Trí tuệ nhân tạo – Tôi không sợ công cụ, nhưng sợ con người

Tác giả: Alexander Demling phỏng vấn Amitai Etzioni và Oren Etzioni – Der Spiegel số 28/2023

Người dịch: Tôn Thất Thông

Tháng Tám 2023

Nhà nghiên cứu Oren Etzioni với cha Amitai trong căn hộ ở Washington DC.:
"Tôi có may mắn,rằng cha tôi vô cùng tò mò

DĐKP: *Xin giới thiệu cùng độc giả một cuộc tranh luận về trí tuệ nhân tạo giữa người cha là Amitai Etzioni, nhà xã hội học tiếng tăm người Mỹ và người con trai là Oren Etzioni, nhà nghiên cứu tiên phong về trí tuệ nhân tạo. Cuộc tranh luận ôn hòa này làm chúng ta nhớ lại cuộc cách mạng khoa học vũ bão cách đây 4 thế kỷ. Nó tạo ra nhiều phồn vinh và vô vàn tiến bộ, đồng thời cũng để lại nhiều vấn nạn chưa có lời giải cho con người và thiên nhiên. Liệu kinh nghiệm đó có giúp chúng ta giảm bớt hệ lụy của trí tuệ nhân tạo cho nhiều thế hệ về sau?*

CUỘC TRANH LUẬN: Amitai Etzioni là một trong những nhà xã hội học quan trọng nhất ở Hoa Kỳ, Oren Etzioni là nhà nghiên cứu trí tuệ nhân tạo. Trong cuộc phỏng vấn cuối cùng trước khi qua đời, người cha thảo luận về sự nguy hiểm và cơ hội của trí tuệ nhân tạo với con trai mình – người cha khá bi quan, người con trai thì khá lạc quan.

Amitai Etzioni là một trong những trí thức Hoa Kỳ có ảnh hưởng nhất trong thế kỷ 20, con trai ông là Oren, 59 tuổi, là người tiên phong trong nghiên cứu về trí tuệ nhân tạo (AI). Cả hai đã trả lời nhiều cuộc phỏng vấn trong cuộc sống của họ, nhưng đây là lần đầu tiên họ tham gia phỏng vấn chung với nhau. Vào giữa tháng năm, họ tiếp chúng tôi trong căn hộ của người cha gần khách sạn Watergate ở Washington, DC, và thảo luận với nhau khoảng hai giờ. Nó trở thành một cuộc trò chuyện về các vấn đề trong cuộc sống của họ: về cuộc cách mạng ChatGPT và hậu quả của nó đối với xã hội của chúng ta. Điều mà không ai biết trước vào thời điểm đó: Đó là cuộc phỏng vấn quan trọng cuối cùng của Amitai Etzioni. Nhà nghiên cứu xã hội, sinh ra ở Cologne, qua đời hai tuần sau đó ở tuổi 94.

Người cha: „Trí tuệ nhân tạo sẽ phá hủy sự thật“

Người con: „Tôi không sợ công cụ, nhưng sợ con người“

***

SPIEGEL: Câu hỏi đầu tiên dành cho cả hai ông, cha và con của gia đình Etzioni. Hai ông có sợ trí tuệ nhân tạo không?

Oren Etzioni (người con): Trí tuệ nhân tạo không phải là sinh vật giống như yêu tinh hay quái vật của Frankenstein, mà chỉ là một công cụ, mặc dù là một công cụ rất mạnh. Tôi không sợ bản thân công cụ, tôi sợ cách mọi người sử dụng nó, thí dụ các tác nhân độc hại hoặc các quốc gia bất hảo chẳng hạn.

Amitai Etzioni (người cha): Vào một thời điểm nào đó trong cuộc Cách mạng Công nghiệp, con người đã mất kiểm soát đối với tiến bộ công nghệ và kinh tế. Từ đó, chúng ta cố gắng giành lại quyền kiểm soát đó. Trí tuệ nhân tạo làm cho cuộc chạy đua này thậm chí còn nhanh hơn và khó thắng hơn. Đồng thời, các phong trào xã hội mới đang nổi lên vốn dĩ không muốn sống trong một xã hội, trong đó các lực lượng nói trên thống trị chúng ta.

SPIEGEL: Hai ôngthấy những phản kháng chống lại trí tuệ nhân tạo hay sao?

