệu múa hình số 8 trong tổ ong tối đen

Trật tự của bầu trời không chỉ nằm ở những loài chim lớn. Nó cũng nằm trong nhịp vỗ cánh của con ong mật nhỏ bé. Cách giao tiếp tinh vi đầu tiên của động vật mà con người giải thích được là điệu múa của ong mật. Học giả người Áo Karl von Frisch đã phát hiện điều này và nhận giải Nobel [3].

Khi ong mật tìm thấy nơi có mật, nó quay về tổ. Rồi nó lắc mình và vẽ hình số 8. Điệu múa này gọi là múa lắc hình số 8 (waggle dance). Bên trong tổ ong tối đen, không có một tia sáng. Ong mật lấy phía trên trên vách tổ làm mốc. Phía trên là hướng ngược với trọng lực.

Ong mật đi thẳng qua giữa hình số 8. Lúc này, nó báo hướng bằng góc mà cơ thể đi thẳng. Mức nghiêng so với phía trên chính là hướng có mật. Nó báo khoảng cách bằng thời gian lắc mình. Lắc càng lâu thì nơi đó càng xa. Chỉ bằng điệu múa này, ong có thể truyền tin đến nơi cách xa vài kilômét.

Nếu mật ở rất gần, ong không múa hình số 8 mà múa vòng tròn. Khi đó, nó không báo hướng thật chi tiết. Vì nơi đó gần, nên không cần báo nhiều. Khi khoảng cách xa hơn, nó chuyển sang múa hình số 8.

Tín hiệu của ong mật không chỉ là động tác cơ thể. Những rung động nhỏ ở ngực và âm thanh của cánh cùng chuyển động. Cũng có rung động truyền qua vách tổ. Cách dùng nhiều tín hiệu cùng lúc như vậy gọi là đa phương thức. Đó là cách dùng đồng thời động tác, âm thanh và rung động. Dù tổ ong tối đen, đồng loại vẫn cảm nhận các tín hiệu này bằng cơ thể.

Những tín hiệu này giống nhiều giác quan của con người. Khi nói chuyện với bạn, chúng ta không chỉ dùng giọng nói. Chúng ta còn dùng ánh mắt và cử chỉ tay. Ong mật cũng dùng cùng lúc động tác, rung động và âm thanh.

Von Frisch quan sát điệu múa này trong thời gian dài và giải ra quy tắc. Ông đánh dấu từng con ong bằng màu sắc. Ông kiểm tra từng con ong đã đi đâu về. Đó là việc con người phải làm bằng tay trong thời gian dài. Trí tuệ nhân tạo đã làm việc đó nhanh hơn nhiều [1].

Nhóm của giáo sư Tim Landgraf tại Free University of Berlin ở Đức đã tạo ra cách đọc tự động điệu múa này. Nhóm nghiên cứu gắn camera vào tổ ong. Đó là cách không làm ong bị thương. Trí tuệ nhân tạo tự tìm ra điệu múa của ong mật trong video.

Trí tuệ nhân tạo này đọc nhanh những chuyển động mà mắt người bỏ sót. Nó đọc hướng ong mật đi thẳng. Nó cũng đo thời gian lắc mình. Gộp hai thông tin này lại, nó đoán vị trí có mật. Máy móc đã làm thay việc trước đây con người làm bằng tay.

Nhóm nghiên cứu không dừng lại ở đó. Họ đưa điệu múa đã đọc được vào ong robot. Robot này là một cỗ máy bắt chước điệu múa hình số 8 trong tổ ong. Nó khác với robot ong nhỏ biết bay của Harvard University. Khi robot múa ở giữa tổ ong, những con ong thợ thật đã làm theo điệu múa đó [2].

Ong thợ đổi hướng bay theo phía mà robot chỉ. Nhóm nghiên cứu tinh chỉnh chuyển động và rung động của robot trong nhiều năm. Họ cũng cho robot tỏa hương hoa. Nhờ vậy, ong làm theo điệu múa của robot tốt hơn.

Thí nghiệm này cho thấy hai điều. Một là con người đã đọc đúng điệu múa của ong. Nếu robot gửi tín hiệu sai, ong đã không làm theo. Điều còn lại là ong tin vào điệu múa của nhau và di chuyển theo đó. Chỉ một động tác của cơ thể côn trùng nhỏ bé đã chứa cả hướng và khoảng cách.

Tài liệu căn cứ liên quan

[1] Wario, F., Wild, B., Rojas, R., & Landgraf, T. (2017). Automatic detection and decoding of honey bee waggle dances. PLOS ONE, 12(12), e0188626. <https://doi.org/10.1371/journal.pone.0188626>

[2] Landgraf, T., et al. (2018). Dancing honeybee robot elicits dance-following and recruits foragers. arXiv:1803.07126. <https://arxiv.org/abs/1803.07126>

[3] Frisch, K. von. (1967). The dance language and orientation of bees. Harvard University Press.

6.4 Cuộc trò chuyện ba chiều mà quạ vẽ trong không trung

Trong các trí tuệ bay trên bầu trời, quạ rất khác biệt. Quạ nổi tiếng là thông minh. Não chim nhỏ hơn não người. Nhưng tế bào thần kinh tập trung rất dày. Vì vậy, người ta gọi quạ là vượn người có lông vũ [1].

Nhóm nghiên cứu của University of León ở Tây Ban Nha và Earth Species Project đã quan sát quạ. Đối tượng là quạ đen ăn xác (carrion crow) sống ở miền bắc Tây Ban Nha. Nhóm nghiên cứu gắn cảm biến rất nhỏ vào lông đuôi chim. Thiết bị nhẹ, chỉ nặng 12,5g. Họ làm để sau một thời gian nó tự rơi ra. Đó là thiết bị không làm chim bị thương.

Trong cảm biến này có micrô nhỏ và bộ phát hiện chuyển động. Micrô ghi âm thanh của chim. Bộ phát hiện ghi lại rung động của cánh và cơ thể. Dữ liệu cho biết chim phát âm thanh nào vào lúc nào và chuyển động ra sao. Nó giống thiết bị đồng hồ đeo tay dùng để đếm số bước chân của con người.

Nhóm nghiên cứu gắn các cảm biến này trong ba năm. Họ thu âm khi chim non lớn được khoảng mười đến mười lăm ngày. Trí tuệ nhân tạo chia âm thanh thu được thành ba loại. Đó là tiếng của chim trưởng thành đeo cảm biến, tiếng của chim trưởng thành khác ở bên cạnh, và tiếng của chim non. Nó đã phân biệt ai kêu vào lúc nào.

Trí tuệ nhân tạo phân loại các âm thanh đã thu được. Earth Species Project tạo ra một công cụ tên là Voxaboxen. Đây là công cụ tìm từng tiếng kêu trong bản ghi dài. Công cụ này tìm được hơn 120.000 tiếng quạ. Có những âm thanh mà cả nhà nghiên cứu đã quan sát quạ hơn 25 năm cũng lần đầu nghe thấy. Đó là âm thanh quá nhỏ nên tai người thường bỏ sót. Công cụ này cũng tách từng tiếng ra ngay cả khi âm thanh chồng lên nhau. Trong bản ghi có nhiều con chim cùng kêu, nó vẫn làm đúng phần việc của mình [4].

Nhóm nghiên cứu xem riêng dữ liệu âm thanh và dữ liệu chuyển động. Việc đặt hai loại dữ liệu này cạnh nhau để xem cùng lúc vẫn là việc của tương lai. Hiện nay họ mới tạo ra dữ liệu có thể xem cùng nhau như vậy. Nhờ dữ liệu này, con đường để nhìn sâu hơn vào cuộc trò chuyện của quạ đã mở ra.

Trí tuệ nhân tạo chưa làm sáng tỏ hết ý nghĩa của quạ. Trí tuệ nhân tạo đã giúp chia các âm thanh nhỏ thành từng nhóm. Những người làm sáng tỏ tâm trí của quạ là các nhà động vật học từ trước đây [2].

Nhóm của giáo sư John Marzluff tại University of Washington ở Hoa Kỳ đã làm rõ việc quạ nhớ khuôn mặt. Nhóm nghiên cứu đeo mặt nạ cao su và bắt quạ. Sau đó, đàn quạ kêu lớn và lao vào người đeo cùng chiếc mặt nạ ấy. Chúng đã nhớ khuôn mặt của người nguy hiểm. Chúng còn báo ký ức này cho những con quạ khác.

Kaeli Swift và giáo sư Marzluff ở cùng trường đại học đã quan sát một hành vi khác. Quạ tụ tập bên cạnh đồng loại đã chết và kêu lớn. Người ta gọi đó là đám tang. Nhưng đó không chỉ là nỗi buồn. Đó là nơi để học điều gì đã làm đồng loại chết. Chúng báo nguy cơ ấy cho cả đàn và nhớ rất lâu [3].

Trước đây, người ta chỉ xem tiếng quạ là tiếng kêu ồn ào. Khi dùng trí tuệ nhân tạo để chia thành từng nhóm, điều đó không còn đúng. Mỗi âm thanh dường như có một ý nghĩa khác nhau. Có vẻ trong đó có tiếng báo phải coi chừng ai, và tiếng rủ đi đâu đó. Tiếng kêu của chim trông gần với lời nói của con người hơn.

Trong một nghiên cứu khác về chim, các nhà khoa học đã tiến thêm một bước. Năm 2026, Earth Species Project giới thiệu một trí tuệ nhân tạo có thể nhận ra từng con vật chỉ bằng âm thanh. Đối tượng của nghiên cứu này không phải quạ, mà là chim manh manh vằn. Con đường để phân biệt giọng của từng con trong đàn đã mở ra. Giờ đây con người có thể lần lại xem từng con chim đã nói gì và khi nào. Họ có thể xem con nào kêu nhiều, con nào yên lặng. Những công cụ như vậy cũng có thể dùng trong nghiên cứu về quạ [5].

Nghiên cứu sẽ nhanh hơn khi dữ liệu được chia sẻ cùng nhau. Tháng 7 năm 2026, Earth Species Project công bố một công cụ mới. Tên của nó là alp-data. Đây là công cụ gom dữ liệu âm thanh động vật đang rải rác ở nhiều nơi thành một khối. Nó đưa hơn 35 bộ sưu tập âm thanh vào cùng một cách. Trước đây, mỗi nhóm nghiên cứu tự chỉnh dữ liệu riêng. Giờ đây, họ dùng chung bằng một công cụ. Nghiên cứu tiếng quạ cũng sẽ nhanh hơn trên nền dữ liệu chung như vậy. Phía sau bầu trời mà trí tuệ nhân tạo đã vén mở, loài chim vẫn đang trò chuyện mỗi ngày theo cách con người chưa từng biết [6].

Tài liệu căn cứ liên quan

- [1] Baglione, V., Canestrari, D., Cusimano, M., Hoffman, B., Moreno, V., & Trapote, E. (2025). Capturing vocal communication in a free-living corvid: high-resolution data from low-impact miniaturized tags. *Animal Cognition*, 28, 85. <https://doi.org/10.1007/s10071-025-02018-0>

- [2] Marzluff, J. M., Walls, J., Cornell, H. N., Withey, J. C., & Craig, D. P. (2010). Lasting recognition of threatening people by wild American crows. *Animal Behaviour*, 79(3), 699-707.
<https://www.sciencedirect.com/science/article/abs/pii/S0003347209005806>
- [3] Swift, K. N., & Marzluff, J. M. (2015). Wild American crows gather around their dead to learn about danger. *Animal Behaviour*, 109, 187-197. <https://www.sciencedirect.com/science/article/abs/pii/S0003347215003188>
- [4] Mahon, L., et al. (2025). Robust detection of overlapping bioacoustic sound events. arXiv:2503.02389.
<https://arxiv.org/abs/2503.02389>
- [5] Earth Species Project. (2026, May 5). From what to who? AI learns to recognize individual animals. Earth Species Project. <https://earthspecies.org/2026/05/05/from-what-to-who-ai-learns-to-recognize-individual-animals/>
- [6] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for Animal Language Processing. Earth Species Project.
<https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>

Chương 7. Từ não động vật đến tấm vải deep learning

7.1 Trí tuệ nhân tạo tự học âm thanh

Ngày xưa, các nhà khoa học nghe tiếng động vật bằng tai và đo bằng tay. Họ ghi từng độ cao thấp và độ dài của âm thanh. Cách này dùng được trong phòng yên tĩnh. Nhưng ở bờ biển hay trong rừng thì không tốt. Vì tiếng gió và tiếng sóng lẫn vào nhau. Cũng có nhiều tiếng động vật mà chưa ai biết nghĩa. Công việc quá lớn nếu chỉ làm bằng tay người [1].

Ngày nay, trí tuệ nhân tạo gọi là học sâu làm việc này. Học sâu có nhiều cách. Có cách con người cho biết đáp án đúng. Cũng có cách tự tìm quy tắc mà không cần đáp án. Trí tuệ nhân tạo xử lý tiếng động vật thường dùng cách học không cần đáp án. Nó nghe rất nhiều tiếng động vật. Rồi nó tự tìm quy tắc ẩn trong đó.

Có loại trí tuệ nhân tạo học rộng như vậy rồi dùng cho nhiều việc. Loại này được gọi là mô hình nền tảng âm thanh động vật. Trong tiếng Anh là Bioacoustic Foundation Model. Foundation có nghĩa là nền tảng. Nếu nền tảng vững, ta có thể dùng ngay cho vấn đề mới [2, 4].

Mô hình nền tảng dùng một nền đã học cho nhiều việc. Nó dùng để chia âm thanh. Nó cũng dùng để xóa tiếng ồn. Khi có vấn đề mới, nó không học lại từ đầu. Chỉ cần thêm một ít lên nền đã học. Những việc nhỏ làm trên nền đó được gọi là tác vụ ứng dụng. Trong tiếng Anh là downstream task [2, 4].

Một trong những trí tuệ nhân tạo đầu tiên dẫn dắt lĩnh vực này là AVES. AVES học mà không cần bảng đáp án do con người gán. Nó nghe khoảng 5.000 giờ âm thanh thu từ nhiều nơi. Dù nghe đủ 24 giờ mỗi ngày, lượng đó cũng hơn 200 ngày.

Cách AVES học giống trò điền chỗ trống. AVES nhận nguyên dạng sóng ban đầu của âm thanh. Nó cắt nhỏ dạng sóng này. Rồi nó cố ý xóa vài chỗ. Nó nhìn âm thanh trước và sau để đoán lại chỗ đã xóa. Nó lặp lại trò này vô số lần. Nó học bằng chính âm thanh, không đổi sang hình ảnh.

Điều AVES học là thứ của âm thanh. Nó xem âm thanh lặp lại đều đến đâu. Nó xem dòng âm thanh cao lên và thấp xuống. Nó xem hình dạng mà âm thanh vẽ ra. Khi biết thứ này, nó có thể xử lý tiếng của bất kỳ loài vật nào. Vì dù khác loài, thứ âm thanh vẫn có điểm chung.

Trước khi AVES xuất hiện, bảng đáp án là rào cản. Mỗi âm thanh phải được con người đặt tên thì máy mới học được. Âm thanh không có nhãn tên thì máy không học được. AVES mở ra con đường học không cần nhãn tên. Vì vậy, nó có thể dùng mọi loại âm thanh trên đời làm dữ liệu.

Sau quá trình này, trí tuệ nhân tạo đổi đặc điểm âm thanh thành số. Nhóm số này được gọi là embedding. Nó đổi âm thanh thành tọa độ số để máy tính nhận biết. Nhờ embedding, tiếng của bất kỳ loài vật nào cũng có thể được xử lý theo cùng một cách.

Tọa độ này giống sơ đồ chỗ ngồi trong lớp. Nhìn sơ đồ, ta biết ai ngồi ở đâu. Nhìn embedding, ta biết âm thanh nào nằm ở đâu. Âm thanh giống nhau tụ lại gần nhau. Âm thanh ở xa nhau là các âm thanh khác nhau.

Cách này có một điểm mạnh lớn. Con người không cần gán từng đáp án một. Gán đáp án là việc mất nhiều thời gian. Muốn đặt tên cho một âm thanh, chuyên gia phải tự nghe. Trí tuệ nhân tạo vẫn tìm được quy tắc âm thanh dù không có đáp án. Vì vậy, nó có thể xử lý cả những âm thanh mà chưa ai biết nghĩa. Điểm này rất khác cách cũ.

Tài liệu căn cứ liên quan

[1] Hagiwara, M. (2023). AVES: Animal vocalization encoder based on self-supervision. ICASSP 2023 - IEEE International Conference on Acoustics, Speech and Signal Processing. arXiv:2210.14493. <https://arxiv.org/abs/2210.14493>

- [2] Robinson, D., Robinson, A., & Akrapongpisak, L. (2023). Transferable models for bioacoustics with human language supervision (BioLingual). arXiv:2308.04978. <https://arxiv.org/abs/2308.04978>
- [4] Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. ICLR 2025. arXiv:2411.07186. <https://arxiv.org/abs/2411.07186>

7.2 Xử lý cùng lúc tiếng của nhiều loài vật

Muốn nuôi một mô hình nền tảng, cần có nhiều dữ liệu âm thanh. Nhưng mỗi bộ dữ liệu lại có định dạng khác nhau, nên phải sửa rất nhiều. Tháng 7 năm 2026, Earth Species Project công bố một công cụ hỗ trợ. Tên của nó là alp-data. Nó gom hơn 35 nguồn dữ liệu tiếng động vật rải rác thành một bộ. Nó chỉnh tốc độ âm thanh và nhãn tên về cùng một định dạng. Nhà nghiên cứu bớt công sức sửa dữ liệu và tập trung vào việc huấn luyện [10].

