DEEP FAKE LÀ GÌ?

LÊ TẤN TÀI

Trí tuệ con người đang dần dần bị "trí tuệ nhân tạo" (AI=Artificial Intelligence) thay thế trong khá nhiều lãnh vực như chơi cờ vua, dịch thuật, y tế trị liệu, xử lý dữ liệu để tự học hỏi... Bài viết sau đây nói về một khả năng khác của AI đang làm con người hoảng sợ.

Bạn có thể nào tin cựu tổng thống Mỹ Barack Obama nói trước hàng triệu khán giả trên Internet "Donald Trump là một kẻ ngốc", hay Mark Zuckerberg, ông chủ Facebook, khoe khoang "Tôi đang toàn quyền kiểm soát dữ liệu của hàng tỷ người dùng trên toàn cầu"?
Chuyện giả mà không giả. Năm 2017, video giả mạo cựu Tổng thống Mỹ Obama được nhóm nghiên cứu đại học Washington dùng trí nhân tạo AI để ghép mặt của Cựu Tổng thống Mỹ Barack Obama với một giọng đọc giả mạo, khiến cả thế giới kinh ngạc vì độ chân thật của video. Khái niệm Deepfake bắt nguồn từ thời điểm này.

Deepfake là gì?
Deepfake (tạm dịch là "giả hình") là kết hợp của "deep learning" (học kỹ, sâu) và "fake" (giả mạo), đề cập đến các video bị thao túng, hoặc các sản phẩm công nghệ được tạo ra bởi Trí tuệ nhân tạo một cách tinh vi, cụ thể là "học sâu" (deep learning), nhằm tạo ra các hình ảnh và âm thanh bịa đặt nhưng rất giống thật. Công nghệ này được xây dựng trên nền tảng machine learning, mã nguồn mở của Google. ("Học máy" là một ngành khoa học nghiên cứu các thuật toán (algorithm) cho phép máy tính có thể học được các khái niệm (concept) như con người).

Deepfake sẽ quét video và ảnh chân dung của một người sau đó hợp nhất vào một video riêng biệt, nhờ AI thay thế các chi tiết trên gương mặt như mắt, miệng, mũi với chuyển động gương mặt, giọng nói như thật. Càng có nhiều hình ảnh gốc thì AI càng có nhiều dữ liệu để "học".
Deepfake có thể gán khuôn mặt của người này sang người khác trong video với độ chân thực đến kinh ngạc.

Ứng dụng tích cực.
Có thể nói các hình thức ứng dụng trợ lý giọng nói bắt chước các yếu tố thân mật của con người trong lời nói, bao gồm ngắt nghỉ và tín hiệu bằng lời nói như "hmmm", có tính thực tế cao, như các cuộc gọi điện thoại trực tiếp, tạo cảm giác với người đối thoại rằng họ đang nói chuyện với một người thực.

Một ví dụ khác, cho thấy sử dụng âm thanh giọng nói để tái tạo giọng nói của người thân yêu đã qua đời là một điều mà mọi người cảm thấy kết nối tốt hơn với người đã khuất. Deepfake giả giọng nói còn có một công năng khác là khôi phục giọng nói của một người khi họ bị mất giọng vì bệnh, hỗ trợ giáo dục bằng cách tái tạo âm thanh của các nhân vật lịch sử, như CereProc tạo ra một phiên bản bài diễn văn cuối cùng của cựu Tổng thống Mỹ John F. Kennedy, người đã bị ám sát năm 1963.

Video deepfake có thể làm sinh động các phòng trưng bày và bảo tàng. Đối với ngành công nghiệp giải trí, công nghệ có thể được sử dụng để cải thiện việc lồng tiếng cho các bộ phim tiếng nước ngoài, v.v… Sự phổ biến của Amazon Alexa và Google Assistant làm cho chúng ta sống thoải mái trong một thế giới hòa trộn giữa thật và giả. Các ứng dụng hứa hẹn nhất của AI đều nằm trong lĩnh vực giải trí.