Amitai Etzioni: Ít nhất là có triệu chứng báo trước: Đối với tôi, họ bao gồm những người từ chối trở lại văn phòng sau đại dịch. Họ nói: Tôi muốn dành nhiều thời gian cho gia đình, chơi

trò chơi điện tử thay vì đánh mất thời giờ trong nạn kẹt xe. Cuối cùng, AI chỉ là phần nổi của tảng băng chìm cho một cuộc xung đột lớn hơn nhiều trong xã hội của chúng ta: Con người hay máy móc sẽ được đặt vào trung tâm?

Oren Etzioni: Ở chỗ này, tôi và bố tôi không hoàn toàn đồng ý. AI trong một trung tâm máy tính làm âm mưu chống lại loài người và hành động sát nhân điên cuồng, tôi nghĩ đó chỉ là khoa học viễn tưởng. Công nghệ chắc chắn không phải là trung lập, nhưng cuối cùng con người đưa ra quyết định. Và những người quyền lực vận hành những công cụ này không phải là chính trị gia, mà là những doanh nhân công nghệ. Mark Zuckerberg và Elon Musk có sức mạnh đó, không phải AlphaGo (chương trình AI đã đánh bại nhân vật số một thế giới trong trò chơi cờ vây năm 2017 – BBT).

SPIEGEL: Đồng nghiệp của ông, Geoffrey Hinton, người đặt cơ sở khoa học cho sự bùng nổ AI hiện nay đã bắt đầu sợ hãi sáng kiến của ông ta.

Oren Etzioni: Đối với các mô hình ngôn ngữ lớn như GPT-4 bạn cần phân biệt giữa trí thông minh và tính tự quyết. Theo trực giác, chúng ta thường ném hai thứ đó lại với nhau vì con người thông minh và tự chủ. GPT-4 rất tinh vi và có phần thông minh, nhưng nó không có ý chí riêng. Nó cần nhận được bao nhiêu quyền tự chủ, đó là một câu hỏi chính trị, không phải là vấn đề công nghệ. Con người quyết định điều đó.

Amitai Etzioni: Tôi phải phản bác điều đó. Mọi loại thuốc mới ở Hoa Kỳ trước tiên phải được [cơ quan thẩm định] FDA chấp thuận. Nhưng nếu một vài chàng trai ở San Francisco đêm nay phát minh ra công nghệ để tạo ra tin giả tốt hơn, thì chúng ta phải chấp nhận việc chung sống với nó – ngay cả khi điều đó đi ngược lại các giá trị và ý chí chính trị của chúng ta. Chúng ta thậm chí không thể cấm họ dù có muốn chăng nữa.

SPIEGEL: Tại sao?

Amitai Etzioni: Đứa trẻ đã ngã vào giếng, và công nghệ đang đứng ngoài thế giới. Ngay cả khi chúng ta có một thỏa thuận với Trung Quốc chẳng hạn, thì Nga có tham gia vào đó không? Venezuela? Yemen?

SPIEGEL: Ngay cả ông chủ OpenAI Sam Altman cũng kêu gọi phải điều tiết ngành công nghiệp của mình.

Amitai Etzioni: Ôi, đó chỉ là tự quảng cáo.

Oren Etzioni: Tôi cũng không nghĩ đó là cách tốt nhất. Chúng ta cần những quy định chính xác cho từng mục đích. Một đạo luật AI duy nhất có đủ sức điều tiết để ngăn ngừa vũ khí hạt nhân hay không? Chắc chắn là không. Tốt hơn là chúng ta thay đổi luật hiện hành về bản quyền, phát minh thuốc và tất cả những lĩnh vực còn lại để thích ứng với hoàn cảnh mới này.

Amitai Etzioni: Khi tôi giảng bài ở Đức, tôi luôn hỏi sinh viên, có ai một lần nào được một công ty hỏi về luật bảo vệ dữ liệu hay không? Họ có được chuyển tiếp dữ liệu hay không. Chưa bao có một bàn tay giơ lên. Vì luật lệ còn có nhiều kẽ hở. Nếu có những luật AI quá rộng trong cuộc sống hàng ngày, tôi sợ rằng các cơ quan có trách nhiệm thực thi chúng sẽ bị ngành công nghiệp chiếm đoạt – giống như các ngân hàng lớn của Hoa Kỳ đã làm với các cơ quan quản lý tài chính ở đây.

SPIEGEL: chúng ta hãy bàn về những nguy hiểm sinh ra bởi những người đang chiếm giữ các công cụ rất mạnh này. Hai ông sợ điều gì?