Earth Species Project cũng đưa ra công cụ để xem dữ liệu bằng mắt. Đây là công cụ giúp xem nhanh dữ liệu âm thanh của nhiều loài. Việc gom và chọn dữ liệu như vậy là bước đầu để nuôi trí tuệ nhân tạo. Dữ liệu càng nhiều và càng tốt, trí tuệ nhân tạo càng thông minh.

Khi dữ liệu đã sẵn sàng, một trí tuệ nhân tạo lớn hơn ra đời. Đó là BioLingual do nhóm của David Robinson tạo ra. Trí tuệ nhân tạo này học tiếng của hàng trăm loài trên cạn và dưới biển. Dữ liệu học là các cặp gồm âm thanh và phần mô tả. Số cặp đó vượt quá 1 triệu [2].

BioLingual nối âm thanh với chữ viết. Nó ghép phần mô tả động vật trong bách khoa toàn thư với âm thanh ghi âm thật. Cách dùng ở đây gọi là học đối chiếu (contrastive learning). Đây là cách gom âm thanh và chữ có cùng nghĩa lại gần nhau trên bản đồ trong máy tính. Âm thanh và chữ khác nghĩa thì bị đẩy xa nhau.

Ý tưởng nối âm thanh với chữ này đến từ trí tuệ nhân tạo tên là CLAP. Các nhà nghiên cứu ở Carnegie Mellon University và Microsoft cùng tạo ra nó. CLAP ghép đủ loại âm thanh do con người tạo ra với phần chữ viết về âm thanh đó. BioLingual đưa ý tưởng này sang tiếng động vật [3].

Nhờ vậy, con người có thể tìm âm thanh bằng lời. Ta viết: "tiếng hót cao của chim robin gọi bạn tình vào mùa xuân". Khi đó, trí tuệ nhân tạo tìm âm thanh ấy trong vô số bản ghi. Không cần dạy riêng từ trước. Làm được ngay việc chưa từng học như vậy gọi là zero-shot. Đó là khả năng đoán đúng ngay, dù chưa được học riêng.

BioLingual làm nhiều việc trong một cơ thể. Nó phân biệt loài vật nào đã phát ra tiếng. Nó tìm chỗ có tiếng động vật trong bản ghi dài. Khi hỏi bằng chữ, nó tìm âm thanh phù hợp. Nếu học thêm một ít dữ liệu mới, kỹ năng của nó tăng thêm. Việc mà nếu là con người cần nhiều người chia nhau làm, nay một trí tuệ nhân tạo đảm nhận.

Âm thanh mà BioLingual học đến từ nhiều kho dữ liệu. Có Xeno-canto, nơi gom tiếng chim. Cũng có iNaturalist, nơi mọi người đăng ảnh và ghi chép về thiên nhiên. Đó là dữ liệu các nhà nghiên cứu ghi âm và tích lũy từng phần trong nhiều năm. Trí tuệ nhân tạo học bằng cách dùng các kho này làm nền.

Loại trí tuệ nhân tạo này giúp ích lớn cho nhà nghiên cứu. Trước đây, để tìm một âm thanh, họ phải nghe bản ghi suốt nhiều ngày. Giờ họ tìm bằng một dòng chữ ngắn. Tốc độ nghiên cứu âm thanh đã nhanh hơn rất nhiều [2, 3].

Tài liệu căn cứ liên quan

- [2] Robinson, D., Robinson, A., & Akrapongpisak, L. (2023). Transferable models for bioacoustics with human language supervision (BioLingual). arXiv:2308.04978. <https://arxiv.org/abs/2308.04978>
- [3] Elizalde, B., Deshmukh, S., Al Ismail, M., & Wang, H. (2023). CLAP: Learning audio concepts from natural language supervision. ICASSP 2023. arXiv:2206.04769. <https://arxiv.org/abs/2206.04769>

[10] Earth Species Project. (2026-07-22). Introducing alp-data: A shared data layer for animal language processing. [earthspecies.org.
https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/](https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/)

7.3 Trí tuệ nhân tạo nghe âm thanh và giải thích bằng lời

Earth Species Project là một nhóm nghiên cứu muốn dùng trí tuệ nhân tạo để giải mã lời nói của động vật. Trí tuệ nhân tạo do nhóm này tạo ra là NatureLM-audio. Trí tuệ nhân tạo này xử lý rộng rãi tiếng chim, thú trên cạn và thú biển [4].

NatureLM-audio nghe âm thanh và trả lời bằng lời. Khi con người cho nó nghe tiếng động vật, nó mô tả bằng chữ con vật đó đang làm gì. Nó đoán đó là loài nào. Nó cũng đếm trong bản ghi có mấy con vật. Nó phân biệt con phát ra tiếng là con non hay con trưởng thành. Nó tự làm được những việc này dù không được dạy riêng.

Trí tuệ nhân tạo này là trí tuệ nhân tạo nghe âm thanh và viết chữ. Nó không phải trí tuệ nhân tạo tạo ra âm thanh mới. Trước đây đã có năng lực hiểu lời người và trả lời bằng chữ. Nay năng lực đó chuyển sang tiếng động vật. Đây là mô hình ngôn ngữ âm thanh lớn đầu tiên được tạo cho động vật.

Năng lực của trí tuệ nhân tạo này được đo bằng bài kiểm tra. Tên bài kiểm tra là BEANS-Zero. Đây là bài kiểm tra chỉ chọn các câu hỏi chưa từng học. Trong đó, trí tuệ nhân tạo này phân biệt được loài. Nó cũng chia được loại tiếng kêu. Nó nghe âm thanh và giải thích bằng chữ chuyện gì đang xảy ra.

Trí tuệ nhân tạo học càng rộng thì càng mạnh trước việc lạ. Nó đạt kết quả tốt hơn trí tuệ nhân tạo chỉ học tiếng động vật. Đó là nhờ nó đã nghe cả tiếng người và đủ loại âm thanh môi trường. Trí tuệ nhân tạo nhỏ học hẹp thì lúng túng trước việc mới. Nền tảng phải rộng thì mới giải được vấn đề mới.

Earth Species Project dùng trí tuệ nhân tạo này cho nhiều nghiên cứu về động vật. Nó được dùng để nghiên cứu tiếng quạ. Nó cũng xem xét tiếng cá voi trắng. Nó xử lý cả tiếng voi. Một trí tuệ nhân tạo nghiên cứu nhiều loài vật cùng lúc. Ý tưởng xử lý tiếng của nhiều loài bằng một phương pháp như vậy gọi là xử lý ngôn ngữ động vật (animal language processing) [11].

Cách dùng trí tuệ nhân tạo này không khó. Ta cho nó nghe âm thanh đã ghi và đặt câu hỏi. Ta hỏi con vật nào đã phát ra âm thanh này. Khi đó, câu trả lời đến bằng chữ. Giống như hỏi bằng lời một công cụ biết âm thanh [4, 12].

Khi trí tuệ nhân tạo nghe âm thanh và trả lời bằng chữ, cần cẩn thận một điểm. Trí tuệ nhân tạo đôi khi bịa ra chuyện không có thật nhưng nghe có vẻ hợp lý. Nó cũng có thể nêu sai tên loài. Nó có thể viết rằng con vật đã làm hành động không xảy ra. Việc tạo ra câu trả lời không đúng sự thật như vậy gọi là ảo giác (hallucination).

Có một cách để giảm lỗi này. Đó là bắt trí tuệ nhân tạo tìm bản ghi thật trước khi trả lời. Nó chỉ được viết những nội dung đã kiểm tra bằng hồ sơ động vật thật. Cách này gọi là phương pháp tăng cường bằng truy xuất (retrieval-augmented). Nó giống như tìm sách trong thư viện để kiểm tra rồi mới nói.

Earth Species Project muốn giải mã lời nói của động vật để bảo vệ thiên nhiên. Khi hiểu âm thanh, ta biết sớm khó khăn mà động vật đang gặp. Ta có thể đọc trong âm thanh tín hiệu cho thấy một đàn đang giảm. Trí tuệ nhân tạo tìm ra tín hiệu đó nhanh hơn con người [4, 11].

Tài liệu căn cứ liên quan

[4] Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. ICLR 2025. arXiv:2411.07186. <https://arxiv.org/abs/2411.07186>

- [11] Earth Species Project. (2026-04-02). Animal language processing: An AI convergence in animal communication. [earthspecies.org. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/](https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/)
- [12] Earth Species Project. (2024-11-06). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. [earthspecies.org. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics](https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics)

7.4 Cách dịch không cần từ điển

Muốn dịch giữa hai ngôn ngữ, thường phải có từ điển. Phải có ai đó cho biết dog trong tiếng Anh là con chó trong tiếng Hàn. Nhưng lời nói của động vật không có từ điển. Vì chưa có ai ghi lại cá voi dùng âm thanh nào để nói điều gì. Làm sao có thể chuyển nghĩa khi không có từ điển [5]?

Năm 2017, các nhà khoa học máy tính đã tìm ra một cách. Đó là cách dịch dù không có một câu ví dụ ghép cặp nào giữa hai ngôn ngữ. Cách này gọi là dịch máy không giám sát (unsupervised machine translation). Nghĩa là trí tuệ nhân tạo tự dịch dù không được cho đáp án đúng.

Việc tạo câu ví dụ ghép cặp cho cá voi gần như là không thể. Vì không ai biết cá voi dùng âm thanh nào để nói điều gì. Muốn ghép cặp, trước hết phải biết nghĩa. Nhưng muốn biết nghĩa thì phải có bản dịch. Dịch không giám sát cho ta một manh mối để vượt qua ngõ cụt này.

Chìa khóa của phương pháp nằm ở hình dạng do các từ tạo ra. Khi đưa rất nhiều từ của một ngôn ngữ vào trí tuệ nhân tạo, nó vẽ một bản đồ. Những từ có nghĩa gần nhau tụ lại gần nhau. Khoảng cách và hướng giữa các từ hình thành quy tắc. Trên bản đồ, khoảng cách giữa nam và nữ cũng giống khoảng cách giữa vua và nữ hoàng. Bản đồ do các từ tạo ra như vậy gọi là không gian tiềm ẩn (latent space). Đó là bản đồ biểu thị nghĩa của từ bằng vị trí.

Ở đây xảy ra một điều kỳ lạ. Khi vẽ bản đồ của các ngôn ngữ khác nhau, hình dạng của chúng gần như giống nhau. Bản đồ tiếng Anh và bản đồ tiếng Phần Lan giống nhau. Vì não người chia thể giới theo cách tương tự. Vì vậy, nếu xoay một bản đồ để chồng lên bản đồ khác, các từ sẽ trùng nhau. Từ koira trong tiếng Phần Lan đặt xuống đúng chỗ của dog trong tiếng Anh. Như vậy, một từ điển dịch xuất hiện mà không cần từ điển có sẵn.

Các nhà nghiên cứu đang thử chuyển phương pháp này sang cá voi. Project CETI làm việc đó. Họ vẽ riêng bản đồ âm thanh của cá nhà táng và bản đồ âm thanh của cá heo mũi chai. Rồi họ nghiên cứu cách xoay và chồng hai bản đồ. Họ dùng toán học để xoay hai bản đồ sao cho khớp với nhau một cách mềm mại. Họ chưa làm rõ hết ý nghĩa. Hiện họ đang ở bước tìm cấu trúc âm thanh và xem chỗ nào có thể chồng bản đồ.

Project CETI đã tìm thấy trong tiếng cá nhà táng một mẫu giống nguyên âm của con người. Người làm rõ mẫu này là nhà ngôn ngữ học Gašper Beguš. Ông nghiên cứu tại University of California, Berkeley. Trong coda, tức mảnh âm thanh của cá nhà táng, có hai loại. Coda là nhóm tiếng lách cách ngắn mà cá voi phát ra khi trò chuyện. Một loại là coda a, loại kia là coda i. Như con người phân biệt a và i bằng hình dạng miệng, cá voi phân biệt bằng độ vang của âm thanh. Nghiên cứu này được đăng trên tạp chí của Royal Society năm 2026 [6].

Project CETI trước đó cũng đã làm rõ cấu trúc của tiếng cá nhà táng. Họ dùng trí tuệ nhân tạo để xem xét 8.700 coda của cá nhà táng. Kết quả cho thấy coda được ghép lại như các khối Lego. Một vài đặc điểm như tốc độ và nhịp được kết hợp theo nhiều cách. Họ ví cấu trúc này với chữ cái trong tiếng người và gọi là bảng chữ cái âm thanh. Tuy vậy, nghiên cứu chưa chỉ ra từng âm có nghĩa gì. Đây là bước cho thấy quy tắc của cấu trúc [7].

Vào tháng 3 cùng năm, Project CETI cũng công bố một phát hiện khác về cá nhà táng. Họ ghi lại rất kỹ khoảnh khắc một cá con được sinh ra. Khi đó, họ thấy nhiều con cùng chăm sóc cá con. Nghiên cứu này

cho thấy cách cá voi cùng nhau chăm con. Đây là một nghiên cứu riêng, khác với nghiên cứu tìm ra mẫu ghép âm thanh [13].

Việc tìm cấu trúc của âm thanh giúp ích cho việc dịch khi không có từ điển. Nếu tiếng cá voi có cấu trúc giống tiếng người, ta sẽ dễ tìm chỗ để đặt hai bản đồ lên nhau. Càng chia nhỏ âm thanh, bản đồ càng chi tiết. Những bản đồ chi tiết thì càng dễ chồng khớp với nhau [5, 6, 7].

Tài liệu căn cứ liên quan

- [5] Goldwasser, S., Gruber, D. F., Kalai, A. T., & Paradise, O. (2023). A theory of unsupervised translation motivated by understanding animal communication. NeurIPS 2023. arXiv:2211.11081. <https://arxiv.org/abs/2211.11081>
- [6] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B. <https://doi.org/10.1098/rspb.2025.2994>
- [7] Sharma, P., Gero, S., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [13] Project CETI / Harvard SEAS. (2026-03). Studies documenting rare sperm whale birth and ancient cooperative care released. EurekAlert! <https://www.eurekalert.org/news-releases/1120823>

7.5 Trí tuệ nhân tạo vượt qua loài và bức tường chưa vượt được

Đôi khi, trí tuệ nhân tạo chỉ học tiếng của một loài vật nhưng vẫn nhận ra tiếng của loài khác. Điều này được gọi là khả năng thích nghi vượt qua loài. Trong tiếng Anh, đó là khái quát hóa liên loài (cross-species generalization). Ví dụ, trí tuệ nhân tạo học tiếng chim nhưng vẫn phân loại được cả tiếng ếch [8, 9].

Ví dụ rõ nhất là Perch của Google. Perch 2.0 ra mắt năm 2025 chủ yếu học từ tiếng chim. Nó đã nghe rất nhiều âm thanh của chim và các động vật sống trên cạn. Số loài động vật trong dữ liệu này vượt quá 14.500 loài. Trí tuệ nhân tạo này hầu như không học âm thanh dưới biển. Nhưng nó vẫn nhận ra tiếng cá voi rất tốt. Nó đạt kết quả cao trong bộ phân loại được làm từ ít dữ liệu. Một trí tuệ nhân tạo học về chim đã nhận ra cá voi [8, 9] [8].

Cá voi không phải bức tường duy nhất mà Perch vượt qua. Perch học từ tiếng chim cũng xử lý được tiếng ếch. Bộ âm thanh tiếng ếch ở Amazon của Brazil được gọi là AnuraSet. Perch đã chia các tiếng ếch này theo từng loại. Nó làm được việc đó dù chưa từng học tiếng ếch [8, 9].

Những kết quả này là mặt sáng. Nhưng trí tuệ nhân tạo vẫn còn những bức tường lớn chưa vượt được.

Thứ nhất, dữ liệu quá ít. Trí tuệ nhân tạo học tiếng người đã lớn lên nhờ hàng trăm nghìn giờ ghi âm. Tiếng của động vật hoang dã, dù gom hết lại, vẫn ít hơn rất nhiều. Hơn nữa, máy ghi âm ngày xưa chỉ thu được những âm lớn. Nó bỏ lỡ những tiếng thì thầm nhỏ mà cá voi trao đổi khi ở gần nhau. Tiếng nhỏ này được gọi là soft click. Nghĩa là những âm quan trọng đã bị thiếu khỏi dữ liệu [6].

Soft click là tiếng nhỏ mà cá voi trao đổi với nhau khi ở gần. Nếu so với con người, nó giống như lời thì thầm vào tai. Tiếng nhỏ này có thể chứa ý nghĩa để chia sẻ cảm xúc. Vì máy ghi âm cũ bỏ lỡ tiếng này, nên những cuộc trò chuyện quan trọng đã bị thiếu.

Thứ hai, thế giới sống của chúng quá khác nhau. Cá nhà táng sống ở vùng biển sâu không có ánh sáng. Nó dùng túi dầu ở trán để phát siêu âm và nhìn xung quanh bằng âm thanh. Con người nhìn thế giới bằng mắt. Cá voi nhìn thế giới bằng âm thanh, còn con người nhìn bằng mắt, nên cách cảm nhận khác nhau. Vì vậy, hai bản đồ có thể không chồng khớp hoàn toàn.