Từ nhiều năm nay, các đạo diễn phim ảnh đã hao tốn rất nhiều trong việc giúp phim hoàn hảo hơn nhờ kỹ xảo đồ hoạ, hay đơn giản là cắt ghép một khung hình, cảnh vật, con người vốn dĩ không hề ở đó nhưng vẫn xuất hiện trên màn ảnh. Chuyện vào năm 2013, khi nam diễn viên Paul Walker đóng phim "Fast and Furious" qua đời vì tai nạn ô-tô sau một buổi tổ chức từ thiện. Bộ phim lúc ấy chưa đóng xong, tuy nhiên hàng triệu người trên toàn thế giới hào hứng bất ngờ khi gương mặt của anh xuất hiện trong phần phim tiếp theo ra rạp. Ngày nay, AI tái tạo lại hình ảnh của nữ diễn viên quá cố Carrie Fisher trong vai Công chúa Leia trong "Chiến tranh giữa các vì sao".

Như vậy khi một diễn viên nổi tiếng qua đời, đạo diễn chỉ cần tạo ra một người giả tiếp tục xuất hiện trong các bộ phim khác. Vấn đề đạo đức ở đây là khả năng tái sinh những người nổi tiếng có thể khiến họ trở thành con rối cho các công ty, được tái tạo để quảng cáo sản phẩm hoặc nhãn hiệu, quyền tôn trọng bản thân của các nhân vật này cần phải được xét đến.

Ứng dụng tiêu cực

Ba mươi năm trước, Photoshop xuất hiện và làm thay đổi cách con người tiếp nhận các dữ kiện vì các hình ảnh nhìn thấy có thể là sản phẩm của một quá trình cắt ghép kỳ công. Người ta nghi ngờ vào độ chân thật của hình ảnh và đặt niềm tin vào video, ghi âm vì đây là những thứ gần như không thể giả mạo. Nhưng Deepfake xuất hiện và phá vỡ thành trì của thế giới Internet.

Người ta có thể 'đưa' bất kỳ chính khách nào tới đâu, làm bất cứ điều gì, khi các video được phổ biến để hủy hoại ai đó. Việc áp dụng Deepfake là một điều thú vị, nhưng cần cảnh giác với nó. Trong thực tế, không nên tạo bất kỳ một video giả mạo nào dù chỉ đề mục đích cho vui! Nó có thể khiến chúng ta gặp những rắc rối pháp lý và ảnh hưởng đến danh tiếng của bản thân.

Salvador Dali, người áp dụng thuật toán vào ảnh của Marilyn Monroe viết: "Internet nói chung và mạng xã hội nói riêng đều là những phát minh mang tiềm năng lớn lao để phát triển vô tận, ẩn giấu nhiều chương sách mới còn chưa được khai phá nhằm mục đích chia sẻ, kết nối con người lẫn nhau trên toàn cầu. Dẫu vậy, vạn vật đều mang những thái cực đối lập song hành lập nhau, đi kèm với viễn cảnh tươi sáng vẫn luôn là những mặt tối phức tạp, trong đó có vấn nạn fake news, lừa đảo bằng tin tức giả mạo."

Những nhân vật tiếng tăm rất dễ có nguy cơ bị tấn công bằng video giả mạo. Ngay cả những người phụ nữ bình thường cũng có thể bị người xấu dùng công nghệ này tạo ra những video khiêu dâm, khỏa thân giả mạo, xúc phạm đến danh dự và phẩm giá của họ. Một khi video đã bị phát tán trên Internet thì gần như không thể ngăn chặn nổi.

Hơn nữa, rất khó để phân biệt tính thật giả của những nội dung này. Những nội dung sai sự thật sẽ hướng dẫn dư luận, làm hại đến uy tín và danh tiếng của các quan chức chính trị, nhà lãnh đạo doanh nghiệp, diễn viên, nghệ sĩ..., bào mòn niềm tin của mọi người với báo chí, cơ quan, tổ chức xã hội... Một vài ví dụ sau đây:
Đầu năm 2019, một nhóm tội phạm mạng đã lừa giám đốc điều hành công ty có trụ sở tại Anh trả cho họ 243.000 USD bằng cách sử dụng âm thanh deepfake giả giọng ông chủ của doanh nghiệp này qua điện thoại.