Oren Etzioni: Trước cuộc bầu cử tổng thống sắp tới. Và mọi cuộc bầu cử tiếp theo sau đó. Máy in của Gutenberg khiến việc sao chép thông tin gần như miễn phí, internet thì hoạt động toàn cầu, và bây giờ nó sẽ sản xuất thông tin [giả] miễn phí. Các cử tri có thể bị ảnh hưởng bởi những thông điệp nghe như được phát biểu bởi Elon Musk hoặc Angela Merkel. Các ứng dụng như vậy cần phải được quản lý, chứ không phải GPT-4.

Amitai Etzioni: Trí tuệ nhân tạo sẽ phá hủy sự thật. Bạn không còn có thể tin vào mắt và tai của chính mình. Đây là một thách thức đối với nền dân chủ và cộng đồng của chúng ta.

SPIEGEL: Giữa Donald Trump và ChatGPT, kẻ dối trá nào nguy hiểm hơn cho nền dân chủ?

Amitai Etzioni: Trong ngắn hạn, điều đó vẫn thuộc về Trump. Ông ấy thực sự có cơ hội để trở thành tổng thống một lần nữa . ChatGPT có thể sẽ không. Nhưng về lâu dài AI nguy hiểm hơn.

Oren Etzioni: Một nhà mị dân lôi cuốn, người có thể thao túng người khác cũng nguy hiểm hơn về lâu dài. ChatGPT không viết những bài phát biểu sôi nổi, ít nhất là đầu ra của nó cũng chỉ ở vị trí khá trung bình. Nhưng nó có thể kích hoạt vô số chatbot thân thiện với Trump trên Facebook và Twitter cùng một lúc. Điều đó đặt ra mối nguy hiểm thực sự: Donald Trump có thể trả tiền cho một triệu Mini-GPT.

Amitai Etzioni: Những người sáng lập nước Mỹ sợ đám đông bị quyến rũ bởi một nhà lãnh đạo lôi cuốn. Vì vậy, họ thành lập một nền dân chủ gián tiếp: Mọi người được phép bỏ phiếu hai năm một lần, giữa các cuộc bỏ phiếu, họ giữ im lặng. Công nghệ thách thức hệ thống này vì giờ đây người lãnh đạo có thể tiếp cận trực tiếp với quần chúng bất cứ lúc nào. Trump đã có thể làm điều này với Twitter và ông ấy sẽ hành động hiệu quả hơn nữa với chatbot.

SPIEGEL: Những hình ảnh do AI tạo ra về giáo hoàng trong chiếc áo khoác ngoài màu trắng và các chương trình sao chép giọng nói từ một số mẫu bài phát biểu đang được lưu hành. Làm thế nào để chúng ta chế ngự những làn sóng thông tin giả?

Oren Etzioni: Không có thuốc chữa bách bệnh. Quy định chỉ có thể là một phần của giải pháp. Ví dụ, người ta có thể yêu cầu phải thẩm định thông tin kỹ thuật số, một loại hình mờ mà tất cả nội dung do AI sản xuất phải có. Công nghệ cho việc này đã có sẵn. Nhưng vấn đề của các cơ quan quản lý vẫn còn đó: Họ làm việc trong tầm quốc gia, nhưng vấn đề là quốc tế. Chúng ta cần những phương tiện khác.

SPIEGEL: Đó sẽ là gì?

Oren Etzioni: Giáo dục. Chúng ta phải học cách sử dụng công nghệ này. Những kẻ lừa đảo dùng hình ảnh cháu trai ngày nay có thể làm giống như cháu trai thực sự. Có một ngày, một đồng nghiệp nhờ vả tôi một chuyện qua tin nhắn. Mãi cho đến khi anh ấy hỏi về thẻ quà tặng của Apple, tôi mới biết thực hư thế nào. Với mỗi cuộc gọi, chúng ta phải tự hỏi: điều này có đáng tin hay không?

Amitai Etzioni: Này Oren, khi con còn là một cậu bé và ngồi trước TV, bố đã nói: Hãy nghĩ xem quảng cáo này đang cố gắng thao túng con như thế nào.

SPIEGEL: Rồi ông vẫn mua đồ ngọt cho con?

Oren Etzioni: (cười) Ngày nay, đó vẫn còn là một thách thức.