Thế giới mà mỗi loài cảm nhận theo cách riêng được gọi là Umwelt. Từ này có nghĩa là thế giới xung quanh. Người tạo ra từ này là nhà sinh học Jakob von Uexküll. Nó có nghĩa là mỗi loài vật có một thế giới riêng, được cảm nhận bằng giác quan của chính loài đó. Nhà triết học Ludwig Wittgenstein cũng để lại

một câu nói tương tự. Ông nói rằng dù sư tử có thể nói, con người cũng không hiểu được. Đó là vì sư tử và con người sống trong những thế giới khác nhau.

Thứ ba, dùng bừa bãi thì nguy hiểm. Trí tuệ nhân tạo có thể bắt chước tiếng chim mái để gọi chim trống. Việc đó làm xáo trộn trật tự của loài chim khi tìm bạn đời và nuôi con. Hành vi dùng âm thanh giả để lừa động vật hoang dã được gọi là sự lừa dối sinh thái (ecological deception) [14].

Nguy hiểm còn đi xa hơn việc bắt chước. Trí tuệ nhân tạo có thể nhìn phản ứng của động vật rồi sửa âm thanh. Nếu động vật đáp lại, nó chỉnh âm thanh theo câu đáp đó. Thí nghiệm thay đổi âm thanh theo phản ứng của động vật trong thời gian thực được gọi là thí nghiệm âm thanh phản ứng. Nếu dùng đúng, nó trở thành chìa khóa để giao tiếp. Nếu dùng sai, nó trở thành công cụ lừa động vật.

Cá voi đã giữ văn hóa âm thanh riêng của mình trong hàng chục triệu năm. Âm thanh giả có thể làm biến dạng văn hóa lâu đời ấy. Một trật tự âm thanh đã bị rối thì khó đưa trở lại như cũ. Vì vậy, ta phải cẩn thận trước khi bắt chước.

Khi đưa trí tuệ nhân tạo lớn vào rừng và biển, cũng có vấn đề. Trí tuệ nhân tạo lớn thì nặng và dùng nhiều năng lượng. Máy ghi âm trong rừng khó chứa được thứ lớn như vậy. Vì thế, các nhà nghiên cứu chuyển phần lõi của trí tuệ nhân tạo lớn sang một hệ nhỏ hơn. Việc này giống như chỉ chuyển phần bên trong của chiếc bát lớn sang chiếc bát nhỏ.

Vì vậy, các nhà nghiên cứu nói rằng cần đặt ra nguyên tắc phải giữ. Đó là nguyên tắc không can thiệp bừa bãi vào tiếng của động vật. Họ gọi đây là nguyên tắc phải được giữ trước tiên. Trước khi bắt chước tiếng động vật, chúng ta cần học cách lắng lẽ lắng nghe trước.

Nếu giải được tiếng cá voi, luật pháp cũng có thể thay đổi. Năm 2026, đã có thảo luận về việc trao quyền pháp lý cho cá voi. Nếu con người biết qua âm thanh rằng động vật cảm thấy gì và trải qua điều gì, con người sẽ khó đối xử tùy tiện với động vật. Việc giải mã âm thanh gắn liền với việc bảo vệ động vật [15].

Trí tuệ nhân tạo đã mở ra con đường mới đi tới ngôn ngữ của động vật. Nó rút ngắn việc con người phải làm trong hàng trăm năm xuống còn vài năm. Nhưng con đường mở ra không có nghĩa là ta đã đến đích. Dữ liệu vẫn còn ít, và thế giới của động vật khác với thế giới của con người. Trí tuệ nhân tạo chỉ là công cụ. Dùng công cụ đó thế nào là việc của con người. Tấm lòng khiêm tốn lắng nghe phải đi trước công nghệ.

Tài liệu căn cứ liên quan

- [6] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B*. <https://doi.org/10.1098/rspb.2025.2994>
- [8] Burns, A., et al. (2025). Perch 2.0: The bittern lesson for bioacoustics. *arXiv:2512.03219*. <https://arxiv.org/abs/2512.03219>
- [9] Choi, C. Q. (2025). AI trained on birdsong can recognize whale calls. *IEEE Spectrum*. <https://spectrum.ieee.org/foundation-models-google-birds-whales>
- [14] Beguš, G. (2024). New patterns in animal communication with AI: Ethical and legal implications. *Interspecies Internet Lectures*.
- [15] Bolling, R. (2026-06-11). Decoding whale language could unlock new legal rights for them. *Boing Boing*. <https://boingboing.net/2026/06/11/decoding-whale-language-could-unlock-new-legal-rights-for-them.html>

Chương 8. Thế giới môi trường của động vật và ranh giới ký hiệu học

8.1 Mỗi loài có một thế giới riêng, Umwelt

Thế giới không chỉ có một. Mỗi sinh vật sống trong thế giới riêng của mình. Có một người đầu tiên nói rõ ý này. Đó là nhà sinh vật học Jakob von Uexküll, sinh ra ở Estonia. Ông là học giả cho rằng mỗi loài động vật cảm nhận thế giới theo một cách khác nhau [1].

Uexküll gọi thế giới này là thế giới môi trường. Trong tiếng Đức, nó được gọi là Umwelt. Umwelt là thế giới riêng mà một sinh vật cảm nhận và tạo ra bằng giác quan của mình. Con người nhìn màu bằng mắt và nghe âm thanh bằng tai. Nhưng không phải động vật nào cũng cảm nhận như con người.

Từ Umwelt không khó hiểu. Dù ở cùng một sân vận động, chó và người cảm nhận những điều khác nhau. Chó đọc thế giới bằng mùi, còn người đọc bằng mắt. Mỗi sinh vật xây nên thế giới bằng giác quan phù hợp với mình.

Hãy nhìn cá nhà táng (*Physeter macrocephalus*). Cá nhà táng lặn xuống biển sâu hàng trăm mét. Nơi đó là bóng tối không có một chút ánh sáng. Trong bóng tối, màu sắc không còn ý nghĩa. Thay vào đó, cá nhà táng phát ra âm thanh mạnh từ phần trán. Nó nghe âm thanh dội lại và đọc thế giới. Thế giới của cá nhà táng không được vẽ bằng màu sắc. Nó được vẽ bằng âm thanh [2].

Con người đã hiểu lầm trong một thời gian dài. Họ nghĩ rằng điều con người không cảm nhận được thì không tồn tại. Con người lấy giác quan của mình làm thước đo duy nhất. Thái độ này được gọi là chủ nghĩa lấy con người làm trung tâm. Đây là thái độ lấy giác quan và suy nghĩ của con người làm tiêu chuẩn của thế giới. Vì thái độ này, con người chỉ xem âm thanh của động vật là tiếng kêu hoặc phản ứng.

Có một ngành học nghiên cứu âm thanh và cử chỉ của động vật. Đó là ký hiệu học động vật (Zoosemiotics). Ký hiệu học động vật nghiên cứu ý nghĩa của các tín hiệu mà động vật trao đổi với nhau. Tín hiệu là âm thanh, mùi và cử chỉ chứa ý nghĩa. Học giả người Mỹ Thomas Sebeok đã đặt tên cho ngành học này. Ngành học này đứng trên ý tưởng Umwelt của Uexküll. Tín hiệu chứa ý nghĩa không chỉ có trong lời nói của con người. Mọi sinh vật đều trao đổi tín hiệu trong thế giới riêng của mình [3].

Ý tưởng của Uexküll dẫn đến một câu hỏi mới. Học giả người Ý Nicola Zengiaro đưa ra ý tưởng gọi là ký hiệu học động vật sinh thành. Ông nói rằng trí tuệ nhân tạo chỉ bắt chước tín hiệu của động vật bằng máy thì chưa thể hiểu được ý nghĩa của nó. Nghĩa là phải nhìn cùng lúc môi trường sống và cử chỉ của động vật thì mới đọc đúng ý nghĩa [4].

Bây giờ trí tuệ nhân tạo đang mở lại vấn đề này theo cách mới. Trí tuệ nhân tạo đọc bằng tính toán các vùng âm thanh rộng hơn con người. Những nơi con người từng nghĩ là im lặng trống rỗng lại đầy âm thanh. Tín hiệu đang qua lại trong biển, rừng và dưới đất. Trí tuệ nhân tạo giúp tìm ra các quy luật ẩn đó.

Vấn đề này cũng nối với xã hội loài người. Aza Raskin, người lập ra Earth Species Project, nói rằng công nghệ làm con người xa cách nhau. Con người cũng đã xa rời tự nhiên. Dùng trí tuệ nhân tạo để nghe âm thanh của động vật là việc nối lại sợi dây đã đứt này [5].

Tài liệu căn cứ liên quan

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press. <https://doi.org/10.2307/j.ctv2hnkcc3>

[2] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press. <https://doi.org/10.2307/j.ctv2hnkcc3>

[3] Sebeok, T. A. (1972). Perspectives in zoosemiotics. Mouton.

[4] Zengiaro, N. (2025). AI and animal communication: A generative zoosemiotics perspective. Linguistic Frontiers, 8(2). <https://doi.org/10.2478/lf-2025-0021>

[5] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>

8.2 Âm thanh và rung động ngoài giác quan con người

Tai người có giới hạn. Con người chỉ nghe được âm thanh từ 20 hertz đến 20.000 hertz. Hertz là đơn vị đếm số lần âm thanh rung trong 1 giây. Âm thanh cao hơn hoặc thấp hơn mức này thì con người không nghe được. Nhưng âm thanh đó không biến mất. Trí tuệ nhân tạo bắt lấy âm thanh này và cho con người thấy [1].

Trước hết hãy nhìn âm thanh của thực vật. Trong thời gian dài, người ta cho rằng thực vật không liên quan đến âm thanh. Nhưng nhóm nghiên cứu của Giáo sư Yossi Yovel và Giáo sư Lilach Hadany tại Đại học Tel Aviv ở Israel đã thay đổi suy nghĩ này. Cây cà chua và cây thuốc lá bị thiếu nước hoặc bị thương ở thân đã phát ra siêu âm. Siêu âm là âm thanh cao mà con người không nghe được. Âm thanh đó lan trong không khí với độ lớn gần bằng khi con người trò chuyện.

Vì sao thực vật phát ra âm thanh. Khi thiếu nước, những bọt khí nhỏ xuất hiện trong các đường nước rất mảnh bên trong thân cây. Khi các bọt này vỡ, chúng tạo ra tiếng tách. Tai người không nghe thấy, nhưng máy móc ghi lại được.

Người ta đưa âm thanh này vào trí tuệ nhân tạo học sâu. Học sâu là trí tuệ nhân tạo tự tìm ra đặc điểm của âm thanh. Khi đó, âm thanh do thiếu nước và âm thanh do bị thương được chia thành hai nhóm khác nhau. Hai âm thanh có dạng sóng khác rõ. Dạng sóng là hình dạng của âm thanh dao động như sóng nước. Những động vật nghe được siêu âm, như côn trùng, có thể nghe âm thanh này.

Thực vật cũng nghe âm thanh. Nhóm của Giáo sư Yovel và Giáo sư Hadany đã quan sát hoa anh thảo chiều. Họ bật âm thanh giống tiếng ong bay bên cạnh hoa. Khi đó, trong vòng 3 phút, hoa làm mật hoa ngọt hơn. Hoa không phản ứng với các tiếng ồn khác. Cánh hoa giống như đã làm nhiệm vụ của tai để gom âm thanh [2].

Nhóm nghiên cứu của Giáo sư Heidi Appel và Giáo sư Rex Cocroft tại Đại học Missouri ở Hoa Kỳ cũng làm một thí nghiệm tương tự. Họ cho cây Arabidopsis, một loài cỏ nhỏ, nghe nhiều âm thanh. Cây vẫn yên khi nghe tiếng mưa gió. Nhưng khi nghe rung động của sâu non gặm lá thì khác. Cây lập tức tạo ra chất xua đuổi côn trùng. Nghĩa là lá đã phân biệt rung động khi bị gặm với các rung động khác [3].

Một nghiên cứu năm 2026 đã đẩy sự thật này đi thêm một bước. Nhóm của Giáo sư Yovel và Giáo sư Hadany quan sát bướm đêm. Bướm đêm cái nghe âm thanh của thực vật khi chọn nơi đẻ trứng. Bướm đêm tránh âm thanh của cây bị thiếu nước và chọn cây khỏe. Đây là bằng chứng đầu tiên cho thấy động vật phản ứng với âm thanh của thực vật [4].

Thế giới dưới đất và thế giới của các sinh vật nhỏ cũng vậy. Ong mật múa hình số 8 trong tổ ong tối. Điệu múa này là tín hiệu cho biết hướng và khoảng cách đến hoa. Nhưng tín hiệu này không chỉ là điệu múa nhìn bằng mắt. Nó còn trộn cả rung động của cánh và rung động truyền qua tổ ong. Nhóm của Giáo sư Tim Landgraf đã đọc điệu múa này bằng trí tuệ nhân tạo. Robot và camera quan sát tổ ong trong hơn một tháng. Trí tuệ nhân tạo tự động giải từng điệu múa trong hình ảnh [5].

Chuột cũng nói bằng âm thanh mà con người không nghe được. Chuột phát siêu âm khi lo lắng hoặc khi vui. Nhóm của Kevin Coffey tại Đại học Washington đã tạo ra trí tuệ nhân tạo tên là DeepSqueak. Trí tuệ nhân tạo này chia nhỏ và đọc âm thanh của chuột. Nhờ đó, người ta có thể đọc trạng thái cảm xúc của chuột qua âm thanh [6].

Tai người không bắt được âm thanh của chuột. Khi sợ, chuột phát âm thanh khoảng 22 kilohertz. Khi vui, chuột phát âm thanh cao ở vùng 50 kilohertz. Kilohertz nghĩa là 1000 hertz. DeepSqueak biến âm thanh này thành hình ảnh và vẽ bản đồ cảm xúc.

Những phát hiện này cùng nói một điều. Không nghe được không có nghĩa là không có âm thanh. Đất, cỏ và các động vật nhỏ vẫn luôn trao đổi tín hiệu. Trí tuệ nhân tạo mở các âm thanh ẩn này ra trước mắt con người.

Tài liệu căn cứ liên quan

- [1] Khait, I., Lewin-Epstein, O., Yovel, Y., & Hadany, L. (2023). Sounds emitted by plants under stress are airborne and informative. *Cell*, 186(7), 1328-1336. <https://doi.org/10.1016/j.cell.2023.03.009>
- [2] Veits, M., et al. (2019). Flowers respond to pollinator sound within minutes by increasing nectar sugar concentration. *Ecology Letters*, 22(9), 1483-1492. <https://doi.org/10.1111/ele.13331>
- [3] Appel, H. M., & Cocco, R. B. (2014). Plants respond to leaf vibrations caused by insect herbivore chewing. *Oecologia*, 175(4), 1257-1266. <https://doi.org/10.1007/s00442-014-2995-6>
- [4] Seltzer, R., Zer Eshel, G., Hadany, L., & Yovel, Y. (2026). Female moths incorporate plant acoustic emissions into their oviposition decision-making process. *eLife*, 14, e104700. <https://doi.org/10.7554/eLife.104700>
- [5] Landgraf, T., et al. (2024). Autonomous tracking of honey bee behaviors over long-term periods with cooperating robots. *Science Robotics*, 9(95). <https://doi.org/10.1126/scirobotics.adn6848>
- [6] Coffey, K. R., Marx, R. G., & Neumaier, J. F. (2019). DeepSqueak: A deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology*, 44(5), 859-868. <https://doi.org/10.1038/s41386-018-0303-6>

8.3 Trí tuệ nhân tạo đọc nhiều giác quan cùng lúc

Động vật không nói chỉ bằng một âm thanh. Âm thanh, cử chỉ và mùi cùng hòa vào nhau để tạo nghĩa. Kiểu giao tiếp dùng nhiều đường truyền cùng lúc như vậy được gọi là đa phương thức (multimodal). Đó là giao tiếp trộn cả ánh mắt, cử chỉ và mùi, chứ không chỉ âm thanh [1].

Nếu chỉ nhìn một âm thanh, ta dễ bỏ lỡ ý nghĩa. Cùng một âm thanh cũng có thể đổi nghĩa theo cử chỉ. Con người cũng khác nhau khi nói cùng một câu trong lúc cười và trong lúc giận. Vì vậy phải nhìn nhiều tín hiệu cùng nhau thì ý nghĩa mới thông suốt.

Nghiên cứu về cá heo đốm ở vùng biển Bahamas là một ví dụ tốt. Tiến sĩ Denise Herzing và Giáo sư Thad Starner đã tạo ra một thiết bị gọi là CHAT. Thiết bị này giúp cá heo và con người trao đổi âm thanh với nhau. Nhưng chỉ âm thanh thì không thể giải được cuộc trò chuyện. Cần ghi lại cả khoảnh khắc cá heo nhìn vào mắt con người và góc mà cơ thể cá heo tạo ra. Phải nhìn nhiều tín hiệu cùng nhau thì ý nghĩa mới thông suốt.