Tháng 6/2019, bộ trưởng chính phủ Malaysia bị cáo buộc xuất hiện trong một video quan hệ tình dục với người đồng giới. Hành vi này là bất hợp pháp ở Malaysia, dù những người ủng hộ ông tin rằng hình ảnh đó là giả mạo nhưng các chuyên gia lại không tìm thấy bằng chứng video bị cắt ghép.

Một ví dụ khác về sự kiện ở Gabon cuối năm 2018. Khi đó, Tổng Thống Ali Bongo Ondimba của nước này đã không xuất hiện trước công chúng trong vài tháng. Dư luận cho rằng tổng thống Ondimba bệnh nặng, thậm chí đã chết. Nhằm dập tắt tin đồn này, chính phủ đã công bố một video cho thấy tổng thống đọc diễn văn chức mừng năm mới.

Trong video, ông Ondimba xuất hiện trông cứng nhắc với nét mặt thiếu tự nhiên. Video lập tức gây nghi ngờ và tranh cãi trên mạng xã hội. Các nhóm chống đối khẳng

định video là sản phẩm của Deepfake và tổng thống đã qua đời. Ông Ondimba sau đó xuất hiện trở lại và tiếp tục lãnh đạo Gabon. Cho đến nay, giới chuyên gia vẫn chưa thể khẳng định đoạn video của ông có phải là giả mạo hay không?

Deepfake ngày nay hiển nhiên trở thành vũ khí hữu hiệu nhất trong chính trị. Thông tin giả mạo có thể dẫn đến biểu tình, bạo động, gây bất ổn... Chuyện gì sẽ xảy ra, nếu trên mạng bỗng dưng xuất hiện một video deepfake mô tả ứng cử viên tổng thống đang quấy rối tình dục trẻ em, hoặc một cảnh sát trưởng đang xúi giục nhân viên thực hiện hành vi bạo lực với người dân tộc thiểu số, hay những người lính có hành động tàn ác trong chiến tranh…

Một giải pháp được nhiều nước cân nhắc là đưa ra luật quy định việc tạo và phát tán nội dung Deepfake là bất hợp pháp. Vào tháng 10, 2019, California quy định rằng việc tạo hoặc chia sẻ video, hình ảnh, giọng nói của các chính trị gia bằng công nghệ Deepfake trước cuộc bầu cử là phạm luật. Tuy nhiên, giải pháp này vẫn không hiệu quả, do tính ẩn danh và không biên giới của Internet. Trong giai đoạn này, các hãng công nghệ lớn như Facebook, Google và Twitter phải hành động để hạn chế sự lan truyền của những video giả mạo.

Nghị Sĩ Marco Rubio nhận xét: "Ngày xưa, nếu muốn đe dọa Hoa Kỳ, đối thủ cần có 10 hàng không mẫu hạm, vũ khí hạt nhân và hỏa tiễn tầm xa. Ngày nay, tất cả những gì bạn cần là khả năng sản xuất một video giả mạo nhưng trông như thật, để gây ảnh hưởng tới kết quả bầu cử, đẩy nước Mỹ vào khủng hoảng và làm suy yếu chúng ta."

Thông thường, các tin xấu và gây tranh cãi luôn lan truyền rất nhanh, nhưng các tin đính chính sau đó lại ít người biết tới. Tuy nhiên, nếu mọi người có xu hướng hoài nghi mọi video họ xem, kể cả thông tin chính thống, ông Hany Farid, một trong những chuyên gia hàng đầu thế giới về deepfake, nói: "Nếu bạn không còn tin những video hay những đoạn âm thanh mà bạn xem, đó thật sự là nguy cơ an ninh quốc gia nghiêm trọng." Ông dự đoán trong tương lai gần, công nghệ Deepfake sẽ phát triển từ một hiện tượng lạ trên Internet thành một công cụ tàn phá xã hội, công kích chính trị. Ông cũng cho rằng mọi người cần có sự chuẩn bị để đối mặt với vấn đề này.

