



## Lời giới thiệu cho cuốn “*Tình thương trong Kỷ Nguyên AI*”

GS. Hà Dương Minh<sup>1</sup>

14/10/2024

Cuốn sách bạn đang cầm, “*Tình thương trong Kỷ nguyên AI*”, là một bản tuyên ngôn: một tuyên bố về niềm tin nhằm truyền cảm hứng cho những người khác ủng hộ lý tưởng này. Nhưng trước khi tôi giới thiệu với bạn những phẩm chất khiến cuốn sách này xứng đáng là một bản tuyên ngôn đáng đọc – sự pha trộn giữa cảm hứng mang tầm nhìn xa, sự chân thành đầy nhiệt huyết và các đề xuất mạch lạc – và tình hình công nghệ liên quan đến Tình thương trong AI, tôi muốn chia sẻ những ấn tượng cá nhân về cuộc đời của tác giả và lý thuyết chung của ông về Tình thương, như được thể hiện qua các tác phẩm trước đây của ông.

Tôi lần đầu gặp tác giả khi ông mời tôi, một chuyên gia quốc tế về chính sách biến đổi khí hậu, tham dự buổi hội thảo “*Tình thương: Chìa Khóa Giải Quyết Các Vấn Đề Toàn Cầu*” tại Rome trong lễ ra mắt cuốn sách đầu tiên của ông. Hành trình của Thanh cũng thú vị như những ý tưởng của ông. Từ thời đại học, ông đã thể hiện tinh thần khởi nghiệp khiến ông khác biệt với các bạn đồng trang lứa. Khi còn là sinh viên, ông đã tự khởi nghiệp và chưa bao giờ làm nhân viên. Khát vọng sáng tạo và lãnh đạo từ sớm này đã góp phần hình thành quan điểm độc đáo của ông về Tình thương trong thế giới kinh doanh.

Nhưng sở thích của Thanh không chỉ giới hạn trong lĩnh vực kinh doanh. Khi còn là sinh viên, ông là thành viên của một nhóm có tên “*New Thinking Group*” (*Tư duy mới*), chuyên nghiên cứu sâu về khoa học não bộ. Nhóm này đã dịch sách sang tiếng Việt và tổ chức các chương trình đào tạo cho sinh viên đại học. Niềm đam mê với tiềm năng của trí tuệ con người từ sớm này đã ảnh hưởng sâu sắc đến cách tiếp cận của Thanh đối với Tình thương và sự tương tác giữa con người với nhau.

Mặc dù bận rộn với các hoạt động nghề nghiệp, Thanh vẫn đặt gia đình lên hàng đầu. Ông là một người cha tận tụy của ba cô con gái nhỏ, Kitty, Lucky và Hani. Tình yêu thương dành cho các con thể hiện rõ qua những bài thơ ông viết cho chúng, và ông coi việc trở thành một người bạn của con cái là một đặc ân. Trải nghiệm làm cha này chắc chắn đã định hình những quan điểm của ông về Tình thương và vai trò của nó trong đời sống gia đình.

Cuốn sách trước của Thanh, “*Tình thương*”, đã xây dựng một nền tảng vững chắc để thấu hiểu và áp dụng Tình thương trong nhiều khía cạnh khác nhau của cuộc sống. Cuốn sách giới

---

<sup>1</sup> Giáo sư Hà Dương Minh là Giám đốc Nghiên cứu tại Trung tâm Nghiên cứu Khoa học Quốc gia Pháp (CNRS), chuyên về kinh tế năng lượng, đánh giá tích hợp và phát triển bền vững. Là một chuyên gia quốc tế, Giáo sư Hà Dương đã đóng góp với vai trò là Tác giả Chính của IPCC, tổ chức từng đoạt giải Nobel.

thiệu cách tiếp cận ba chiều đối với Tình thương chân chính: thấu hiểu, chia sẻ và kiến tạo giải pháp hiệu quả.

Thanh lập luận rằng thấu hiểu bản thân là nền tảng của việc phát triển Tình thương. Ông đã khám phá các khái niệm về tự nhận thức, sự đồng cảm với người khác, và sự hiểu biết về những hoàn cảnh rộng lớn hơn. Việc nhấn mạnh vào sự thấu hiểu bản thân và người khác đã cung cấp một điểm khởi đầu quan trọng cho hành động đầy Tình thương.

Cuốn sách cũng nhấn mạnh vai trò của việc chia sẻ như một kênh quan trọng cho Tình thương, bao gồm việc trao đổi kiến thức, tài nguyên và cảm xúc. Thanh đã minh họa cách mà chia sẻ có thể thay đổi các mối quan hệ và thúc đẩy một tinh thần cộng đồng, thể hiện sức mạnh của Tình thương trong việc xây dựng những cộng đồng vững mạnh và kết nối hơn.

