

Quyền tự do ngôn luận và trách nhiệm trong chuyển đổi số - Nhận diện ranh giới từ thực tiễn

16/06/2026 09:10

(LSVN) - Chuyển đổi số ngày càng phát triển, tuy nhiên, cùng với sự mở rộng của quyền tự do ngôn luận là những nguy cơ ngày càng gia tăng từ thông tin sai lệch, ngôn từ kích động thù hằn và việc lợi dụng không gian mạng để xuyên tạc, trục lợi hoặc gây ảnh hưởng tiêu cực đến xã hội. Thực tiễn trên thế giới và tại Việt Nam cho thấy quyền tự do ngôn luận trong môi trường số không thể tách rời trách nhiệm pháp lý và trách nhiệm xã hội. Do đó, việc nhận diện rõ ranh giới giữa tự do biểu đạt và hành vi lạm dụng quyền tự do ngôn luận trên thế giới trở thành yêu cầu cần thiết nhằm xây dựng một môi trường số an toàn, minh bạch và phát triển bền vững, đây cũng là bài học cho Việt Nam.

1. Tổng quan tình hình trên thế giới

Chuyển đổi số đã và đang làm thay đổi sâu sắc cách con người tiếp cận, trao đổi và lan truyền thông tin trên toàn cầu. Sự phát triển mạnh mẽ của internet, mạng xã hội và các nền tảng truyền thông trực tuyến đã mở rộng đáng kể quyền tự do ngôn luận, cho phép mỗi cá nhân dễ dàng bày tỏ quan điểm, chia sẻ thông tin và tham gia phản biện xã hội theo cách trực tiếp, nhanh chóng và rộng khắp hơn trước đây.

Theo Reuters Institute cho biết hơn 58% người được khảo sát trên toàn cầu bày tỏ lo ngại về khả năng phân biệt đúng – sai đối với thông tin trên internet; tại Mỹ và nhiều quốc gia châu Phi, tỷ lệ này lên tới 73%. UNESCO cũng công bố khảo sát quốc tế cho thấy với hơn 8.000 người tham gia tại 16 quốc gia sẽ tổ chức bầu cử vào năm 2024. Kết quả cho thấy 85% người dân lo ngại về tác động của thông tin sai lệch trực tuyến, trong bối cảnh các nền tảng mạng xã hội đã trở thành nguồn thông tin chính đối với phần lớn người dân. Cùng một cuộc khảo sát cho thấy 87% người dân tin rằng thông tin sai lệch này đã ảnh hưởng lớn đến đời sống chính trị của đất nước họ và lo ngại rằng nó sẽ ảnh hưởng đến kết quả bầu cử năm tới. Do đó, 88% kêu gọi chính phủ và các cơ quan quản lý nhanh chóng giải quyết vấn đề này bằng cách kiểm soát mạng xã hội.

Sự phát triển nhanh của trí tuệ nhân tạo (AI) và công nghệ deepfake càng làm cho môi trường thông tin trở nên phức tạp hơn. Các nội dung giả mạo bằng AI ngày càng khó nhận biết và có khả năng lan truyền với tốc độ rất nhanh. Nhiều nghiên cứu quốc tế cảnh báo AI-generated misinformation đang trở thành nguy cơ lớn đối với niềm tin xã hội, sức khỏe cộng đồng và an ninh thông tin toàn cầu.

Liên Hợp Quốc đang lên án các nền tảng số đã bị sử dụng để phát tán ngôn từ thù hận trong các cuộc xung đột sắc tộc tại Myanmar và Ethiopia. Nhiều tổ chức quốc tế nhận định misinformation và disinformation hiện là những nguy cơ lớn nhất

đối với ổn định xã hội và dân chủ toàn cầu. Báo cáo Global Risks Report 2025 của Diễn đàn Kinh tế Thế giới (WEF) cũng xếp misinformation và disinformation vào nhóm rủi ro nghiêm trọng hàng đầu trong ngắn hạn.