Ảnh ghép phim khiêu dâm

Ác mộng mà Deepfake mang lại là sự giả mạo ghép ảnh phụ nữ khoả thân. Ghép mặt người khác vào nhân vật phim khiêu dâm đang ngày càng phổ biến, đặt ra những câu hỏi mới về vấn đề "lạm dụng công nghệ" tại Trung cộng. Cuộc điều tra của tờ The Beijing News đã phát hiện ra nhiều dịch vụ làm việc này với giá chưa tới 1 USD.

Theo thống kê, tính đến tháng 9/2019, có 96% video deepfake chứa nội dung khiêu dâm. Có một số trang web chuyên phát những nội dung này và thu hút rất nhiều lượt xem trong suốt hai năm gần đây. Những nội dung trong đó hầu hết được tổng hợp từ những video với sự ghép mặt của những người nổi tiếng.

Ai cũng có thể là nạn nhân của Deepfake. Nếu một ngày nào đó bỗng nhiên bạn thấy mặt mình xuất hiện trong một bộ phim khiêu dâm và được lan truyền trên mạng, với tốc độ lan truyền nhanh như hiện nay thì việc một video khiêu dâm sẽ nhanh chóng đến tay bạn bè và người thân của bạn. Khi đó danh dự và mọi người sẽ nhìn và đánh giá bạn ra sao? Đây là ví dụ rõ ràng nhất cho thấy sự xuất hiện của công nghệ Deepfake đã khiến công chúng càng khó phân biệt đâu là thật giả. Và những người có ý đồ dẫn dắt dư luận sẽ cố gắng khai thác điều này, làm cho tình hình ngày càng trở nên tồi tệ hơn.Cách tạo một Deepfake.

Mọi sản phẩm sử dụng trí tuệ nhân tạo đều trải qua hai bước chính: nạp dữ liệu đầu vào, sau đó dựng lên mô hình và lựa chọn một thuật toán để liên tục xử lý, "học" từ các mô hình đó.

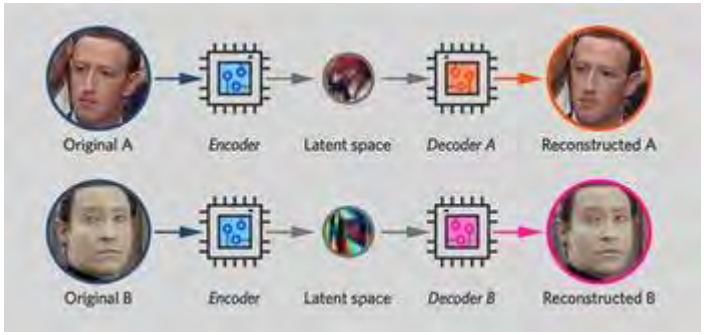
Theo Reddit, dữ liệu đầu vào để tạo nên một video ghép mặt giả mạo rất đơn giản, nó chính là những bức ảnh công khai của diễn viên có trên mạng từ Google, những clip video có sẵn từ Youtube. Cách làm và công cụ cũng có sẵn, người dùng "deepfakes" chỉ cần thực hiện vài thuật toán mã nguồn mở như Google TensorFlow hoặc Keras để cho cỗ máy "học" và ghép khuôn mặt với độ giống cao. Quá trình "học" chính là đóng góp mấu chốt của trí tuệ nhân tạo.

Nhóm nghiên cứu Samsung hợp tác với Viện Khoa học và Công nghệ Skolkovo đã phát triển được một phương thức để hoạt họa hóa các bức chân dung cổ điển, khiến chúng trở nên sống động hơn bao giờ hết, phát triển một hệ thống dựng hình mặt người sử dụng với nguồn hình ảnh tĩnh cực ít, kể cả chỉ duy nhất một tấm hình gốc cũng có thể giúp nó hoạt động và làm giả được.