Nghiên cứu về quạ cũng cho thấy nhiều tín hiệu rất quan trọng. Nhóm của Giáo sư John Marzluff tại Đại học Washington ở Hoa Kỳ đã quan sát quạ. Nhóm nghiên cứu đeo một chiếc mặt nạ đặc biệt, bắt quạ rồi thả ra. Sau đó, khi một người đeo cùng chiếc mặt nạ xuất hiện, quạ kêu lớn và kéo đến. Quạ đã nhớ khuôn mặt của người từng làm hại mình. Khi người đó đến, quạ phát tiếng cảnh báo để báo nguy hiểm cho đồng loại [2].

Trí nhớ này kéo dài lâu. Quạ không quên khuôn mặt đó trong hơn 2 năm. Chỉ nghe âm thanh thì khó tìm ra sự thật này. Phải nhìn cả âm thanh và hành động của quạ thì ý nghĩa mới được giải ra.

Earth Species Project đã tham gia để mở rộng các nghiên cứu như vậy. Tổ chức này là một nhóm nghiên cứu muốn dùng trí tuệ nhân tạo để giải tiếng nói của động vật. Họ nói rằng cần rời khỏi cách chỉ nhìn âm thanh. Vì có dữ liệu do cảm biến gắn trên cơ thể động vật thu thập. Cảm biến đo âm thanh, chuyển động, độ sâu dưới nước và những điều khác [3][4].

Cảm biến gắn trên cơ thể động vật đo được nhiều thứ. Nó thu khoảng cách bơi, độ sâu dưới nước và nhịp tim. Khi thêm dữ liệu này vào âm thanh, ý nghĩa trở nên rõ hơn. Ta có thể phân biệt đó là âm thanh khi đói hay âm thanh khi sợ. Trí tuệ nhân tạo đọc các dữ liệu khác nhau này cùng nhau theo thứ tự thời gian.

Trong một bài viết năm 2026, Earth Species Project giới thiệu một tên gọi mới. Đó là xử lý ngôn ngữ động vật (Animal Language Processing). Cụm từ này được dùng lần đầu trong một công bố năm 2024. Cách này không ép ngữ pháp của tiếng người lên động vật. Dù là âm thanh hay cử chỉ, nó xem nhiều tín hiệu là dữ liệu có quy luật và phân tích chúng. Đây là thái độ đọc thế giới của động vật đúng như nó vốn có, không lấy lời nói của con người làm thước đo.

Một kho dữ liệu giúp nghiên cứu này cũng đã xuất hiện. Năm 2026, Earth Species Project mở một công cụ tên là alp-data. Công cụ này gom hơn 35 bộ dữ liệu âm thanh của chim, cá voi và côn trùng vào một nơi. Hiện nay dữ liệu âm thanh là trung tâm, còn dữ liệu chuyển động và hình ảnh sẽ được thêm sau. Công cụ này được mở để ai cũng có thể nghiên cứu bằng cùng một dữ liệu. Nhờ vậy nghiên cứu mới tích lũy nhanh.

Tài liệu căn cứ liên quan

- [1] Herzing, D. L., Starner, T., et al. (2025). CHAT Jr: A wearable bioacoustics system for two-way communication research between humans and cetaceans. NeurIPS 2025. <https://neurips.cc/virtual/2025/131501>
- [2] Marzluff, J. M., Walls, J., Cornell, H. N., Withey, J. C., & Craig, D. P. (2010). Lasting recognition of threatening people by wild American crows. Animal Behaviour, 79(3), 699-707. <https://doi.org/10.1016/j.anbehav.2009.12.022>
- [3] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>
- [4] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. <https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>

8.4 Bảng chữ cái âm thanh của cá nhà táng và trí tuệ nhân tạo

Cá nhà táng nói bằng tiếng tách. Một nhóm nhiều tiếng tách ngắn được gọi là coda. Coda là cụm tiếng tách ngắn mà cá nhà táng trao đổi với nhau. Project CETI nghiên cứu các coda này. CETI là kế hoạch nghiên cứu nhằm dùng trí tuệ nhân tạo để giải tiếng nói của cá voi [1].

Năm 2024, CETI và nhóm nghiên cứu của Viện Công nghệ Massachusetts (MIT) ở Hoa Kỳ đã có một phát hiện lớn. Họ phân tích hơn 8700 mảnh âm thanh cá voi bằng trí tuệ nhân tạo. Khi đó, coda được cho thấy là một hệ thống âm thanh có quy luật. Nhóm nghiên cứu gọi quy luật này là bảng chữ cái âm thanh của cá nhà táng [2].

Bảng chữ cái này có bốn yếu tố. Thứ nhất là nhịp điệu, tức nhịp đều giữa các tiếng tách. Thứ hai là tốc độ chung, gọi là tempo. Thứ ba là rubato, tức việc tăng giảm tốc độ theo dòng trò chuyện. Thứ tư là âm trang trí, tức thêm một tiếng tách ở cuối. Khi trộn các yếu tố này, nhiều coda được tạo ra. Nhóm nghiên cứu tìm thấy hơn 143 tổ hợp thường xuất hiện. Cấu trúc này giống như lời nói của con người có phụ âm và nguyên âm.

Nghiên cứu năm 2025 tiến thêm một bước. Nhóm của Gašper Beguš tại Đại học Berkeley ở Hoa Kỳ và CETI đã xem xét coda thật kỹ. Trong tiếng tách có âm thanh giống nguyên âm của con người. Nhóm nghiên cứu tìm thấy hai nguyên âm cá voi khác nhau. Chúng là hai âm khác nhau, giống như âm a và âm i của con người. Điều đó có nghĩa là cá nhà táng chia âm thanh rất tinh vi giống con người [3].

Hình dạng của coda thay đổi theo tình huống trò chuyện. Con cá voi phía sau đổi âm thanh theo âm thanh do con cá voi phía trước phát ra. Điều này giống con người đáp lời nhau khi trò chuyện. Dòng chảy như vậy hiện rõ hơn nhờ nhiều dữ liệu và phân tích tính toán.

Project CETI cũng báo tin mới vào năm 2026. Nhóm nghiên cứu ghi lại kỹ cảnh cá nhà táng sinh con. Họ thu âm dưới nước hơn 6 giờ và thu cả hình ảnh quay từ trên không. Từ dữ liệu đó, người ta thấy cả đàn cùng giúp cá con ra đời. Đây là trường hợp đầu tiên dùng số liệu để chứng minh sự chăm sóc hợp tác như vậy ở động vật không thuộc nhóm người và khỉ [4].

Việc đổi tiếng cá voi thành tiếng người không dễ. Vì không có từ điển ghi song song hai ngôn ngữ. Nhóm của Beguš đang tìm một cách mới. Đó là cách tự tìm quy luật của âm thanh mà không có bảng đáp án, rồi đoán ý nghĩa [5].

Nghiên cứu như vậy rất khó làm nếu không có trí tuệ nhân tạo. Tai người khó tìm ra quy luật trong rất nhiều tiếng lách cách ấy. Trí tuệ nhân tạo so sánh vô số âm thanh rất nhanh và tìm ra quy luật ẩn bên trong.

Có một điều không được quên. Tìm ra quy luật của coda không có nghĩa là đã biết hết ý nghĩa. Biết cấu trúc của âm thanh khác với biết ý nghĩa. Cá nhà táng đang nói gì thì vẫn chưa được xác nhận. Nghiên cứu mới chỉ ở bước mở cánh cửa đầu tiên.

Tài liệu căn cứ liên quan

- [1] Project CETI. (2024, May 7). Sperm whale phonetic alphabet proposed for the first time. Project CETI. <https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time>
- [2] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [3] Beguš, G., Sprouse, R. L., Leban, A., Silva, M., & Gero, S. (2025). Vowel- and diphthong-like spectral patterns in sperm whale codas. *Open Mind*, 9. <https://doi.org/10.1162/OPMI.a.252>
- [4] Project CETI. (2026, March). Studies documenting rare sperm whale birth and ancient cooperative care released. *EurekaAlert!* <https://www.eurekaalert.org/news-releases/1120823>
- [5] Beguš, G., Leban, A., & Gero, S. (2026). Approaching an unknown communication system by latent space exploration and causal inference. *Royal Society Open Science*, 13(8), 250829. <https://doi.org/10.1098/rsos.250829>

8.5 Ranh giới ký hiệu học và ngoại giao giữa các loài

Trí tuệ nhân tạo mở ra cho con người một thế giới cảm giác rộng hơn. Vậy âm thanh của sự sống dưới biển được thu thập bằng gì. Con người dùng máy nghe dưới nước. Máy nghe dưới nước là micro bắt âm thanh trong nước. Trên cơ thể cá voi hoặc cá heo, người ta gắn thiết bị ghi nhỏ. Ở nhiều nơi trên biển, người ta thả mạng cảm biến phao để bắt âm thanh. Trí tuệ nhân tạo đọc những âm thanh được thu như vậy [1].

Cũng có ý tưởng về công nghệ sẽ xuất hiện trong tương lai. Đó là đề xuất dùng mạng truyền thông thế hệ thứ 6 sắp tới (6G) như một bộ cảm biến. 6G là mạng không dây nhanh hơn hiện nay rất nhiều. Ý tưởng là nếu đọc được sóng vô tuyến của mạng này khi chúng va vào vật thể, ta có thể biết chuyển động. Nhưng đây vẫn chỉ là ý tưởng ở giai đoạn đề xuất. Hơn nữa, sóng vô tuyến hầu như không đi qua được nước biển. Vì vậy, cách này được nghĩ để quan sát trên mặt đất và trong không khí. Nó không phù hợp với việc thu âm thanh dưới biển [2].

Ở đây xuất hiện một ranh giới quan trọng. Ranh giới ký hiệu học là đường phân chia xem ta được đọc tín hiệu và ý nghĩa đến đâu. Giả sử trí tuệ nhân tạo đã thu thập mọi tín hiệu của động vật. Nhưng như vậy có được ép các tín hiệu đó vào khuôn lời nói của con người không. Ví dụ, có được ghép từng âm thanh cá voi với nhạc của con người không. Cách ghép như vậy là áp suy nghĩ của con người lên động vật [3].

Hãy lấy ví dụ về cách ghép sai. Có người có thể nói rằng trong 10 con cá heo, 3 con dùng danh từ. Có người cũng có thể ghép từng tiếng lách cách của cá voi với nhạc của con người. Nhưng đó là ép khuôn của con người lên chúng. Âm thanh của động vật phải được đọc trong thế giới của chính động vật.

Việc giải mã âm thanh đi kèm vấn đề đạo đức. Các học giả nói rằng con người không được tùy tiện bắt chước lời của động vật. Nếu dùng sai, việc đó có thể dùng để lừa hoặc quấy rối động vật. Vì vậy, họ nói cần đặt ra quy tắc trước [4].

Ý tưởng Umwelt của Uexküll đến nay vẫn là kim chỉ nam. Động vật và thực vật không sống để trở thành từ dịch trong từ điển của con người. Tín hiệu của chúng là văn hóa riêng đã kéo dài qua thời gian rất lâu. Các nhà nghiên cứu hiện đại tiếp nối ý tưởng của Uexküll và nói rằng không được dùng trí tuệ nhân tạo để tùy tiện bắt chước động vật. Mục tiêu của trí tuệ nhân tạo không phải là tạo ra đồ chơi động vật biết nói.

Có thí nghiệm ghi âm tiếng động vật rồi phát lại cho chúng nghe. Việc này được gọi là phát lại (playback). Đây là thí nghiệm dùng loa phát âm thanh đã ghi và quan sát động vật phản ứng thế nào. Nếu dùng công cụ này tùy tiện, ta có thể làm xáo trộn thế giới của động vật. Con vật bị giật mình vì âm thanh phát lại có thể nhầm lẫn tín hiệu thật với tín hiệu giả. Vì vậy, các nhà nghiên cứu nói rằng phải cẩn thận.

Con đường chúng ta cần đi không phải là chinh phục, mà là lắng nghe. Đó là việc con người bước xuống khỏi vị trí người diễn giải duy nhất của Trái Đất. Đó là việc bảo vệ để âm thanh của động vật tiếp tục vang lên lâu dài trên đất liền và trong biển của chúng. Các nhà nghiên cứu gọi điều này là ngoại giao giữa các loài (interspecies diplomacy). Ngoại giao giữa các loài là lời hứa để con người và các sự sống khác cùng tồn tại mà không làm hại nhau [5].

Lắng nghe không dừng lại ở việc im lặng nghe. Nó còn nối tiếp đến việc bảo vệ biển và rừng nơi động vật sống. Muốn bảo vệ âm thanh, ta phải bảo vệ cả nơi phát ra âm thanh. Đó là ý nghĩa thật sự của ngoại giao giữa các loài.

Tài liệu căn cứ liên quan

- [1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press. <https://doi.org/10.2307/j.ctv2hnkcc3>
- [2] Ericsson. (2025). Integrated sensing and communication (ISAC) explained. Ericsson. <https://www.ericsson.com/en/6g/isac>
- [3] Zengiaro, N. (2025). AI and animal communication: A generative zoosemiotics perspective. Linguistic Frontiers, 8(2). <https://doi.org/10.2478/lf-2025-0021>
- [4] Alcantara, L., & Andrews, K. (2026). Can we talk to the animals? The ethics of using machine learning to decode animal communication. Topoi. <https://doi.org/10.1007/s11245-026-10409-2>
- [5] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>

Chương 9. Ngã rẽ giữa lừa dối và bảo tồn

9.1 Sự rối loạn do âm thanh giả gây ra

Giờ đây, trí tuệ nhân tạo có thể bắt chước âm thanh của động vật rất giống thật. Nó tạo ra tiếng chim hót, bài hát của cá voi, và tiếng kêu của voi. Giống như con người học nói, trí tuệ nhân tạo cũng nghe âm thanh và học quy luật. Sau khi học như vậy, nó tự tạo ra tiếng chim. Công nghệ này có thể dùng để giúp động vật. Nhưng nếu dùng sai, nó sẽ làm thiên nhiên rối loạn nặng [1].

Trong nghiên cứu động vật có một phương pháp gọi là 'phát lại(playback)'. Đó là thí nghiệm cho động vật nghe lại âm thanh đã ghi và quan sát phản ứng của chúng. Trước đây, người ta chỉ dùng âm thanh thật đã ghi lại. Bản ghi cũ có lẫn tiếng ồn, nên động vật đôi khi nhận ra. Bây giờ trí tuệ nhân tạo tạo ra âm thanh giả rất trơn tru. Vì vậy động vật dễ bị lừa hơn. Âm thanh giả này làm động vật nhầm lẫn trong việc giao phối và chăm sóc con non [7].

Các học giả ở nhiều nước đã cảnh báo nguy cơ này. Giáo sư Christian Rutz và các đồng nghiệp là những người tiêu biểu. Họ cho rằng âm thanh giả làm thay đổi hành vi của động vật hoang dã. Việc lừa động vật bằng âm thanh như vậy được gọi là 'lừa dối sinh thái'. Nghĩa là lừa động vật trong tự nhiên để làm chúng đổi hành vi. Chim nghe tiếng giao phối giả sẽ đi tìm bạn đời vô ích. Khi đó nó mất sức và thời gian, rồi bị mất cả lúc chăm sóc con non.

Giáo sư Karen Bakker cũng đã cảnh báo nguy cơ này từ lâu. Giáo sư Bakker là học giả dẫn đầu nghiên cứu dùng công nghệ số để nghe âm thanh của động vật. Bà đưa ý nghĩa đó vào cuốn sách 'Âm thanh của sự sống(The Sounds of Life)' xuất bản năm 2022. Giáo sư Bakker qua đời năm 2023. Lời cảnh báo bà để lại vẫn còn tiếp tục đến nay [2].

Con người thường gán cảm xúc của mình vào âm thanh của động vật. Ta dễ nghĩ bài hát của cá voi là bài hát tình yêu. Nhưng chưa ai biết ý nghĩa thật của âm thanh đó. Nếu đoán bừa, ta sẽ sai. Các học giả nói rằng cũng phải cẩn thận với thói quen này.

Có một câu chuyện về tiến sĩ Joyce Poole, người nghiên cứu voi trong thời gian dài. Chuyện xảy ra ở Amboseli, Kenya. Nhóm nghiên cứu cho đàn voi nghe tiếng kêu của một voi mẹ đã chết. Những con voi nghe âm thanh nhớ thương ấy và chạy khắp nơi để tìm. Trong số đó, con gái của voi mẹ đã chết gọi mẹ nhiều lần và chờ câu trả lời. Nhưng chủ nhân của âm thanh ấy không ở đâu cả. Voi nhớ cả tiếng của người thân đã chết. Voi nói chuyện với nhau từ xa bằng âm trầm mà tai người khó nghe rõ. Nhờ vậy, chúng nhớ riêng tiếng của hàng chục con voi. Chỉ một âm thanh đã làm cả đàn dao động mạnh [3][4].

Cá voi cũng cần được cẩn thận. Cá voi lưng gù là loài cá voi biết hát. Khi một bài hát mới xuất hiện ở một vùng biển, nó lan ra như bài hát thịnh hành. Bài hát ở vùng biển trước Australia có khi lan tới tận bên kia Nam Thái Bình Dương. Bài hát này lan chậm trong vài tháng đến vài năm. Nó đi chậm qua khoảng 6,000 kilômét. Đó là quãng đường rất xa, như từ Seoul đến một đất nước rất xa. Cá voi con cũng nghe bài hát của cá voi trưởng thành rồi học theo. Điều đó giống cách bài hát thịnh hành lan qua điện thoại thông minh. Việc học bài hát của nhau như vậy rất hiếm ở loài ngoài con người. Nếu phát bài hát cá voi giả do trí tuệ nhân tạo tạo ra xuống biển, vấn đề lớn sẽ xảy ra. Truyền thống bài hát đã kéo dài lâu đời có thể bị rối tung. Như vậy là làm hỏng văn hóa của chúng trước khi giải được ngôn ngữ của động vật [5].