Trước thực trạng đó, nhiều quốc gia và tổ chức quốc tế đã tăng cường xây dựng cơ chế kiểm soát thông tin trên không gian mạng nhằm cân bằng giữa quyền tự do ngôn luận và trách nhiệm xã hội. Liên minh châu Âu đã ban hành Digital Services Act nhằm yêu cầu các nền tảng số tăng cường minh bạch, xử lý nội dung độc hại và hạn chế lan truyền thông tin sai lệch.

2. Thực trạng lạm dụng quyền tự do ngôn luận trong môi trường số

Thực tiễn cho thấy quyền tự do ngôn luận trên không gian mạng đang bị một số cá nhân và tổ chức lợi dụng để thực hiện các hành vi vi phạm pháp luật đã và đang gây ảnh hưởng tiêu cực đến xã hội.

2.1. Xuyên tạc sự thật, bóp méo thông tin

Hiện nay hành vi xuyên tạc sự thật, bóp méo thông tin hoặc phát tán thông tin sai lệch nhằm dẫn dắt dư luận theo hướng tiêu cực có xu hướng nguy hiểm và nghiêm trọng nhất. Nó được thực hiện thông qua việc cắt ghép hình ảnh, video, thay đổi bối cảnh sự việc hoặc suy diễn quá mức để làm sai lệch bản chất thông tin ban đầu.

Theo Reuters Institute Digital News Report 2025, gần 60% người dùng internet toàn cầu cho biết họ lo ngại về khả năng phân biệt thông tin thật – giả trên môi trường trực tuyến. Báo cáo khảo sát gần 100.000 người tại 48 quốc gia cũng cho thấy mạng xã hội đang trở thành nguồn tiếp cận thông tin chủ yếu của nhiều người dùng, trong khi mức độ tin tưởng vào nội dung trực tuyến có xu hướng giảm.

UNESCO cũng đưa ra cảnh báo rằng sự phát triển nhanh của trí tuệ nhân tạo (AI) và công nghệ deepfake đang khiến misinformation và disinformation trở nên tinh vi hơn bao giờ hết.

Một ví dụ điển hình là vụ việc các video deepfake giả mạo bác sĩ thật để quảng bá thông tin y tế sai lệch trên TikTok và nhiều nền tảng mạng xã hội khác. Theo đó, The Guardian đăng tải năm 2025, hàng trăm video AI đã sử dụng hình ảnh và giọng nói giả mạo của các chuyên gia y tế nhằm quảng bá các loại thực phẩm chức năng chưa được kiểm chứng. Các video này lan truyền rộng rãi trong nhiều tuần trước khi bị xử lý, gây nguy cơ ảnh hưởng trực tiếp đến sức khỏe cộng đồng và niềm tin của người dân đối với thông tin y tế chính thống.

Ngoài lĩnh vực y tế, deepfake còn bị sử dụng để thao túng dư luận trong các vấn đề chính trị và xã hội. Theo các nghiên cứu quốc tế về deepfake, nội dung giả mạo bằng AI ngày càng khó phân biệt với nội dung thật, đặc biệt trong các video liên quan đến nhân vật chính trị hoặc các sự kiện nhạy cảm. Một số nghiên cứu cho

thấy ngay cả các công cụ phát hiện deepfake hiện nay cũng gặp nhiều khó khăn khi xử lý các video giả mạo được tạo bằng công nghệ AI thế hệ mới.

Bên cạnh đó, Meta Oversight Board năm 2026 cũng chỉ ra rằng cơ chế kiểm soát deepfake hiện nay của các nền tảng mạng xã hội chưa đủ hiệu quả để ngăn chặn tốc độ lan truyền của misinformation trong các tình huống khủng hoảng hoặc xung đột.