Rất khó tạo một deepfake nếu chỉ sử dụng một chiếc máy tính thông thường. Hầu hết các sản phẩm deepfake được làm nên từ các máy tính cao cấp với card đồ họa (hay còn gọi là card màn hình - Graphics card) mạnh mẽ, hay cao hơn nữa là sử dụng điện toán đám mây (cloud computing).

Deep-fakes chất lượng kém dễ dàng bị phát hiện. Những khuôn mặt trên deepfake không thể chớp mắt như bình thường, vì thuật toán không bao giờ thực sự "học" về việc chớp mắt. Đồng bộ môi xấu, hoặc màu da loang lổ có thể giúp nhận ra đâu là video giả. Các chi tiết như tóc đặc biệt khó để deepfake có thể "render" (kết xuất đồ họa hay quá trình tập hợp các mô hình thành một hình ảnh) một cách mượt mà.

Đồ trang sức hay răng làm ẩu cũng là một điểm cần chú ý, hay các hiệu ứng ánh sáng kỳ lạ, chẳng hạn như chiếu sáng không nhất quán và phản chiếu trên mống mắt sẽ là một căn cứ quan trọng để phân biệt. Các nhà nghiên cứu đã giới thiệu một công cụ cho phép nhận diện các video deepfake.



Công cụ này còn có thể phân tích được những chi tiết mà mắt người không thể nhận ra, như phân tích phổ hoặc ánh sáng của bức ảnh để nhận ra vị trí khác biệt. Tuy nhiên, họ cũng thừa nhận phải liên tục phát triển để chạy đua với những kỹ thuật làm giả mới nhất. Đáng sợ hơn, Deepfake đang ngày càng được cải tiến và hoàn thiện vượt qua trí tưởng tượng của người thường, những video deepfake xuất hiện ngày càng nhiều và gia tăng với tốc độ chóng mặt trên mạng Internet.

Theo thống kê của Deep-trace, tính đến đầu năm 2019, có 7.964 video deepfake xuất hiện trực tuyến. Chỉ sau 9 tháng, còn số này đã tăng lên đến 14.678 và tất nhiên vẫn tiếp tục tăng một cách nhanh chóng.

Đạo đức khoa học

Phát minh nguyên tử năng hay Deepfake không mang lại lợi ích cho nhân loại bao nhiêu mà tai họa thì rất lớn.

Năng lượng hạt nhân là giải pháp mới để sản xuất ra điện năng so với các nguồn năng lượng khác và được cho khá an toàn, nhưng trong lịch sử đã chứng kiến nhiều sự cố về các nhà máy điện hạt nhân. Chất thải của nhà máy nguyên tử rất độc hại và tồn tại cả ngàn năm. Khả năng rủi ro rò rỉ phóng xạ cao là một hiểm họa cho sự sống muôn loài. Chưa nói đến vũ khí hạt nhân thật sự là một đại thảm họa cho nhân loại.

Tính cho tới nay, chỉ có hai quả bom hạt nhân đã được sử dụng trong Thế chiến II tại Hiroshima và Nagasaki (Nhật Bản), đã cho thế giới thấy sức công phá và hủy diệt khủng khiếp của chúng. Ngày nay cả thế giới có trên 15.000 loại vũ khí hạt nhân và hãy tưởng tượng nếu có một lãnh tụ điên khùng nào ra lệnh nhấn nút khai hỏa thì loài người sẽ tuyệt chủng, những người còn lại mức sống và tuổi thọ chỉ tương đương với thời kỳ trước Trung cổ. Ngoài ra, nó cũng sẽ hủy diệt hệ sinh thái và tác động khủng khiếp đến khí hậu Trái Đất.Vũ khí hạt nhân tiêu diệt thế giới vật chất. Vũ khí Deepfake tiêu diệt giá trị tinh thần.