Vì vậy các nhà nghiên cứu cũng làm thí nghiệm này một cách thận trọng. Họ chỉ làm ngắn khi thật cần thiết. Khi thí nghiệm kết thúc, họ xem động vật có ổn không. Cùng một công nghệ, tùy cách dùng mà có thể giúp ích hoặc gây hại.

Tài liệu căn cứ liên quan

- [1] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. *Science*, 381(6654), 152-155. <https://doi.org/10.1126/science.adg7314>
- [2] Bakker, K. (2022). *The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants*. Princeton University Press.
- [3] Poole, J. (1996). *Coming of age with elephants: A memoir*. Hyperion.
- [4] McComb, K., Moss, C., Sayialel, S., & Baker, L. (2000). Unusually extensive networks of vocal recognition in African elephants. *Animal Behaviour*, 59(6), 1103-1109. <https://doi.org/10.1006/anbe.2000.1406>
- [5] Garland, E. C., Goldizen, A. W., Rekdahl, M. L., et al. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current Biology*, 21(8), 687-691. <https://doi.org/10.1016/j.cub.2011.03.019>
- [7] Kückkuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. *AI and Ethics*, 5, 6379-6391. <https://doi.org/10.1007/s43681-025-00828-z>

9.2 Mạng giám sát ở khắp nơi và đời sống riêng của động vật

Ngày xưa, rừng sâu và biển xa là nơi ẩn náu của động vật. Mắt và tai của con người không chạm tới đó. Bây giờ thì khác. Những cảm biến âm thanh nhỏ và thông minh được đặt ở nhiều nơi trong rừng và biển. Thiết bị này nghe âm thanh suốt 24 giờ. Trí tuệ nhân tạo phân tích ngay các âm thanh đó. Thiết bị ghi âm tự động đặt trong rừng và biển như vậy được gọi là 'mạng quan sát âm thanh thụ động' [6].

Giờ đây, động vật hoang dã ở trước những con mắt theo dõi ở khắp nơi. Như thể những tháp canh vô hình luôn bật sáng. Aza Raskin, người đồng sáng lập Earth Species Project, đã chỉ ra thay đổi này. Ông nói nhờ trí tuệ nhân tạo, con người có thể theo dõi rừng và biển không ngừng nghỉ. Đồng thời, ông cảnh báo rằng cần đặt ra quy tắc để sức mạnh này không trở thành sự giám sát siết chặt động vật [8].

Thiết bị này cũng được dùng cho việc tốt. Cá voi thường mất mạng vì va vào tàu lớn. Người ta dùng cảm biến âm thanh để tìm nơi có cá voi rồi đổi tuyến tàu. Như vậy tàu lớn không va vào cá voi. Thiết bị cũng phát hiện tiếng cựa của người lén chặt cây trong rừng. Đó là cách bảo vệ động vật đang biến mất và ngăn săn trộm.

Nhưng nếu rơi vào tay người xấu, nó trở thành vũ khí. Điều này giống như con dao vừa có thể nấu ăn, vừa có thể làm người khác bị thương. Người ta cũng có thể bắt chước giả tiếng giao phối hoặc tiếng thức ăn để dụ động vật ra. Việc này chưa xảy ra rộng rãi, nhưng các học giả lo về nguy cơ đó. Công nghệ tạo ra để giúp động vật có thể bị dùng để bắt động vật [7].

Đời sống riêng của động vật cũng là vấn đề. Đời sống riêng là vùng của riêng mình mà ta không muốn cho người khác thấy. Giáo sư Karen Bakker nói rằng phải xem cá voi và cá heo là những tồn tại có quyền. Có khi con người nghe lén cả những âm thanh khế giữa mẹ và con non. Họ còn tự ý gom các âm thanh đó rồi mở ra cho ai cũng dùng. Các học giả cho rằng phải xem việc này là xâm phạm đời sống riêng. Nó giống như người khác lén đọc nhật ký của ta rồi đăng lên Internet [2].

Người bản địa đã gìn giữ rừng và biển từ lâu cũng có quyền của họ. Có khi công ty lớn lén lấy âm thanh phát ra từ đất của họ. Họ dùng âm thanh đó để dạy trí tuệ nhân tạo. Dữ liệu âm thanh cũng là tài sản của vùng đất ấy. Không được lấy khi chưa có phép của chủ nhân. Quyền để người địa phương tự quyết định dữ liệu phát sinh từ đất của họ được gọi là 'chủ quyền dữ liệu của người bản địa'.

Trên Internet cũng có chuyện kể rằng động vật chống lại sự giám sát. Có chuyện nói một đàn cá nhà táng đồng loạt ngừng phát âm thanh. Cũng có chuyện nói quạ tạo ra tiếng chim để làm trí tuệ nhân tạo nhầm lẫn. Nhưng những việc này chưa được khoa học xác nhận. Chúng chỉ là chuyện trong video trên Internet, nên không được tin như sự thật.

Cách dùng tốt và cách dùng xấu đều xuất phát từ cùng một công nghệ. Vì vậy điều quan trọng là ai dùng nó và dùng với ý định gì. Động vật cũng có không gian riêng cần được bảo vệ.

Tài liệu căn cứ liên quan

- [2] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.
- [6] Earth Species Project. (2024). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. <https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics>
- [7] Kükükuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. AI and Ethics, 5, 6379-6391. <https://doi.org/10.1007/s43681-025-00828-z>
- [8] Faunalytics (Hanes, A.). (2025, October 30). The ethics of using AI to talk to animals. Faunalytics. <https://faunalytics.org/the-ethics-of-using-ai-to-talk-to-animals/>

9.3 Dấu hiệu của con người giấu trong âm thanh: thủy vân âm thanh

Cần có cách nhận ra âm thanh giả. Các học giả và kỹ sư đã đưa ra một cách. Đó là giấu một 'dấu hiệu' trong âm thanh do trí tuệ nhân tạo tạo ra. Cách này được gọi là 'thủy vân âm thanh(audio watermarking)' [9].

Thủy vân là công nghệ khắc dấu hiệu không nhìn thấy vào trong tệp âm thanh. Dấu hiệu này là tín hiệu máy tính ẩn trong tệp âm thanh. Con người khó nhận ra tín hiệu này. Nhưng chương trình kiểm tra đọc được tín hiệu đó ngay. SynthID do Google tạo ra là một ví dụ về dấu hiệu như vậy. Dấu hiệu này được rải ở nhiều chỗ trong âm thanh. Vì vậy rất khó cắt bỏ hoặc xóa đi. Dù giảm âm lượng hoặc đổi tốc độ, dấu hiệu vẫn còn.

Dấu hiệu này cho biết nguồn gốc của âm thanh. Máy có thể kiểm tra âm thanh nào là âm thanh giả do con người tạo ra. Khi đó ta có thể tách âm thanh thật của động vật và âm thanh giả để xử lý riêng. Tách thật và giả là bước đầu để bảo vệ thiên nhiên. Nếu không có dấu hiệu, rất khó biết cái gì là thật [1].

Dấu hiệu này cũng có thể dùng để bảo vệ dữ liệu nghiên cứu. Trong rừng và biển có hàng nghìn thiết bị thu âm. Nếu âm thanh giả lẫn vào đó, kết quả nghiên cứu sẽ rối loạn. Nếu đếm lần tiếng chim thật và tiếng chim do máy tạo ra, câu trả lời sẽ sai. Có dấu hiệu thì có thể lọc âm thanh giả. Khi đó ta có thể tin và dùng dữ liệu nghiên cứu [7].

Cũng có ý kiến dùng cách này để ngăn săn trộm. Đó là cách đưa dấu hiệu vào âm thanh dùng cho nghiên cứu để xác nhận âm thanh đến từ đâu. Nó giúp tách âm thanh nghiên cứu được công nhận khỏi âm thanh giả không rõ nguồn gốc. Tuy nhiên, cách này có giới hạn. Nếu người xấu tạo âm thanh không có dấu hiệu, ta không thể lọc ra. Nếu họ dùng âm thanh đã xóa dấu hiệu, rất khó phát hiện. Vì vậy chỉ một công nghệ không thể chặn hết săn trộm [8].

Cách này vẫn còn ở giai đoạn đề xuất. Đó là ý tưởng tốt, nhưng không tự nó thành công. Nó chỉ có sức mạnh khi mọi quốc gia và công ty hứa sẽ đưa dấu hiệu vào âm thanh. Nếu chỉ một nơi phát âm thanh giả không có dấu hiệu, sẽ xuất hiện lỗ hổng. Giống như chỉ cần một lỗ trên lưới, cá cũng thoát ra được. Vì vậy chỉ công nghệ là chưa đủ. Cần có quy tắc để cùng nhau giữ.

Công nghệ này cũng có thể dùng cho các âm thanh khác do con người tạo ra. Đó là đưa dấu hiệu vào giọng nói hoặc bài hát do trí tuệ nhân tạo tạo ra. Khi đó có thể giảm việc bị giọng nói giả lừa. Cách làm tạo ra để bảo vệ động vật cũng bảo vệ con người. Ngày nay, bài hát và giọng nói do trí tuệ nhân tạo tạo ra tràn ngập Internet. Vì vậy dấu hiệu như thế sẽ còn có ích hơn trong tương lai.

Tài liệu căn cứ liên quan

- [1] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. <https://doi.org/10.1126/science.adg7314>
- [7] Kükükuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. AI and Ethics, 5, 6379-6391. <https://doi.org/10.1007/s43681-025-00828-z>

[8] Faunalytics (Hanes, A.). (2025, October 30). The ethics of using AI to talk to animals. Faunalytics.

<https://faunalytics.org/the-ethics-of-using-ai-to-talk-to-animals/>

[9] Google DeepMind. (2024). SynthID: Identifying AI-generated content. <https://deepmind.google/models/synthid/>

9.4 Tạo quy tắc giữa con người và động vật

Việc xử lý âm thanh của động vật bằng trí tuệ nhân tạo cần có quy tắc. Vì người có sức mạnh không được tùy ý làm thiên nhiên dao động. Âm thanh vượt qua biên giới và lan đi xa. Vì vậy chỉ một quốc gia giữ quy tắc là chưa đủ. Các học giả nói rằng nhiều quốc gia nên cùng tạo ra quy tắc để cùng tuân theo. Đó là lời hứa cần giữ giữa con người và động vật [10].

Ở Trường Luật Đại học New York, Hoa Kỳ, có một nhóm nghiên cứu tên là 'Sự sống vượt lên trên con người (More Than Human Life)'. Gọi tắt là MOTH. Nhóm này được thành lập năm 2022. Các học giả luật và nhà khoa học cùng tụ họp. Nhóm này nói rằng phải xem động vật không phải là dữ liệu, mà là những tồn tại có đời sống riêng. Và nhóm đã đưa ra quy tắc gồm bốn lời hứa. Đó là lời hứa chuẩn bị, cùng bàn bạc, ngăn tổn hại, và bảo vệ.

Thứ nhất là quy tắc xem xét trước. Có thí nghiệm trong đó trí tuệ nhân tạo và động vật hoang dã trao đổi âm thanh. Trước khi làm thí nghiệm như vậy, phải kiểm tra nguy cơ trước. Phải xác nhận rằng nó không làm hại văn hóa giao phối và chăm sóc con non. Chỉ sau đó mới cho phép thí nghiệm. Điều này giống như phải kiểm tra an toàn trước khi bắt đầu công trình.

Thứ hai là quy tắc ngăn việc dùng tùy tiện. Công nghệ bắt chước âm thanh của động vật không được để ai cũng dùng. Nhóm nói rằng công nghệ dùng cho săn bắn phải bị cấm hẳn. Chỉ những nơi đã qua kiểm tra mới được phép dùng một cách thận trọng. Điều này giống như không để dụng cụ nguy hiểm ở bất cứ đâu. Nhóm cũng nói rằng cần ngăn việc dùng công nghệ cốt lõi tùy tiện để kiếm tiền.

Thứ ba là quy tắc ngăn tin tức phóng đại. Có những tin thổi phồng rằng "trí tuệ nhân tạo đã tạo ra máy dịch có thể trò chuyện tự do với động vật". Các học giả cho rằng những tin như vậy có hại. Vì chúng làm con người xem động vật như đồ chơi. Vì vậy họ nói rằng cần thông báo trung thực những việc công nghệ vẫn chưa làm được [11].

Quy tắc không phải để ngăn nghiên cứu. Nó là hàng rào giúp nghiên cứu đi đúng hướng. Giống như sân chơi phải có vạch kẻ thì trẻ em mới yên tâm chạy chơi. Để tạo quy tắc, cần tiếng nói của nhiều người. Không chỉ có nhà khoa học. Người hiểu luật, người bản địa sống trong rừng và biển, và người yêu quý động vật phải cùng quyết định. Khi có nhiều con mắt cùng nhìn, chỗ bị bỏ sót sẽ giảm. Mục tiêu của quy tắc chỉ có một. Đó không phải là thống trị động vật, mà là tôn trọng chúng [7][8].

Tài liệu căn cứ liên quan

[7] Kűcűkuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. *AI and Ethics*, 5, 6379-6391. <https://doi.org/10.1007/s43681-025-00828-z>

[8] Faunalytics (Hanes, A.). (2025, October 30). The ethics of using AI to talk to animals. Faunalytics. <https://faunalytics.org/the-ethics-of-using-ai-to-talk-to-animals/>

[10] Rodríguez-Garavito, C. (2025). Listening to the more-than-human world: Legal and ethical principles for nonhuman animal communication technologies. *More-Than-Human Life (MOTH) Program*, NYU School of Law. <https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai>

[11] Kluskens, C. (2025). The responsible media representation of artificial intelligence for non-human animal communication. *NeurIPS 2025 Workshop on AI for Non-Human Animal Communication*.

9.5 Tin mới nhất và ngã rẽ chúng ta đang đứng

Năm 2026, nghiên cứu ngôn ngữ động vật vẫn tiến nhanh. Earth Species Project đã gọi lĩnh vực này là 'Xử lý ngôn ngữ động vật (Animal Language Processing)'. Đây là cách gom nhiều âm thanh của động vật rồi phân tích, giống như máy tính hiểu lời nói của con người. Cách này rời khỏi cách nhìn riêng từng loài. Đó là cách xem xét âm thanh của nhiều loài động vật cùng một lúc. Tên gọi này không phải được đặt lần đầu, mà là gom các nghiên cứu rải rác lại thành một nhóm. Công bố này xuất hiện vào tháng 4 năm 2026 [12].

Vào tháng 7 cùng năm, một kho dữ liệu mới cũng được mở. Tên của nó là 'alp-data'. Đây là nơi gom dữ liệu âm thanh động vật để nhiều nhà nghiên cứu cùng dùng. Nó chứa hơn 35 bộ dữ liệu. Khi chia sẻ dữ liệu, nghiên cứu sẽ nhanh hơn. Điều này giống như bạn trong lớp chia sẻ đồ dùng học tập thì công việc kết thúc nhanh hơn [13].

Nền tảng của nghiên cứu này là 'NatureLM-audio'. Đây là một trí tuệ nhân tạo lớn được làm cho âm thanh của động vật. Trí tuệ nhân tạo này chỉ nghe âm thanh rồi đoán đó là loài động vật nào. Nó cũng đếm được trong âm thanh có bao nhiêu con vật. Nó còn thử phân biệt giai đoạn trưởng thành, như chim non hay chim đã lớn. Nó cho thấy khả năng nhận ra cả âm thanh của loài chưa từng học. Trí tuệ nhân tạo như vậy làm trong vài phút việc mà con người phải mất nhiều ngày. Vì thế, các nhà nghiên cứu có thể xem xét âm thanh nhanh hơn nhiều [6].

Nghiên cứu về cá voi cũng tiến lên. Project CETI đã thu thập coda, tức những mảnh âm thanh của cá nhà táng. Coda là một nhóm tiếng nhấp nối tiếp nhau như "tách, tách, tách". Nhóm nghiên cứu đã phân tích hơn 8,700 coda. Kết quả là họ tìm ra các mảnh cơ bản tạo nên âm thanh. Các nhà nghiên cứu chia các mảnh cơ bản này thành bốn loại. Họ gọi chúng là 'chữ cái âm thanh của cá nhà táng'. Điều này giống như việc chia phụ âm và nguyên âm trong chữ viết của con người. Cá voi thay đổi ý nghĩa bằng cách đổi tốc độ và nhịp của âm thanh. Chúng cũng đổi ý nghĩa bằng cách thêm tiếng nhấp. Các học giả còn tìm thấy quy tắc ghép âm thanh, giống như nguyên âm của con người. Điều đó không có nghĩa là cá voi phát ra âm 'a' hay 'i' như con người. Đây là một cách ví von. Giống như con người ghép phụ âm và nguyên âm để tạo nhiều từ, cá voi cũng ghép âm thanh theo quy tắc [14][15][16].