SPIEGEL: ChatGPT không chỉ nói dối khi người dùng ra lệnh cho nó. Ví dụ, chatbot cáo buộc một thị trưởng Úc về vụ bê bối tham nhũng mà người đàn ông đối thoại tiết lộ rằng tự mình tìm ra sự thật. Tại sao AI thường tạo ra những điều hoàn toàn vô nghĩa với nhiệt tình cao độ?

Oren Etzioni: Thật là phức tạp để hiệu chỉnh sự tự nhận thức của mô hình vì nó không có khái niệm về thực tế. Khi ChatGPT quyết định diễn đạt bầu trời có màu xanh lam hay xám, nó sẽ tính toán với các xác suất rất ít khác nhau. Theo thời gian, mô hình phải học, khi nào thì nên nói 'Tôi không biết'. Một điều không được quên: Công nghệ này còn rất trẻ. Sẽ có những bước đi sai lầm. Hãy nghĩ về công nghệ hạt nhân. Tôi nghĩ bây giờ mọi người đều đồng ý rằng Hiroshima và Nagasaki là những bước đi sai lầm.

Amitai Etzioni Chúng ta cần tìm cách dạy đạo đức cho các mô hình. Không phải là dạy những gì một triết gia đặt ra như vậy, mà là những giá trị mà các nhóm tôn giáo hoặc xã hội khác nhau thực sự sống với chúng. Các mô hình khác nhau có thể phản ánh sự đồng thuận khác nhau của họ về các giá trị.

SPIEGEL: Elon Musk bị cáo buộc là muốn phát triển TruthGPT vì ông cho rằng ChatGPT quá đúng về mặt chính trị.

Oren Etzioni: Vâng, cuối cùng những mô hình này chủ yếu là những tấm gương. Khi chúng ta giơ tấm gương này lên trước xã hội của mình, chúng ta sẽ thấy những người Cộng hòa, Dân chủ và tất cả những người không đảng phái.

.

SPIEGEL: Điều này cũng là do các mô hình ngôn ngữ lớn có lẽ đã được đào tạo với vô số bài đăng từ Twitter và Reddit – tức là nơi không nhất thiết mọi người gặp nhau để thể hiện mặt tốt nhất của họ.

Oren Etzioni: Là những lập trình viên, chúng tôi nói: »Rác vào, Rác ra« – nếu bạn nạp rác vào một hệ thống, nó cũng sẽ phun ra rác. Nhưng vấn đề có thể giải quyết được: mọi người liên tục đánh giá đầu ra của mô hình và do đó, theo thời gian, dạy cho nó các giá trị của chúng. Nhưng không thể biến ChatGPT thành một khuôn khổ tối thượng về giá trị, bởi vì chúng ta với tư cách là một xã hội cũng chưa đồng thuận về nó.

SPIEGEL: Công ty phân tích dữ liệu lớn Palantir cung cấp cho quân đội một nền tảng AI giúp phát triển các kế hoạch chiến đấu và phân tích các mục tiêu của kẻ thù. Ông chủ của Palantir, Alex Karp gọi AI sáng tạo là một »cuộc cách mạng, để nâng những con tàu và đánh chìm tàu«.

Amitai Etzioni: Hãy tưởng tượng rằng: Công nghệ AI kết hợp với vũ khí hạt nhân và công nghệ vũ trụ có thể xác định mọi điểm trên Trái đất. Nhân loại phải tìm ra một thỏa thuận quốc tế nói rằng: Với những vũ khí tấn công có sức mạnh hủy diệt, trí tuệ nhân tạo không được phép đơn độc đưa ra quyết định không thể đảo ngược. Con người phải có quyết định cuối cùng ở đây. Nhưng chúng ta đã ở trong một cuộc chiến gián tiếp với Trung Quốc. Do đó, tôi không đánh giá cơ hội của một thỏa thuận như vậy là rất cao. Điều đó làm tôi lo lắng.

Oren Etzioni: Bạn phải phân biệt rõ ràng giữa vũ khí phòng thủ và tấn công. Vũ khí tấn công tự động đe dọa an ninh của chúng ta. Nhưng ngược lại, vũ khí phòng thủ thông minh, thậm chí là vũ khí tự động, có khả năng cứu mạng sống. Chúng bảo vệ các trường mẫu giáo và bệnh viện trong cuộc xung đột giữa Israel và Palestine hay ở Ukraine.

SPIEGEL: Trước đây vài năm,một số người ở Google bao gồm cả các chuyên gia AI đã đình công, vì họ phải làm việc theo hợp đồng từ quân đội Mỹ.