Sự giả mạo và dối trá

Con người ngày nay sống trong một thế giới đầy sự giả mạo, dối trá, không ai còn tin những giá trị thật của lời nói, con người thật, sản phẩm hàng hóa thật. Các bộ phận trong con người như tóc, tai, mắt, mũi, chân mày, lông mi, đến cả ngực, mông đều có thể làm giả. Các sản phẩm mỹ thuật như cây, hoa, thú vật kiểng cũng được làm giả bằng các loại nhựa.

Các mặt hàng giả, phẩm chất kém, đa dạng, đầy dẫy, gây thiệt mạng cho người tiêu thụ như phụ tùng an toàn xe hơi, mỹ phẩm... Thực phẩm giả ẩn chứa nhiều rủi ro cho sức khỏe như trứng giả chứa nhiều thành phần gây hại; mì giả làm bằng ngũ cốc hư thối; nước mắm giả được chế từ nước lã, muối, chất hóa học tạo màu, hương vị; thuốc chữa bệnh giả chứa độc chất nguy hiểm, kim loại nặng, hoặc gây chết người. Từ những thứ giả mạo đó, xã hội sinh sản ra những hạng người giả: bác sĩ giả (học

dỏm, trường dỏm), giáo sư giả (bằng cấp dỏm), thể tháo giả (dùng thuốc cường lực), tu sĩ giả (khẩu phật tâm xà) v.v…

Tương tác giữa người với người

Mối quan hệ giữa người với người luôn tồn tại hai hiện tượng song hành đó là thật và giả. Thật, giả có lúc rất rõ ràng, dễ nhận biết, nhưng cũng có lúc lẫn lộn, phức tạp, khó nhận biết. Sự thật khách quan được hiểu cái gì đó là đúng hoặc có thể được chứng minh với bằng chứng cụ thể.

Sự giả dối có thể nhận thức được từ ý chí chủ quan của một người. Người tạo ra sự giả dối đều có mục đích riêng của nó. Bà Vian Bakir, giáo sư truyền thông chính trị và báo chí tại Đại học Bangor Xứ Wales, viết: "Điều đặc biệt tồi tệ về thời điểm hiện tại là một số chính trị gia nổi bật... và các nhà lãnh đạo chuyên quyền trên khắp thế giới đã biến việc nói dối trơ trẽn trở thành thói quen và đương nhiên họ không quan tâm liệu họ có bị phát hiện hay không."

Bà nói thêm: "Tôn giáo trở thành tồi tệ khi người ta dùng nó làm chính trị – để bảo vệ quyền lợi của giáo hội, để tấn công "kẻ xấu", để ủng hộ chiến tranh và xâm lăng, để tận diệt các văn hóa và tôn giáo khác, để bảo vệ giáo pháp, để thay trời hành đạo… nói chung là mọi việc có tính cách chính trị – dù các từ ngữ dùng nghe cao siêu đến thế nào."

Quan hệ xã hội, kể cả các quan hệ thân thuộc như cha mẹ, vợ chồng, con cái, anh em, bà con , thầy trò, chủ tớ... phần lớn đều được giao tiếp qua sự giả dối. Phải nhận rằng, con người hiện nay sống theo quan niệm người khác, suy nghĩ theo người khác, nói theo người khác, hành động theo người khác, rất hiếm cái nào là thực tế khách quan, chân thật.

Hình 15:

Deepfake làm mất niềm tin, làm người ta không phân biệt phải trái, chân giả, cuốn vào guồng quay điên cuồng của nó. Phải nhận biết được sự thật, chấp nhận và đối diện với sự thật, lên án, bài trừ sự giả dối thì xã hội loài người mới phát triển theo quy luật tiến hoá của nhân loại. Nuôi dưỡng sự giả dối, tư tưởng dễ mất phương hướng, thiếu niềm tin chân lý, chìm ngập trong mơ hồ, ảo tưởng, làm nghiêm trọng thêm căn bệnh chủ quan, đưa đến hoang tưởng, cực đoan.

Chừng nào con người trở về với chân tâm của mình thì chừng đó tâm mới an, xã hội mới bình.

LÊ TẤN TÀI