Trí tuệ nhân tạo cũng được dùng để bảo vệ các loài động vật đang biến mất. Quạ Hawaii là một ví dụ. Loài chim này đã sống trong nơi nuôi nhốt lâu năm rồi trở lại tự nhiên. Trong thời gian đó, các nhà nghiên cứu xem chúng có quên một số âm thanh của đàn hay không. Họ dùng trí tuệ nhân tạo để so sánh âm thanh cũ với âm thanh hiện nay. Mục đích là giúp chúng tìm lại những âm thanh đã quên. Earth Species Project cũng nghiên cứu âm thanh của quạ, cá voi trắng và voi. Khi xem xét nhiều loài động vật cùng lúc, ta có thể thấy các quy tắc chung của âm thanh.

Bây giờ chúng ta đang đứng ở ngã rẽ. Một con đường là lừa dối. Đó là con đường dùng âm thanh giả để đánh lừa và làm động vật rối loạn. Con đường còn lại là bảo tồn. Đó là con đường lắng lẽ lắng nghe để bảo vệ động vật. Hai con đường này nằm trên cùng một công nghệ. Đi theo con đường nào tùy thuộc vào lựa chọn của con người. Trí tuệ nhân tạo chỉ là công cụ. Sức nghe âm thanh càng lớn, quy tắc bảo vệ sức mạnh ấy cũng phải càng lớn. Khi sức mạnh tăng lên, ta cũng phải học cách dùng sức mạnh đó. Khi chúng ta bỏ xuống sự kiêu ngạo và mở tai lắng nghe, con người và động vật có thể sống bên nhau [10][1].

Tài liệu căn cứ liên quan

- [1] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. *Science*, 381(6654), 152-155. <https://doi.org/10.1126/science.adg7314>
- [6] Earth Species Project. (2024). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. <https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics>

- [10] Rodríguez-Garavito, C. (2025). Listening to the more-than-human world: Legal and ethical principles for nonhuman animal communication technologies. More-Than-Human Life (MOTH) Program, NYU School of Law.
<https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai>
- [12] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication.
<https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>
- [13] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing.
<https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>
- [14] Project CETI. (2024). Sperm whale phonetic alphabet proposed for the first time.
<https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time>
- [15] Sharma, P., Gero, S., Payne, R., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [16] National Geographic. (2024). We're one step closer to understanding the sperm whale 'alphabet'.
<https://www.nationalgeographic.com/animals/article/sperm-whale-vowel-patterns-communication-discovery>

Chương 10. Nhân vị phi con người và quyền được máy móc hỗ trợ

10.1 Từ im lặng đến lời chứng

Luật của con người từ lâu chỉ trao quyền cho con người. Luật cũng trao quyền cho các tổ chức do con người lập ra, như công ty. Những tổ chức như vậy được gọi là pháp nhân(法人). Pháp nhân là con người giả do luật tạo ra. Công ty và trường học thuộc loại này. Pháp nhân không phải là người. Nhưng trước pháp luật, pháp nhân được đối xử như người [3].

Luật ngày xưa tin vào một điều. Đó là chỉ bên biết nghĩa vụ của mình và có thể lập cam kết mới có quyền. Động vật không thể tự bào chữa bằng lời nói. Vì vậy luật không xem động vật là chủ thể của quyền. Luật xem động vật là đồ vật và tài sản mà con người có thể mua bán.

Bức tường cũ này nay đang lung lay. Vì trí tuệ nhân tạo và âm sinh học đã kết hợp với nhau. Âm sinh học là ngành nghiên cứu âm thanh do động vật tạo ra. Hai công nghệ gặp nhau và bắt đầu cho thấy nội tâm của động vật bằng con số [2].

Ngày xưa, con người chỉ nghe tiếng động vật bằng tai. Vì vậy có nhiều âm thanh bị bỏ lỡ. Ngày nay, người ta đặt nhiều máy ghi âm dưới biển trên diện rộng. Project CETI(Cetacean Translation Initiative) làm việc này. Dịch sang tiếng Việt, đó là kế hoạch dịch ngôn ngữ cá voi. Kế hoạch này đã thu thập âm thanh của cá nhà táng trong thời gian dài. Trí tuệ nhân tạo xem xét từng âm thanh đã thu được. Máy tính còn tìm ra cả những quy tắc mà con người bỏ lỡ [4].

Cá nhà táng trao đổi các mẫu âm thanh gọi là coda(coda). Coda là âm thanh gồm nhiều tiếng tách ngắn nối lại. Trong đó có identity coda, tức coda cho biết bản sắc của đàn. Nó giống như một dấu hiệu âm thanh dùng chung trong cùng một dòng họ. Cá nhà táng nhận ra cùng đàn bằng âm thanh này. Khó xem nó giống hệt tên riêng của con người. Nó gần với khẩu hiệu lớp mà cả lớp cùng dùng hơn.

Trong âm thanh của cá nhà táng có những quy tắc ẩn bên trong. Chúng nối vài đơn vị âm thanh theo nhiều cách để tạo ra nhiều dạng. Lời nói của con người cũng vận hành theo cách đó. Con người nối vài phụ âm và nguyên âm để tạo ra vô số từ. Tuy nhiên, mỗi coda có nghĩa chính xác là gì thì vẫn đang được nghiên cứu. Cấu trúc âm thanh đã được làm rõ. Nhưng ý nghĩa thì vẫn chưa được giải hết.

Các nhà nghiên cứu não người giải thích ngôn ngữ theo cách mới. Đó là nhóm nghiên cứu của Giáo sư Evelina Fedorenko tại Viện Công nghệ Massachusetts(MIT). Năm 2024, họ đăng nghiên cứu trên tạp chí Nature. Họ cho rằng ngôn ngữ không phải điều kiện bắt buộc để tạo ra suy nghĩ sâu. Điều đó có nghĩa là ngôn ngữ chỉ là công cụ truyền nội tâm cho người khác. Suy nghĩ và lời nói là hai việc khác nhau. Không thể nói rằng không biết nói thì không có suy nghĩ [1].

Cách nghĩ này làm lung lay sự cố chấp lâu đời của luật. Không thể bỏ qua trí tuệ của động vật chỉ vì chúng không nói như con người. Động vật cũng trao đổi những âm thanh tinh vi theo cách của chúng. Trí tuệ nhân tạo giúp tìm ra quy tắc âm thanh mà con người không thấy.

Ở đây cần tách một điều ra để xem. Biết rằng âm thanh có quy tắc và thay đổi luật là hai việc khác nhau. Dù khoa học phát hiện sự thật mới, luật không đổi ngay. Luật chỉ chuyển động khi có đồng thuận xã hội, tranh luận lâu dài và án lệ mới. Dù vậy, nếu tiếng động vật được chứng minh là giao tiếp có ý nghĩa, luật cũng sẽ thay đổi. Luật sẽ khó tiếp tục xem động vật chỉ là đồ vật.

Tài liệu căn cứ liên quan

[1] Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature*, 630, 575-586. <https://doi.org/10.1038/s41586-024-07522-w>

- [2] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617.
<https://doi.org/10.1038/s41467-024-47221-8>
- [3] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. *Ecology Law Quarterly*, 52(1).
<https://doi.org/10.15779/Z383X83N5Q>
- [4] Project CETI. (2024, May 7). Sperm whale phonetic alphabet proposed for the first time. Project CETI.
<https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time>

10.2 Chương trình MOTH của NYU và quyền được máy móc hỗ trợ

Có một trung tâm đang cố đưa thay đổi này vào luật. Đó là chương trình về sự sống vượt ngoài con người tại Trường Luật Đại học New York (NYU). Tiếng Anh gọi là MOTH (More Than Human Life). Chương trình này do Giáo sư César Rodríguez-Garavito, một học giả luật môi trường, dẫn dắt [1].

Giáo sư Garavito nghiên cứu cùng nhiều học giả. Tiến sĩ Gašper Beguš của Đại học California, Berkeley (UC Berkeley) cũng tham gia. Họ đưa ra một ý tưởng mới. Đó là dùng âm thanh động vật do trí tuệ nhân tạo giải mã làm chứng cứ trước tòa. Họ gọi điều này là Machine-Assisted Rights, tức quyền được máy móc hỗ trợ. Đây chưa phải là thuật ngữ được dùng rộng rãi. Đây là đề xuất mới của họ.

Có thể giải thích quyền được máy móc hỗ trợ như sau. Trí tuệ nhân tạo đọc cách động vật giao tiếp và môi trường sống của chúng. Nó đo bằng con số thiệt hại mà động vật chịu do con người. Rồi những con số đó được đưa ra tòa như chứng cứ thay cho động vật. Nó giống như một thông dịch viên và người đại diện cho động vật không thể nói. Điều này giống việc người lớn giúp trẻ em khi trẻ không thể tự nộp giấy tờ.

Cách này muốn đi vòng qua điều kiện cũ của luật. Luật cũ chỉ trao tư cách khởi kiện cho con người. Tư cách này gọi là tư cách có thể nộp đơn kiện, tức nguyên đơn hợp lệ. Động vật không có tư cách này. Vì vậy các học giả muốn dùng thiệt hại do trí tuệ nhân tạo đo được làm chứng cứ. Bằng cách đó, họ muốn gõ cửa tòa án vì động vật.

Nhóm nghiên cứu của Giáo sư Beguš đã có một phát hiện lớn. Họ tìm thấy cấu trúc formant (formant) trong âm thanh của cá nhà táng. Formant là đoạn âm trong đó năng lượng dồn lại và vang lên rõ hơn hẳn. Nó giống điều xảy ra khi con người đổi hình miệng để phát âm nguyên âm. Trong công bố năm 2025, nhóm nghiên cứu cho biết trong coda của cá nhà táng có coda a(a) và coda i(i). Đó là những âm thanh khác nhau, giống nguyên âm của con người [2].

Nhưng bản đồ âm thanh do trí tuệ nhân tạo khôi phục cũng cho biết một sự thật buồn. Tiếng chân vịt của tàu lớn và tiếng nổ thăm dò tài nguyên lấp đầy biển. Những âm thanh này được gọi là tiếng ồn biển do con người tạo ra. Một nghiên cứu đăng trên tạp chí Science năm 2021 đã bàn về điều này. Nội dung là sau khi công nghiệp phát triển, biển đã ồn hơn nhiều. Cũng có nghiên cứu cho rằng tiếng ồn tàu thuyền có thể ảnh hưởng đến cấu trúc coda của cá nhà táng [3].

Chương trình MOTH cho rằng thiệt hại này có thể được dùng làm chứng cứ trước tòa. Họ nói không cần dịch hoàn hảo ý nghĩa của âm thanh. Chỉ cần cho thấy bằng con số việc giao tiếp của đàn cá nhà táng bị phá hỏng đến mức nào. Nghĩa là lấy chính biểu đồ âm thanh bị méo làm chứng cứ. Khi đó, tòa án có thể buộc tàu đổi tuyến hoặc chặn thiết bị âm thanh quân sự. Lập luận này nói đến hai quyền của cá nhà táng. Đó là quyền thoát khỏi đau đớn và quyền giữ gìn văn hóa [4].

Ý tưởng này cũng làm luật hiện có mạnh hơn. Ở Hoa Kỳ có luật bảo vệ động vật đang có nguy cơ biến mất. Cũng có luật riêng bảo vệ động vật có vú sống ở biển. Những con số thiệt hại do trí tuệ nhân tạo đo được trở thành sức mạnh khiến các luật này được thực thi. Điều đó có nghĩa là ngay trước khi lập quyền mới, ta vẫn có thể dùng luật hiện có mạnh hơn.

Ở đây cũng cần tách ra để xem. Biết rằng âm thanh bị phá hỏng và thay đổi luật là hai việc khác nhau. Vẫn chưa xác nhận có vụ kiện nào đã thật sự nộp chứng cứ đó ra tòa. Các học giả hiện mới ở bước đề xuất rằng có thể dùng như vậy.

Tài liệu căn cứ liên quan

- [1] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. *Ecology Law Quarterly*, 52(1). <https://doi.org/10.15779/Z383X83N5Q>
- [2] Beguš, G., Sprouss, C., Lebanidze, T., Gero, S., & Andreas, J. (2025). Vowel- and diphthong-like spectral patterns in sperm whale codas. *Open Mind*, 9, 1849-1874. <https://doi.org/10.1162/OPMI.a.252>
- [3] Duarte, C. M., Chapuis, L., Collin, S. P., et al. (2021). The soundscape of the Anthropocene ocean. *Science*, 371(6529), eaba4658. <https://doi.org/10.1126/science.aba4658>
- [4] Surma, K. (2025, October 29). AI is decoding whales' communications. Could that be a turning point in the push for their rights? *Inside Climate News*. <https://insideclimatenews.org/news/29102025/ai-sperm-whale-communications-legal-rights/>

10.3 Lời chứng ở vùng Caribe và trí tuệ của người bản địa

Cuộc thảo luận về nhân vị phi con người không chỉ nằm trong máy tính của các học giả phương Tây. Nhân vị phi con người là một tồn tại không phải con người nhưng có thể được luật bảo vệ. Cuộc thảo luận này gặp trí tuệ của người bản địa, những người đã bảo vệ biển từ lâu, và phát triển lớn hơn [2].

Ngày 28 tháng 3 năm 2024, một việc lịch sử đã diễn ra tại Rarotonga thuộc Quần đảo Cook (Cook Islands) ở Thái Bình Dương. Vua Māori và các nhà lãnh đạo Quần đảo Cook gặp nhau tại một nơi. Các nhà lãnh đạo Tahiti, Tonga và Hawaii cũng tham gia. Họ công nhận cá voi là chủ thể độc lập có quyền. Họ tuyên bố cá voi là tồn tại không phải con người nhưng được luật bảo vệ. Tên của tuyên bố chung này là He Whakaputanga Moana, tức Tuyên bố Đại Dương. Tuyên bố này được nhắc lại rộng rãi tại Waitangi, New Zealand, vào tháng 2 năm 2026 [1][4].

Tuyên bố này là tuyên bố đầu tiên do người bản địa dẫn dắt công nhận cá voi là sinh vật biết cảm nhận. Tuyên bố nói rằng cá voi có quyền bơi tự do và quyền sống trong biển khỏe mạnh.

Tuyên bố này bắt nguồn từ tri thức sinh thái truyền thống. Tri thức sinh thái truyền thống là hiểu biết lâu đời xem thiên nhiên như họ hàng cùng sống. Người bản địa đã truyền lại tri thức này qua thời gian dài. Họ biết bài hát của cá voi không phải là tiếng ồn ngẫu nhiên. Họ tin rằng bài hát ấy chứa ký ức của biển và lịch sử của dòng họ.

Trí tuệ nhân tạo hiện đại chỉ cùng hướng với trực giác lâu đời này. Project CETI đã quan sát cá voi ngoài khơi Dominica ở vùng Caribe trong thời gian dài. Nhóm nghiên cứu xem xét coda của những con cá voi này bằng trí tuệ nhân tạo. Khi ghép vài đặc điểm lại, số coda có thể phân biệt tăng lên nhiều. Trước đây, người ta cho rằng có khoảng hai mươi mốt loại coda. Nhưng khi làm rõ cách kết hợp, số dạng có thể phân biệt vượt quá một trăm. Điều này giống việc con người dùng ít chữ cái để tạo ra nhiều từ. Nó cũng phù hợp với niềm tin của người bản địa rằng bài hát mang theo ý nghĩa [3].

Các nhà lãnh đạo Quần đảo Cook từ lâu đã xem cá voi là họ hàng. Họ đối xử với cá voi như bậc trưởng lão lớn của biển. Tầm lòng này đã có từ rất lâu trước khi khoa học xuất hiện. Nay trí tuệ nhân tạo bổ sung cho tầm lòng ấy một căn cứ bằng con số.

Vì vậy người bản địa và các nhà khoa học đã cùng tạo ra một cách mới. Đó là đặt phép tính của trí tuệ nhân tạo cạnh tri thức biển của người bản địa. Ở đây, người bản địa không phải trợ lý nghiên cứu. Họ là nhân vật chính cùng quyết định từ lúc thiết kế trí tuệ nhân tạo. Nhà khoa học và người bản địa nắm tay ngang hàng để bảo vệ biển.

Thỏa thuận như vậy không chỉ ở trong một nước. Vì đại dương trải qua vùng biển của nhiều nước. Cá voi không biết biên giới và đi lại trên biển rộng. Vì vậy luật của một nước không thể bảo vệ hết cá voi. Tuyên bố Đại Dương là bước đầu để nhiều nước và bộ tộc cùng nắm tay.

Tài liệu căn cứ liên quan

- [1] He Whakaputanga Moana. (2024, March 28). Whale protection and legal personhood declaration. Rarotonga, Cook Islands. <https://www.whaledclaration.com/>
- [2] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. *Ecology Law Quarterly*, 52(1). <https://doi.org/10.15779/Z383X83N5Q>
- [3] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>
- [4] Inside Climate News. (2026, February 5). How a 'powerful vision' on whales and oceans could change worldviews. Inside Climate News. <https://insideclimatenews.org/news/05022026/maori-scientist-daniel-hikuroa-legal-rights-of-whales/>

10.4 Quyền tiến hóa của loài

Ở cuối con đường mà trí tuệ nhân tạo mở ra có một ý tưởng lớn hơn. Đó là ý tưởng vượt qua lệnh đối tuyến đường hàng hải. Đó là quyền để trí tuệ phi con người tự đi về tương lai. Một số học giả gọi điều này là quyền tiến hóa của loài. Đây chưa phải quyền được luật quy định. Nó là đề xuất của các học giả [1].