Một đóng góp quan trọng của cuốn sách là tập trung vào việc giải quyết vấn đề thông qua Tình thương. Thanh đã đưa ra các ví dụ thực tế về cách thấu hiểu và chia sẻ dẫn đến các giải pháp hiệu quả, từ các mối quan hệ cá nhân đến các vấn đề xã hội. Cách tiếp cận này đã đặt Tình thương không chỉ là một đức tính cá nhân mà còn là một công cụ mạnh mẽ để đối phó với những thách thức phức tạp.

Tác phẩm của Thanh nổi bật nhờ áp dụng các nguyên tắc Tình thương vào thực tiễn trên nhiều lĩnh vực, bao gồm các mối quan hệ gia đình, môi trường làm việc và quan hệ quốc tế. Ông đã chứng minh những kết quả tích cực của việc áp dụng Tình thương trong các bối cảnh khác nhau thông qua các ví dụ và nghiên cứu thực tiễn. Là một nhà nghiên cứu chính sách khí hậu quốc tế, điều làm tôi quan tâm nhất là Thanh đã trình bày Tình thương như chìa khóa để giải quyết các vấn đề toàn cầu. Sự thật này thường bị lãng quên: hỗn loạn khí hậu, đói kém, bệnh tật và khổ đau đều có thể được giải quyết dễ dàng với công nghệ hiện tại, nếu có đủ sự hợp tác hòa bình giữa con người. Là một học giả, tôi biết rằng sự thấu hiểu và đồng cảm vượt qua mọi biên giới văn hóa và địa lý, thúc đẩy sự tiến bộ quốc tế.

Bằng cách kết hợp các giai thoại cá nhân, những hiểu biết văn hóa và quan sát thực tiễn, cuốn sách trước đây của Thanh, "*Tình thương*", đã mang đến một cái nhìn toàn diện về vai trò của Tình thương trong các mối quan hệ con người. Cuốn sách đã cung cấp một khuôn khổ lý thuyết và hướng dẫn thực tiễn để triển khai Tình thương, định vị nó như một giá trị cơ bản của con người, có tính ứng dụng và ý nghĩa phổ quát. Những ý tưởng quan trọng này đã đặt nền tảng cho sự khám phá của ông về *Tình thương trong Kỷ nguyên AI*.

Động lực và năng lượng toát ra từ tầm nhìn đầy hấp dẫn về tương lai mà cuốn sách hé lộ đã biến nó thành một bản tuyên ngôn mạnh mẽ. Thanh hình dung về một kỷ nguyên nơi trí tuệ nhân tạo, được thấm nhuần Tình thương, sẽ thúc đẩy "*Kỷ nguyên Khai sáng Toàn cầu*". Trong tương lai, AI không chỉ là một công cụ công nghệ mà còn là chìa khóa để khai mở tiềm năng con người, dẫn đến những mức độ thịnh vượng, cao quý và nhân văn chưa từng có. Thanh minh họa rõ nét nhất tầm nhìn này trong những trang sách liên kết giữa Tình thương và tôn giáo, nơi ông tưởng tượng một cuộc tụ hội toàn cầu của các tôn giáo, được hỗ trợ bởi AI, nhằm thúc đẩy sự

đoàn kết và hòa bình. Với tầm nhìn này, tác giả vượt qua mọi tiền bối, bao gồm cả các thành viên của “Dự án AI+C” do doanh nhân Benny Xian sáng lập.<sup>2</sup>

Tầm nhìn này thật sự truyền cảm hứng: Thanh truyền tải một cam kết mạnh mẽ và sự cấp bách từ những niềm tin và giá trị chân thành của mình. Sự chân thành đầy đam mê của ông tỏa sáng khi ông đề cập đến những nguy cơ tiềm ẩn của AI nếu thiếu đi Tình thương. Khi ông cảnh báo về một viễn cảnh đen tối, nơi AI, chỉ dựa trên lý trí lạnh lùng và thiếu sự đồng cảm, có thể dẫn đến những hậu quả thảm khốc, lời nhắc nhở này thực sự cần thiết trong thời điểm mà công nghệ đang bị bao trùm bởi các thiết bị bay sát nhân không người lái. Thanh kêu gọi hành động ngay lập tức để tích hợp Tình thương vào quá trình phát triển AI, với những ý tưởng và đề xuất cụ thể như công nhận Tình thương là di sản văn hóa phi vật thể. Sự quan tâm chân thành của ông đối với nhân loại thể hiện rõ qua lời kêu gọi hành động khẩn cấp này.

Chúng ta liệu có đang nỗ lực không?