Oren Etzioni: Tôi không thể hiểu nổi những đồng nghiệp này. Tôi không hài lòng khi ngày càng có nhiều hệ thống vũ khí mạnh hơn được phát triển. Nhưng chúng nên ở trong tay chúng ta thay vì trong tay các chế độ toàn trị. Xã hội phải xác định nơi nào là ranh giới của chúng ta. Các chuyên viên công nghệ như chúng tôi có trách nhiệm tham gia vào cuộc thảo luận này với một góc nhìn tinh tế, thay vì vùi đầu vào cát.

SPIEGEL: Sự mê hoặc toàn cầu với ChatGPT đã kích hoạt một cuộc chạy đua giữa Microsoft, Google và Meta cho các mô hình AI tốt nhất. Có phải công nghệ này quá nguy hiểm để giao khoán cho các doanh nhân?

Amitai Etzioni: Chúng ta sẽ thấy sự tập trung của cải thậm chí còn lớn hơn vào tay một số ít người. Việc cuối cùng Microsoft hay Google thắng thế không phải là điều quyết định. Điều quan trọng là liệu toàn bộ ngành công nghiệp có thể ngăn cản Quốc hội Hoa Kỳ trong việc đưa ra các quy tắc hiệu quả cho nó hay không.

Oren Etzioni: Sự thật cơ bản của công nghệ vẫn còn: nó đã giải thoát cho hàng trăm triệu người thoát khỏi nghèo.

Amitai Etzioni: Câu hỏi đặt ra là liệu nó có tiếp tục như thế này hay không. Trong vài thế kỷ qua, công nghệ đã đưa chúng ta ra khỏi thời Trung cổ để biến thành một xã hội giàu có. Những thế kỷ tiếp theo sẽ mang lại điều gì? Chỉ có nhiều tiện ích hơn?

SPIEGEL: Ông có thấy trí tuệ nhân tạo chủ yếu là mối đe dọa đối với xã hội của chúng ta không?

Amitai Etzioni: Không hề. Cảm xúc giả vờ có thể hữu ích trong việc xây dựng máy tính đồng cảm. Cho đến nay, chúng ta đã để người già cô đơn trong nhà dưỡng già và để họ ở một mình ở đó vì dễ quản lý hơn. Có phải là tốt hơn, nếu có máy tính biết nói “Tôi rất vui được lắng nghe bạn”? Vâng, tôi thích điều đó lắm.

SPIEGEL: Người bạn đồng hành ảo cho đến nay vẫn là chuyện khoa học viễn tưởng.

Amitai Etzioni: Không có gì ảnh hưởng mạnh mẽ đến sức khỏe tinh thần của chúng ta bằng sự cô đơn. Tại sao chúng ta chỉ nên giúp đỡ mọi người khi họ hoang mang và tiến dần đến trạng thái trầm cảm? Vì lý do này, trí tuệ nhân tạo không thể bị cấm đoán một cách đơn giản. Chúng ta cần suy nghĩ về sự hữu ích của các quy tắc cũng như tác hại của chúng.

SPIEGEL: Các ông đã cùng xuất bản các bài báo về trí tuệ nhân tạo. Làm thế nào để hai ông thảo luận những vấn đề này với nhau?

Oren Etzioni: Với tư cách là những học giả, chúng tôi đã quen với việc hoạt động trong khuôn khổ chật hẹp của ngành học và biệt ngữ riêng của chúng tôi. May mắn cho tôi là, cha tôi cực kỳ tò mò và đặt nhiều câu hỏi. Đôi khi chúng tôi không đồng ý, nhưng vì thế mà chúng tôi thảo luận về nó.

Amitai Etzioni: Chúng tôi có một đại gia đình và hàng năm chúng tôi cùng nhau đi nghỉ hai tuần. Chúng tôi cần thời gian này để thảo luận với nhau những gì vượt ra ngoài ranh giới của nghề nghiệp. Nhờ thế, chúng tôi có một loạt bài tiểu luận về thiết kế hệ thống trí tuệ nhân tạo, và đặt tên cho nó là Hồ sơ Marbella vì chúng được tạo ra khi đi nghỉ ở Tây Ban Nha.

SPIEGEL: Xin cảm ơn hai ông về cuộc phỏng vấn này.

Phỏng vấn và biên tập: Alexander Demling. Der Spiegel số 28/2023 – 08/07/2023, trang 16-19 (báo giấy, không có link).

https://diendankhaiphong.org