Lời nói của con người đã lớn lên qua thời gian dài. Nó bắt đầu từ vài âm thanh rồi chín dần thành ngữ pháp phức tạp. Văn hóa âm thanh của động vật cũng không phải mẫu vật đứng yên. Nó là tài sản sống, được học, bắt chước, thay đổi và lớn lên qua các thế hệ. Đây cũng là lý do bài hát cá voi khác nhau đôi chút tùy vùng. Nó giống việc lời nói của con người chia thành phương ngữ theo từng vùng [3].

Có thể hiểu điều này khi nhìn cá nhà táng mẹ. Cá mẹ dạy âm thanh của đàn cho con non mới sinh. Con non bắt chước rất lâu bằng âm thanh vụng về như trẻ bập bẹ. Nó giống em bé học nói và bắt chước còn vụng. Điều đó có nghĩa là âm thanh của cá nhà táng không phải bản năng được khắc sẵn khi sinh ra. Nó là di sản văn hóa truyền từ thế hệ này sang thế hệ khác. Như con người truyền lại sách và bài hát, cá voi truyền lại âm thanh.

Nhưng tiếng ồn mà con người đổ vào biển ngăn việc học này. Tiếng ồn công nghiệp và tiếng nổ quân sự ngày đêm lấp đầy biển. Cá voi con không nghe rõ âm thanh của cha mẹ. Vì dải cao độ quan trọng để học âm thanh bị trộn lẫn. Điều này giống như trong lớp quá ồn thì không nghe được lời thầy cô. Việc này không chỉ gây hại cho một con. Nó cắt đứt mạch văn hóa và ký ức của cả một đàn [2].

Vì vậy một số nhà triết học sinh thái và luật học đưa ra một lập luận. Họ muốn bảo vệ quyền để động vật học và nối tiếp âm thanh một cách an toàn. Lúc này, thế giới nơi động vật sống được gọi là thế giới riêng mà mỗi loài cảm nhận bằng giác quan của mình. Thế giới mà dơi cảm nhận và thế giới mà cá voi cảm nhận khác nhau. Họ đề xuất lập quyền để chúng tự lớn lên, không bị tiếng ồn hay âm thanh giả của con người làm phiền. Đây chưa phải luật. Nó còn ở giai đoạn thảo luận của học giả.

Âm thanh cá voi giả do trí tuệ nhân tạo tạo ra cũng cần được thận trọng. Âm thanh giả do con người bật bừa có thể làm cá con lẫn lộn khi học. Vì vậy đề xuất này cũng có phần ngăn âm thanh giả. Nếu dùng sai, công nghệ tốt cũng có thể trở thành tiếng ồn mới. Máy thông dịch phải là công cụ lắng nghe thế giới của cá voi. Nó không được trở thành công cụ làm rối thế giới của cá voi.

Nếu giảm tiếng ồn biển, tiếng ồn sẽ nhanh chóng lắng xuống. Nghiên cứu Science năm 2021 đã làm rõ điểm này. Tiếng ồn phát ra ngay tại chỗ. Vì vậy nếu giảm nguyên nhân, biển sẽ sớm yên tĩnh. Nếu chúng ta giảm tiếng ồn thương mại, cá voi có thể nghe rõ tiếng của nhau trở lại.

Tài liệu căn cứ liên quan

- [1] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. *Ecology Law Quarterly*, 52(1). <https://doi.org/10.15779/Z383X83N5Q>
- [2] Duarte, C. M., Chapuis, L., Collin, S. P., et al. (2021). The soundscape of the Anthropocene ocean. *Science*, 371(6529), eaba4658. <https://doi.org/10.1126/science.aba4658>
- [3] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617. <https://doi.org/10.1038/s41467-024-47221-8>

10.5 Những chứng cứ mới và câu chuyện tiếp theo

Lĩnh vực này vẫn đang lớn lên nhanh chóng. Mỗi năm lại có nghiên cứu mới và công cụ mới xuất hiện. Đó là những việc vài năm trước còn khó tưởng tượng [1].

Nghiên cứu của Berkeley và Project CETI là một ví dụ tốt. Như đã thấy ở trên, nhóm nghiên cứu tìm thấy những âm giống nguyên âm trong coda của cá nhà táng. Năm 2026, một nghiên cứu tiếp theo được đăng trên *Proceedings of the Royal Society B*. Nội dung nói rằng những âm giống nguyên âm của cá nhà táng cũng theo quy tắc như nguyên âm của con người. Bằng chứng cho thấy cá nhà táng tạo âm thanh giống con người đã vững chắc hơn.

Earth Species Project mở ra một con đường khác. Tổ chức này đã tạo ra một trí tuệ nhân tạo tên là NatureLM-audio. Đây là mô hình ngôn ngữ âm thanh lớn đầu tiên dành cho tiếng động vật. Mô hình này chỉ nghe âm thanh mà đoán được đó là loài vật nào. Nó cũng đếm được trong bản ghi có mấy con vật. Nó còn nhận ra ở mức nhất định cả tiếng của loài vật mà nó chưa từng học [2].

Năm 2026, tổ chức này mở rộng cách nghiên cứu. Thay vì chỉ đào sâu vào một loài, họ gom tiếng của nhiều loài lại cùng một lúc. Bằng cách này, họ mở rộng phương pháp dùng máy tính để phân tích tiếng của nhiều loài vật cùng nhau. Tháng 7 năm 2026, họ công bố alp-data, một kho dữ liệu chung cho việc đó. Đây là nền tảng để các nhà nghiên cứu trên thế giới chia sẻ cùng một dữ liệu. Nó giống như trong thư viện, ai cũng có thể mượn cùng một cuốn sách [3].

Tháng 3 năm 2026, Project CETI công bố hai nghiên cứu mới. Một nghiên cứu ghi lại rất kỹ việc sinh con của cá nhà táng. Đây là bản ghi đầu tiên mô tả việc sinh con dưới biển chi tiết đến mức này. Nghiên cứu còn lại cho thấy cả đàn giúp đỡ khi sinh con. Khi cá mẹ sinh con, những con cá voi khác ở bên cạnh và bảo vệ. Đây là nghiên cứu cho thấy rõ việc chăm sóc hợp tác như vậy ở động vật không phải linh trưởng. Có học giả nhìn những hành vi này rồi nói về quyền của cá voi. Họ cho rằng tính xã hội phức tạp và hành vi chăm sóc có thể là tài liệu tham khảo khi bàn về quyền. Tuy nhiên, khó có thể xem ngay hành vi giúp nhau là nghĩa vụ pháp lý. Cá voi giúp bạn và con người tuân theo pháp luật là hai việc khác nhau [4].

Còn rất lâu nữa pháp luật mới đối xử với động vật như con người. Cho đến nay, tòa án ở nhiều nước thường không công nhận tư cách khởi kiện của động vật. Phán quyết năm 2022 của tòa án bang New York, Hoa Kỳ, là một ví dụ. Một tổ chức bảo vệ quyền động vật đã kiện thay cho con voi Happy. Họ yêu cầu thả Happy khỏi chuồng hẹp. Tòa án không chấp nhận yêu cầu đó. Lý do là không thể dùng nguyên xi chế độ dành cho con người cho voi. Như vậy, biết sự thật và thay đổi pháp luật là hai việc khác nhau [5].

Tất cả các bằng chứng này chỉ về một hướng. Đó là tiếng động vật không phải tiếng ồn, mà là giao tiếp có ý nghĩa. Càng có nhiều bằng chứng, pháp luật càng khó giữ động vật ở vị trí như đồ vật. Khoa học đi trước, còn pháp luật chậm rãi đi theo sau. Trí tuệ nhân tạo đang trở thành thông dịch viên đưa tiếng nói của động vật vào tòa án. Việc dịch ngôn ngữ của sự sống trên Trái Đất nối liền với việc xây dựng quyền cho chúng.

Tài liệu căn cứ liên quan

- [1] Beguš, G., Sprous, C., Lebanidze, T., Gero, S., & Andreas, J. (2026). The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B*, 293(2044). <https://doi.org/10.1098/rspb.2025.2994>
- [2] Robinson, D., Miron, M., Hagiwara, M., Weck, B., Keen, S., Alizadeh, M., Narula, G., Geist, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. *International Conference on Learning Representations (ICLR) 2025*. <https://arxiv.org/abs/2411.07186>
- [3] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. <https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>
- [4] Project CETI. (2026, March). Studies documenting rare sperm whale birth and ancient cooperative care released. *EurekAlert!* <https://www.eurekalert.org/news-releases/1120823>
- [5] Nonhuman Rights Project, Inc. v. Breheny. (2022). 38 N.Y.3d 555, New York Court of Appeals. <https://www.nycourts.gov/ctapps/Decisions/2022/Jun22/52opn22-Decision.pdf>

Lời kết: Internet giữa các loài và Trái Đất cùng chung sống

E.1 Tiếng click của cá nhà táng và bảng chữ cái nơi biển sâu

Dưới biển sâu luôn tối. Ở độ sâu hàng nghìn mét, nơi ánh nắng không tới, mọi thứ tối như đêm đen. Trong bóng tối ấy, cá nhà táng nhìn thế giới bằng âm thanh. Cá nhà táng bắn ra tiếng lách cách từ phần trán. Mỗi âm thanh ngắn này được gọi là click [3].

Cá nhà táng phát ra nhiều click nối tiếp nhau. Nhóm âm thanh nối tiếp như vậy được gọi là coda [2, 3]. Click là một âm thanh. Coda là một chùm âm thanh gồm nhiều click gộp lại. Coda được tạo bằng cách nối khoảng 3 đến 40 click. Với tai người, coda chỉ nghe như những tiếng lách cách giống nhau. Trong thời gian dài, con người coi âm thanh này là tiếng ồn vô nghĩa [2].

Nhưng trí tuệ nhân tạo đã nghe lại âm thanh ấy. Một nhóm nghiên cứu tên là Project CETI đã bắt tay vào việc. CETI là nhóm gồm các học giả từ nhiều nước, muốn giải mã lời nói của cá voi. Nhà nghiên cứu Pratyusha Sharma và giáo sư Jacob Andreas của nhóm này đã phân tích âm thanh. Tiến sĩ Shane Gero, người đã quan sát cá nhà táng Dominica trong thời gian dài, cũng cùng tham gia [1].

Nhóm nghiên cứu đã thu thập và xem xét rất nhiều âm thanh. Họ xem kỹ từng coda trong 8.719 coda của cá nhà táng. Trong đó, họ tìm thấy bốn yếu tố cơ bản tạo nên coda. Đó là những thứ như nhịp điệu, tốc độ của âm thanh, và các click nhỏ được thêm vào. Khi thay đổi các yếu tố này theo nhiều cách, các coda khác nhau được tạo ra.

Trí tuệ nhân tạo đã tìm ra quy luật trong các click tưởng như giống hệt nhau. Cá nhà táng làm khoảng cách giữa các click dài ra hoặc ngắn lại từng chút một. Chúng cũng thay đổi tốc độ âm thanh. Chúng còn thêm âm thanh nhỏ vào trước hoặc sau nhóm click. Những khác biệt nhỏ này hợp lại và tạo nên các cấu trúc âm thanh khác nhau.

Cá nhà táng cũng điều chỉnh tốc độ âm thanh. Chúng có khi phát cùng một coda dài hơn một chút. Có khi chúng rút ngắn lại và phát nhanh hơn. Điều này giống như khi hát, ta kéo dài hoặc rút ngắn nhịp. Sự kéo dài và rút ngắn nhỏ này cũng là chìa khóa làm thay đổi âm thanh.

Nghiên cứu mới còn đi thêm một bước. Các học giả xem trong coda, tiếng vang ở độ cao nào vang mạnh. Khi con người phát âm nguyên âm 'a' và 'i', cũng có chỗ vang mạnh. Trong coda của cá nhà táng cũng xuất hiện những chỗ như vậy. Nhóm CETI đã tìm thấy hai loại nguyên âm cá voi như thế. Chúng giống âm 'a' và 'i' do con người phát ra [4].

Cá nhà táng tạo ra những nguyên âm này bằng cách nào. Cá nhà táng tạo âm thanh bằng môi âm thanh ở bên trong đầu. Rồi chúng chỉnh âm đó bằng các túi khí trong đầu. Điều này giống như con người phát âm từ dây thanh và đổi hình miệng để tạo nguyên âm. Có thể nói nơi tạo âm và nơi chỉnh âm được tách riêng.

Nhóm nghiên cứu cho biết cá nhà táng phát ra khoảng 156 loại coda khác nhau. Những coda này không phải là chữ cái riêng lẻ. Chúng là các nhóm âm thanh được tạo bằng cách nối vài yếu tố như nhịp điệu và tốc độ. Ở biển sâu cũng có khả năng tạo nhiều âm thanh từ ít vật liệu. Tuy vậy, ý nghĩa của từng coda vẫn chưa được làm rõ. Nhóm nghiên cứu mới chỉ tìm ra cấu trúc của âm thanh.

Cùng một coda cũng khác nhau đôi chút giữa các đàn. Điều này giống tiếng địa phương của con người. Trong cùng một đàn, chúng phát âm chồng lên nhau và giữ nhịp với nhau. Khi một con cá voi phát âm, con khác nối âm ngay sau đó. Như vậy, cá nhà táng cùng nhau tạo nên dòng chảy của cuộc trò chuyện.

Một số coda đóng vai trò giống như thẻ tên. Cá nhà táng sống thành các đàn gia đình lớn. Mỗi đàn có những coda hay dùng riêng. Nghe coda đó, ta có thể biết cá voi thuộc đàn nào. Điều này giống như đoán quê của một người qua tiếng địa phương.

Điều quan trọng là âm thanh được tạo nên nhờ các khác biệt nhỏ tích lại. Cá nhà táng nối vài yếu tố âm thanh theo nhiều cách. Tùy cách nối, các coda khác nhau được tạo ra. Điều này giống con người dùng vài chữ cái để tạo vô số từ. Ở biển sâu cũng có một cấu trúc như thế ẩn bên trong.

Muốn bắt được âm thanh này, cần thiết bị ghi âm tốt. Nhóm CETI đã làm ra máy ghi âm nhỏ gắn lên cơ thể cá voi. Thiết bị này ghi lại âm thanh của cá voi cùng với tình hình dưới nước. Vì phải xem âm thanh được phát ra trong hoàn cảnh nào thì mới giải được ý nghĩa. Chỉ sau khi con người có công cụ mới là trí tuệ nhân tạo, bảng chữ cái của biển sâu mới hiện ra.

Tài liệu căn cứ liên quan

- [1] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, 15, 3617.
<https://doi.org/10.1038/s41467-024-47221-8>
- [2] Oceanographic. (2026). Scientists listen in on sperm whale communication with new tech. *Oceanographic Magazine*.
<https://oceanographicmagazine.com/news/scientists-listen-in-on-sperm-whale-communication-with-new-tech/>
- [3] Project CETI. (2026). Research. Project CETI. <https://www.projectceti.org/research/index>
- [4] Kemeny, R. (2026, April 14). We're one step closer to understanding the sperm whale 'alphabet'. *National Geographic*.
<https://www.nationalgeographic.com/animals/article/sperm-whale-vowel-patterns-communication-discovery>

E.2 Vượt qua ý nghĩ rằng chỉ con người mới biết nói

Trong thời gian dài, con người tự coi mình là đặc biệt. Họ tin rằng trên Trái Đất, chỉ con người biết nói và chia sẻ ý nghĩa. Âm thanh của động vật không giống lời nói của con người bị xem chỉ là tiếng ồn. Ý nghĩ rằng chỉ con người đặc biệt đã kéo dài rất lâu [1].

Niềm tin này vững như một bức tường lâu đời. Vì bức tường ấy, âm thanh của động vật bị chôn trong bóng tối. Con người không muốn lắng nghe những sinh mệnh bên cạnh đang nghĩ gì. Chúng ta không biết trong thời gian đó đã có bao nhiêu câu chuyện trôi qua mất.

Bức tường này lớn lên từ thói quen đo động vật bằng cách nói của con người. Con người nghĩ rằng phải có từ và câu thì mới là lời nói. Những âm thanh không vừa với khuôn ấy bị gạt ra là vô nghĩa. Nhưng âm thanh của động vật không có lý do gì phải vừa với khuôn lời nói của con người. Vì chúng đi ra từ cơ thể khác nhau và đời sống khác nhau.

Trí tuệ nhân tạo đang từng chút phá bỏ bức tường này [3, 4]. Trí tuệ nhân tạo sắp xếp âm thanh của động vật trên một bản đồ ý nghĩ. Các học giả gọi bản đồ này là không gian tiềm ẩn (latent space). Trên bản đồ này, âm thanh giống nhau thì các điểm cũng tụ lại gần nhau. Âm thanh khác nhau thì các điểm cũng cách xa nhau [4].

Các học giả đặt bản đồ lời nói của con người và bản đồ âm thanh của động vật cạnh nhau. Hai bản đồ không hoàn toàn xa lạ với nhau. Khi chỉnh hình dạng của hai bản đồ rồi chồng lên nhau, những vị trí giống nhau xuất hiện. Có khi một từ của con người và một tín hiệu của động vật nằm ở cùng một chỗ.