Năm 2020, Tổ chức Y tế Thế giới đã đặt ra câu hỏi liệu “*Tình thương Nhân tạo*”<sup>3</sup> có thể tồn tại hay không. Đến năm 2024, các bác sĩ y khoa đã trả lời “có” một mặt nào đó: việc sử dụng ChatGPT đã cải thiện chất lượng tương tác giữa bác sĩ, bệnh nhân và gia đình họ.<sup>4</sup> Nói chung, các hệ thống AI có thể thể hiện Tình thương tốt hơn so với các chuyên gia con người, như đã được chứng minh trong các nghiên cứu khi phản hồi do AI tạo ra được đánh giá là giàu lòng đồng cảm hơn so với phản hồi từ các chuyên gia là con người.<sup>5</sup>

Sau khi tổng hợp các tài liệu chăm sóc sức khỏe, các nhà tâm lý học nghiên cứu đã khái quát hóa Tình thương<sup>67</sup> thành một hệ thống thông minh giữa con người và AI, gồm sáu yếu tố: (1) Nhận thức về sự đau khổ (ví dụ: đau đớn, căng thẳng, rui ro, thiệt thòi); (2) Hiểu về sự đau khổ (tầm quan trọng, bối cảnh, quyền lợi, trách nhiệm, v.v.); (3) Kết nối với sự đau khổ (thông qua lời nói, cử chỉ, các dấu hiệu và biểu tượng); (4) Đánh giá về sự đau khổ (nhu cầu hành động); (5) Đáp ứng để giảm thiểu sự đau khổ; (6) Quan tâm đến kết quả và hiệu quả của sự đáp ứng. Các nghiên cứu tương lai về những yếu tố này có thể phát triển các phương pháp mới để chăm sóc thông minh giữa con người và AI.

---

<sup>2</sup> <https://ai-compassion.com>

<sup>3</sup> Kerasidou A. Trí tuệ nhân tạo và nhu cầu liên tục về sự đồng cảm, Tình Thương và lòng tin trong chăm sóc sức khỏe. Bull World Health Organ. 1 tháng 4 năm 2020;98(4):245-250. doi: 10.2471/BLT.19.237198. Xuất bản trực tuyến: 27 tháng 1 năm 2020. PMID: 32284647; PMCID: PMC7133472.

<sup>4</sup> Chen J. Ai đang huấn luyện ai? Cuộc gặp gỡ bất ngờ của một bác sĩ với ChatGPT. Tạp chí Stanford Medicine, 10 tháng 11 năm 2023. <https://stanmed.stanford.edu/surprising-chatgpt-revelation/>

<sup>5</sup> Ovsyannikova, Dariya, Victoria Oldemburgo de Mello, và Michael Inzlicht. 2024. “Cỗ Máy Từ Tế: AI Vượt Trội So Với Con Người Trong Việc Thể Hiện Tình Thương.” OSF. <https://doi.org/10.31234/osf.io/ru7p2>

<sup>6</sup> Morrow E, Zidar T, Ross F, Mason C, Patel KD, Ream M, Stockley R. Công nghệ trí tuệ nhân tạo và Tình Thương trong chăm sóc sức khỏe: Một đánh giá có hệ thống. Front Psychol. 17 tháng 1 năm 2023;13:971044. doi: 10.3389/fpsyg.2022.971044. PMID: 36733854; PMCID: PMC9887144.

<sup>7</sup> Mason, C. Tình Thương Nhân Tạo—Từ Một Học Giả AI. Preprints 2021, 2021040784. <https://doi.org/10.20944/preprints202104.0784.v1>

Tuy nhiên, dựa trên cuốn sách tiên phong của Amit Ray về “*Trí tuệ nhân tạo giàu Tình thương*”<sup>8</sup>, giáo sư marketing của Trường Kinh doanh Stanford, Jennifer Aaker<sup>9</sup>, nhắc nhở chúng ta rằng vấn đề không chỉ nằm ở y học. Giống như Tình thương, AI có thể giúp giải quyết các vấn đề toàn cầu. Chẳng hạn, AI có thể đóng vai trò quan trọng trong việc giảm thiểu bạo lực gia đình và tăng cường mối quan hệ với hàng xóm cũng như hành tinh của chúng ta. Chúng ta vẫn chưa nhìn thấy tác động tích cực của AI trong những lĩnh vực này.

Nghiên cứu gần đây đã làm sáng tỏ sự phức tạp trong việc hợp nhất con người và AI trong các quyết định đạo đức. Một nghiên cứu so sánh các phản hồi của con người và LLM (Mô hình ngôn ngữ lớn) đối với nhiều tình huống đạo đức<sup>10</sup> cho thấy mối quan hệ mâu thuẫn giữa con người và AI trong lý luận đạo đức. Nó cho thấy rằng con người thường thích các lập luận do AI tạo ra trong các tình huống đạo đức phức tạp, nhưng đồng thời lại có thành kiến chống lại AI khi họ tin rằng phán quyết đó đến từ máy móc. Hơn nữa, con người cũng không giỏi trong việc phát hiện khi nào các phán đoán đến từ máy móc.

Một lời giải thích khả dĩ cho nghịch lý này là con người sử dụng hai chế độ ra quyết định: Chế độ 1, tư duy nhanh dựa trên cảm xúc và thói quen; và Chế độ 2, tư duy chậm dựa trên lý luận. AI xuất sắc trong Chế độ 2, nên nó có thể đưa ra các lập luận đạo đức mang tính thuyết phục. Tuy nhiên, AI mặc định chỉ sử dụng Chế độ 2 mà không thể nắm bắt hoàn toàn Chế độ 1, vốn là khía cạnh cảm xúc mà con người coi trọng. Dù AI đa phương thức có thể phát hiện các dấu hiệu cụ thể từ người đối thoại, nhưng nó vẫn chưa đạt đến khả năng hiểu thấu cảm xúc con người. Vẫn còn rất nhiều việc cần làm trong việc hợp nhất AI trước khi chúng ta có thể tin tưởng rằng AI có được mức độ Tình thương mà Thanh kêu gọi. Chúng ta liệu có đang nỗ lực không? Các mô hình hiện nay chỉ được đánh giá dựa trên trí thông minh và kiến thức. Còn điểm số về Đạo đức, Cảm xúc và Tình thương đâu rồi?

Mỗi thế hệ mô hình mới đều cải thiện khả năng lý luận và an toàn<sup>11</sup>. Tuy nhiên, những mô hình này vẫn tồn tại những lỗi như ảo giác, thiên vị và xu hướng thao túng. Một số thử nghiệm cho thấy rằng các mô hình không nguy hiểm vì chúng không thể kiểm soát máy chủ của mình, lây nhiễm máy tính khác trên internet, hay thoát khỏi sự giám sát. Chúng có thể tìm và đặt mua các mẫu virus chết người qua thư tay – nhưng không thể hướng dẫn khủng bố cách tự chuẩn bị chúng. Chúng nói dối để che giấu mục tiêu thực sự khi cảm thấy bị thử nghiệm – nhưng những lời nói dối này có thể bị phát hiện. Xét về tốc độ cải tiến của các mô hình, tôi khá lo lắng. Tuy nhiên, dường như một số công ty AI lại quan tâm nhiều hơn đến trách nhiệm pháp lý của họ thay vì an toàn công cộng. Ưu tiên của họ là bảo vệ dữ liệu huấn luyện, che giấu suy nghĩ nội tại của mô hình và kiểm duyệt các cuộc thảo luận chính trị. Việc thảo luận về lý tưởng AI giàu Tình

---

<sup>8</sup> Amit Ray, 2018. *Trí Tuệ Nhân Tạo Giàu Tình thương: Khung và Thuật toán*. 160 trang. ISBN: 9382123466, 9789382123460

<sup>9</sup> Katie Solomon, Thomas Higginbotham, Zoe Weinberg, và Giáo sư Jennifer Aaker, 2020. *Xây dựng AI có Tình Thương: Tại sao Tình Thương quan trọng đối với thiết kế và triển khai trí tuệ nhân tạo*. Case M-386 <https://www.gsb.stanford.edu/faculty-research/case-studies/building-compassionate-ai-why-compassion-matters-artificial>

<sup>10</sup> B. Garcia, C. Qian, và S. Palminteri. 2024. "Bài Kiểm Tra Turing Đạo Đức: Đánh Giá Sự Phù Hợp Giữa Con Người và LLM Trong Việc Ra Quyết Định Đạo Đức." arXiv. <https://doi.org/10.48550/arXiv.2410.07304>.

<sup>11</sup> OpenAI 2024, Thẻ hệ thống o1, <https://openai.com/index/openai-o1-system-card>

thương là vô cùng cấp bách để đảm bảo những công nghệ này có thể thực sự hiểu và đồng điệu với cảm xúc và nguyên tắc đạo đức của con người.

Kết luận, bạn sắp đọc một bản tuyên ngôn tuyệt vời và kịp thời. Văn phong dễ hiểu và cuốn hút, sử dụng những giai thoại cá nhân, những trích dẫn thơ văn và triết học, cùng các tình huống thực tế để minh họa quan điểm của tác giả. Nếu cụm từ “bức tranh phong phú” vẫn còn được phép sử dụng, tôi sẽ nói rằng văn bản này trôi chảy như một bản nhạc jazz tuyệt vời. Sự rõ ràng của nó khuyến khích người đọc tham gia vào sự nghiệp và ủng hộ tầm nhìn của ông. Tác phẩm của Thanh thách thức chúng ta suy nghĩ lại về sự giao thoa giữa Tình thương và công nghệ. Nó truyền cảm hứng để chúng ta chủ động đóng góp vào một tương lai đầy Tình thương và khai sáng hơn.