Ở đây có một phương pháp thú vị. Khi chồng hai bản đồ lên nhau, không cần bảng dịch làm sẵn. Các học giả chỉ nhìn hình dạng của hai bản đồ để tìm vị trí khớp nhau. Điều này giống như chồng hai tờ giấy lên cửa sổ rồi chỉnh cho hoa văn khớp. Các học giả hiện đang thử cách chồng theo hình dạng như vậy.

Có những học giả dẫn dắt nghiên cứu này. Nhà ngôn ngữ học người Mỹ Gašper Beguš dùng trí tuệ nhân tạo để tìm quy luật trong âm thanh động vật. Họ không ép quy luật lời nói của con người lên động vật. Họ chờ quy luật tự hiện ra từ hình dạng của chính âm thanh.

Từ đây, một ý nghĩ mới lớn lên. Động vật không phải là đối tượng chờ lời nói của con người. Mỗi loài cảm nhận thế giới riêng của mình và đã vẽ thế giới đó bằng âm thanh. Thế giới mà động vật cảm nhận bằng giác quan riêng được gọi là Umwelt. Dơi nhìn thế giới bằng âm thanh, chó đọc thế giới bằng mũi. Cá voi có thể giới của cá voi, và chim có thể giới của chim.

Cách chồng bản đồ này có sức mạnh trong tình huống khó. Giữa con người và động vật không có từ điển làm sẵn. Cũng không có bảng ghép từng từ với nhau. Dù vậy, trí tuệ nhân tạo vẫn cố tìm cặp tương ứng chỉ bằng hình dạng của hai bản đồ. Nó có thể trở thành cây cầu mới nối các thế giới khác nhau.

Vì vậy, chúng ta không nên đặt âm thanh động vật thấp hơn lời nói của con người. Ta cần tôn trọng suy nghĩ và nhịp điệu của động vật ngang hàng với mình. Đó là tấm lòng coi âm thanh của mọi sự sống đều quý như nhau. Nếu con người muốn lớn lên thành một thành viên trong gia đình Trái Đất, trước hết cần có tấm lòng này [2].

Tài liệu căn cứ liên quan

- [1] Rodríguez-Garavito, C. (2026, February 19). AI-enabled decoding of whale communication could bolster animal rights. NYU School of Law Ideas. <https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai>
- [2] MOTH - More Than Human Life. (2026). MOTH x CETI: Exploring how understanding sperm whale communications can be a force for good. mothrights.org. <https://mothrights.org/news/moth-x-ceti-exploring-how-understanding-sperm-whale-communications-can-be-a-force-for-good/>
- [3] Bioneers. (2026). The quest to decode whale-speak. Bioneers. <https://bioneers.org/the-quest-to-decode-whale-speak/>
- [4] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. <https://doi.org/10.1126/science.adg7314>

E.3 Internet giữa các loài nối sự sống trên Trái Đất

Giữa con người có internet. Internet là mạng lưới nối những người ở xa để họ trao đổi tin tức. Nhưng trên Trái Đất không chỉ có con người sống. Cá voi, voi, chim, ong mật, và cả cây cối cũng cùng sống. Có một giấc mơ muốn nối các sự sống này thành một. Giấc mơ ấy được gọi là Internet giữa các loài (Interspecies Internet) [1].

Giấc mơ này xuất hiện lần đầu vào năm 2013. Vint Cerf, người tạo ra internet, và nhạc sĩ Peter Gabriel đã đi đầu. Học giả Neil Gershenfeld và Diana Reiss cũng cùng tham gia. Họ nói rằng đừng để internet chỉ thuộc về con người. Ý nghĩ của họ là mở mạng lưới ấy cho động vật nữa.

Internet giữa các loài không dừng lại ở việc dịch âm thanh động vật bằng ứng dụng. Nó có nghĩa là một mạng lưới lớn để mọi sự sống trên Trái Đất chia sẻ tin tức và nguy hiểm của nhau. Giống như dây thần kinh của con người nối khắp cơ thể, mạng lưới này nối sự sống trên toàn Trái Đất.

Thật ra, các mảnh của sự kết nối này đã có sẵn trong tự nhiên. Nhiều loài động vật nghe lén tiếng cảnh báo của loài khác. Khi meerkat báo nguy, động vật khác ở gần hiểu âm thanh đó và cùng tránh đi. Các nhà khoa học đã xác nhận cách nghe lén này ở hơn 70 loài động vật. Như vậy, các loài động vật khác nhau đang chia sẻ cùng một thông tin nguy hiểm [2].

Cá heo cũng cho thấy sự chia sẻ như vậy. Khi đàn này và đàn kia gặp nhau trên biển, cá heo trao đổi tiếng huýt sáo. Mỗi tiếng huýt sáo khác nhau này đóng vai trò giống như thẻ tên. Chúng xác nhận bằng âm thanh xem nhau là ai. Con đường âm thanh không nhìn thấy này đã tạo thành một mạng lưới [3].

Mạng lưới này không chỉ nổi âm thanh. Cây và cỏ cũng phát tín hiệu. Khi thiếu nước, các loài thực vật như cà chua và thuốc lá phát ra âm thanh tai người không nghe được. Một cây bị căng thẳng phát tiếng lách cách khoảng 30 đến 50 lần trong một giờ. Trí tuệ nhân tạo có thể gom các âm thanh như vậy và đọc được trạng thái của cây [4].

Trí tuệ nhân tạo giúp nối các mảnh này lại với nhau. Mạng thông tin mới trong tương lai cũng có thể góp sức. Mạng 6G có thể cho sóng vô tuyến bật vào vật thể để đọc chuyển động xung quanh. Ý tưởng dùng mạng thông tin như cảm biến này được gọi là network as a sensor. Tuy vậy, việc nối công nghệ này với mạng giao tiếp của động vật hiện vẫn chỉ là khả năng [5].

Nếu dùng tốt mạng lưới này, ta có thể bảo vệ thiên nhiên. Hãy giả sử có người bắt đầu chặt cây bừa bãi trong rừng mưa nhiệt đới. Chuyển động tản ra của chim và tín hiệu bất thường của cây sẽ đi theo mạng lưới và tập trung về trung tâm. Con người có thể nhận tin ngay và ngăn lại. Internet giữa các loài có thể trở thành công cụ để cùng sống, không phải công cụ giám sát. Nó sẽ trở thành mạng thần kinh lớn, nơi sự sống trên Trái Đất chia sẻ nguy hiểm và chăm sóc lẫn nhau.

Tài liệu căn cứ liên quan

- [1] Interspecies Internet. (n.d.). About. Interspecies Internet. <https://www.interspecies.io/about>
- [2] Magrath, R. D., Haff, T. M., Fallow, P. M., & Radford, A. N. (2015). Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. *Biological Reviews*, 90(2), 560-586. <https://doi.org/10.1111/brv.12122>
- [3] Quick, N. J., & Janik, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. *Proceedings of the Royal Society B*, 279(1738), 2539-2545. <https://doi.org/10.1098/rspb.2011.2537>
- [4] Khait, I., Lewin-Epstein, O., Sharon, R., et al. (2023). Sounds emitted by plants under stress are airborne and informative. *Cell*, 186(7), 1328-1336. <https://doi.org/10.1016/j.cell.2023.03.009>
- [5] Nokia Bell Labs. (n.d.). Network as a sensor. Nokia. <https://www.nokia.com/bell-labs/research/6g-networks/6g-technologies/network-as-a-sensor/>

E.4 Những điều công cụ mới làm sáng tỏ: trí tuệ nhân tạo học lời nói của tự nhiên

Trí tuệ nhân tạo cũng tiến tới việc học âm thanh động vật. Earth Species Project đã công bố một trí tuệ nhân tạo mới. Tên của nó là NatureLM-audio. Đây là trí tuệ nhân tạo âm thanh lớn đầu tiên được tạo cho âm thanh động vật. Nó được công bố lần đầu vào tháng 11 năm 2024, và được giới thiệu chính thức tại hội nghị quốc tế ICLR năm 2025 [1].

Trí tuệ nhân tạo này làm nhiều việc cùng lúc. Nó nghe âm thanh và đoán đó là động vật nào. Nó cũng phân biệt con vật đó là con non hay con trưởng thành. Nó chia loại âm thanh và gắn lời giải thích cho âm thanh. Nó còn đếm trong bản ghi có bao nhiêu con vật.

Điều đáng chú ý là nó làm được việc con người không dạy. Thông thường, trí tuệ nhân tạo chỉ làm được việc đã học. Nhưng trí tuệ nhân tạo này tự có thêm kỹ năng mới trong lúc học rất nhiều âm thanh. Việc đếm có bao nhiêu con cũng không phải là điều nó học từ đầu. Vì nghe nhiều, nó tự làm được.

Earth Species Project xem xét lời nói của nhiều loài động vật cùng nhau. Hiện nay, họ đang nghiên cứu sâu âm thanh của quạ, cá voi trắng và voi. Họ hợp tác với các nhà sinh vật học trên khắp thế giới để thu thập dữ liệu. Cách làm của nhóm này là không chỉ nhìn một loài, mà nhìn nhiều loài cạnh nhau.

Vào tháng 7 năm 2026, một công cụ mới nữa ra đời. Tên của nó là alp-data. Đây là công cụ gom dữ liệu âm thanh của nhiều loài động vật vào một nơi. Nó chứa 35 bộ dữ liệu về tiếng chim, tiếng cá voi và tiếng côn trùng. Giờ đây, các học giả có thể lấy âm thanh của nhiều loài động vật từ một cửa chung [2].

Earth Species Project gọi dòng nghiên cứu này là Xử lý ngôn ngữ động vật (Animal Language Processing). Đó là cách nghiên cứu giao tiếp của động vật bằng trí tuệ nhân tạo. Nó không áp nguyên xi

quy luật lời nói của con người lên động vật. Trước hết, nó thừa nhận rằng mỗi loài động vật có thể có cách giao tiếp khác nhau. Trên nền đó, trí tuệ nhân tạo tìm quy luật từ dữ liệu [3].

Nhóm này cũng hỏi suy nghĩ của mọi người. Năm 2025, họ hỏi 1.057 người ở 67 quốc gia trên thế giới. Nhiều người vui mừng trước việc con người có thể hiểu lời nói của động vật. Đồng thời, họ cũng bày tỏ lo lắng rằng không được đối xử tùy tiện với động vật. Khi công nghệ lớn lên, tầm lòng thận trọng cũng lớn lên [4].

Những công cụ mới này làm thay đổi cách làm việc của các nhà khoa học. Trước đây, một nhà khoa học thường nghiên cứu lâu dài chỉ một loài vật. Dữ liệu cũng nằm rải rác, mỗi nơi một kiểu. Bây giờ, dữ liệu của nhiều loài vật được xem cùng nhau qua một cửa chung. Những nghiên cứu từng tách rời đang gom lại thành một dòng chảy lớn.

Mục tiêu tiếp theo là thử nghiệm phát lại âm thanh. Đây là thử nghiệm dùng âm thanh do trí tuệ nhân tạo tạo ra để thử nói chuyện với động vật. Kế hoạch này chủ yếu do Project CETI công bố. Earth Species Project tập trung vào việc thu thập và xem xét dữ liệu hơn là phát âm thanh một cách tùy tiện. Việc động vật có hiểu âm thanh do con người tạo ra hay không vẫn chưa được xác nhận [5].

Tài liệu căn cứ liên quan

- [1] Earth Species Project. (2024, November 11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. <https://earthspecies.org/2024/11/11/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics/>
- [2] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. <https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/>
- [3] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. <https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/>
- [4] Earth Species Project. (2025, September 29). What the world thinks about AI and animal communication: Findings from our first global survey. Earth Species Project. <https://earthspecies.org/2025/09/29/what-the-world-thinks-about-ai-and-animal-communication-findings-from-our-first-global-survey/>
- [5] Project CETI. (2026). Research. Project CETI. <https://www.projectceti.org/research/index>

E.5 Lắng nghe cẩn thận: công cụ có trách nhiệm và lời hứa cùng tồn tại

Công cụ mới có sức mạnh lớn. Công cụ mạnh nếu dùng sai sẽ gây hại. Trí tuệ nhân tạo giờ có thể tạo ra âm thanh động vật gần như y hệt. Nếu dùng sức mạnh này vào việc xấu, con người có thể lừa động vật. Hãy thử nghĩ đến việc dùng âm thanh giả để dụ động vật hoang dã ra ngoài. Kẻ săn trộm cũng có thể dùng âm thanh giả để đi săn [1][2].

Vì vậy, chúng ta phải giữ ba lời hứa.

Thứ nhất là lời hứa không can thiệp bừa bãi. Các đàn động vật đã xây dựng văn hóa âm thanh riêng trong thời gian rất dài. Con người không được đưa âm thanh giả vào để buôn bán hay thống trị, rồi làm rối trật tự ấy. Trí tuệ nhân tạo không phải là vũ khí để cướp tiếng nói của động vật. Nó phải là công cụ giúp bảo vệ văn hóa của động vật theo ý của chính chúng.

Vì sao cần lời hứa như vậy? Con người từ lâu chỉ muốn thắng thiên nhiên. Sức mạnh hiểu âm thanh cũng có thể biến thành sức mạnh điều khiển đối phương. Nếu can thiệp bừa bãi vào một đàn, đời sống của động vật sẽ bị rối loạn. Vì vậy, càng biết nhiều thì cũng càng phải biết lùi lại.

Thứ hai là lời hứa để lại dấu trên âm thanh. Các nhà khoa học đề xuất gắn một dấu vô hình vào âm thanh động vật do trí tuệ nhân tạo tạo ra. Dấu này gọi là audio watermark. Giống như hình ẩn trên tiền

giả, tai người không nghe thấy nhưng máy có thể nhận ra. Nhờ vậy, ta có thể phân biệt âm thanh nào là thật và âm thanh nào là giả [3].

Dấu này cũng được dùng để bảo vệ dữ liệu. Các nhà khoa học đặt máy ghi âm lâu ngày trong biển và rừng. Việc chỉ thu âm mà không để con người can thiệp như vậy gọi là quan sát âm thanh thụ động. Nếu âm thanh giả lẫn vào, dữ liệu này sẽ bị bẩn. Nếu gắn dấu vào âm thanh, ta có thể lọc âm thanh giả và giữ dữ liệu sạch.

Thứ ba là lời hứa cùng nhau đặt ra quy tắc. Chương trình MOTH của Trường Luật, Đại học New York đang đi đầu trong việc này. MOTH là nhóm xem xét quyền của những sự sống không phải con người. Dữ liệu do máy tìm ra phải gặp gỡ trí tuệ lâu đời của người bản địa, những người đã gìn giữ biển. Kiến thức về tự nhiên mà người bản địa có được qua quan sát lâu dài gọi là tri thức sinh thái truyền thống. Không được đặt quy tắc chỉ dựa trên lợi ích của con người.

Ba lời hứa này xuất phát từ cùng một tấm lòng. Đó là tấm lòng không tùy ý điều khiển đối phương chỉ vì có sức mạnh mới. Con người từ lâu đã coi thiên nhiên chỉ như một công cụ. Bây giờ là lúc nhìn lại thói quen đó. Sức mạnh mới mang tên trí tuệ nhân tạo không phải là điều để khoe. Nó là một trách nhiệm nặng nề.

Nếu hiểu được lời nói của cá nhà táng, ta cũng có thể nghĩ đến các quyền mới. Các nhà khoa học nói đến quyền của cá voi không bị quấy rối. Họ cũng nói đến quyền được giữ văn hóa riêng. Năm 2024, các lãnh đạo bản địa ở New Zealand và Quần đảo Cook tuyên bố rằng nên xem cá voi là những tồn tại có quyền như con người. Điều đó có nghĩa là cá voi không phải là đồ vật để con người tùy ý sử dụng.

Cuốn sách này kể câu chuyện dịch ngôn ngữ của sự sống trên Trái Đất. Trí tuệ nhân tạo đã lần lượt giải mã tiếng click của cá nhà táng, tiếng hót của chim và tiếng rung trầm của voi. Dịch ngôn ngữ không phải là việc thắng đối phương. Đó là việc hiểu đối phương. Trí tuệ nhân tạo là cây cầu dịch ngôn ngữ của sự sống trên Trái Đất. Khi bước qua cây cầu ấy một cách cẩn thận, chúng ta có thể tạo nên một Trái Đất nơi con người sống cùng các sự sống láng giềng. Việc dịch ngôn ngữ không kết thúc ở đây. Nó chỉ vừa mới bắt đầu.

Tài liệu căn cứ liên quan

- [1] Inside Climate News. (2025, October 29). AI is decoding whales' communications. Could that be a turning point in the push for their rights? Inside Climate News. <https://insideclimatenews.org/news/29102025/ai-sperm-whale-communications-legal-rights/>
- [2] MOTH - More Than Human Life. (2026). MOTH x CETI: Exploring how understanding sperm whale communications can be a force for good. mothrights.org. <https://mothrights.org/news/moth-x-ceti-exploring-how-understanding-sperm-whale-communications-can-be-a-force-for-good/>
- [3] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. <https://doi.org/10.1126/science.adg7314>

Ủng hộ cuốn sách Thư viện AI này

Nếu cuốn sách này hữu ích với bạn, bạn có thể ủng hộ các bản dịch tiếp theo, các ấn bản công khai và những dự án tri thức mở về trí tuệ nhân tạo qua PayPal.

PayPal

<https://paypal.me/kimkjkj>



Quét mã QR hoặc mở đường dẫn ở trên.

Luật sư Kim Kyung-jin