Những thực trạng trên khiến quyền tự do ngôn luận trong môi trường số không phải là quyền tuyệt đối. Việc phát tán thông tin sai lệch, xuyên tạc hoặc thao túng dư luận không chỉ xâm phạm quyền tiếp cận thông tin đúng sự thật của công chúng mà còn ảnh hưởng trực tiếp đến an ninh xã hội, quyền lợi cá nhân và niềm tin cộng đồng. Vì vậy, tự do ngôn luận cần luôn đi kèm với trách nhiệm pháp lý, trách nhiệm đạo đức và ý thức kiểm chứng thông tin trước khi chia sẻ trên không gian mạng.

2.2. Kích động thù hận và phân cực xã hội

Bên cạnh tình trạng xuyên tạc thông tin, môi trường số còn làm gia tăng các hành vi kích động thù hận, cực đoan hóa quan điểm và phân cực xã hội. Đặc điểm lan truyền nhanh, khả năng ẩn danh cao cùng cơ chế đề xuất nội dung của các nền tảng mạng xã hội khiến các phát ngôn gây tranh cãi, kích động hoặc mang tính thù địch dễ thu hút tương tác và lan rộng hơn các nội dung trung lập.

Theo Báo cáo Global Risks Report 2025 của Diễn đàn Kinh tế Thế giới (WEF), misinformation và disinformation tiếp tục được xếp vào nhóm rủi ro nghiêm trọng hàng đầu đối với thế giới trong ngắn hạn năm thứ hai liên tiếp. Báo cáo nhấn mạnh rằng thông tin sai lệch đang làm suy giảm niềm tin xã hội, gia tăng chia rẽ và thúc đẩy phân cực giữa các nhóm cộng đồng và quốc gia.

Một trong những ví dụ điển hình nhất về tác động nguy hiểm của ngôn từ thù hận trên mạng xã hội là cuộc khủng hoảng người Rohingya tại Myanmar. Năm 2018, phái đoàn điều tra của Liên Hợp Quốc kết luận rằng Facebook đã đóng “vai trò quyết định” trong việc lan truyền ngôn từ thù hận và kích động bạo lực chống lại cộng đồng người Rohingya. Theo báo cáo của Liên Hợp Quốc, hơn 670.000 người Rohingya đã phải chạy sang Bangladesh sau các cuộc đàn áp bạo lực tại bang Rakhine.

Các cuộc điều tra sau đó cho thấy nhiều tài khoản và trang liên quan đến quân đội Myanmar đã sử dụng Facebook để phát tán thông tin sai lệch, nội dung kích động thù hận và tuyên truyền chống người Rohingya. Theo điều tra nội bộ của Facebook được TIME công bố năm 2018, mạng lưới này bao gồm hàng chục tài khoản và 52 trang có khả năng tiếp cận tới khoảng 12 triệu người dùng.

Ngoài ra, báo cáo phân tích của Independent Investigative Mechanism for Myanmar (IIMM) xác định hơn 10.000 bài đăng chứa nội dung thù hận chống

người Rohingya đã được phát tán trên Facebook trong giai đoạn bạo lực diễn ra mạnh nhất. Một số bài đăng thu hút hàng trăm bình luận kêu gọi bạo lực hoặc tiêu diệt cộng đồng này. Ví dụ như, người Rohingya là một dân tộc thiểu số chủ yếu theo đạo Hồi, sinh sống tại bang Rakhine phía bắc Myanmar. Vào tháng 8 năm 2017, hơn 700.000 người Rohingya đã bỏ chạy khỏi Rakhine khi lực lượng an ninh Myanmar phát động một chiến dịch có chủ đích, với các hành động giết người, hãm hiếp và đốt nhà trên diện rộng và có hệ thống. Bạo lực này diễn ra sau nhiều thập kỷ phân biệt đối xử, đàn áp và bức hại người Rohingya do nhà nước bảo trợ, tương đương với chế độ phân biệt chủng tộc.

Ngoài ra, Liên Hợp Quốc và nhiều tổ chức quốc tế cũng cảnh báo rằng ngôn từ thù hận trên mạng xã hội đang góp phần thúc đẩy cực đoan hóa và bạo lực tại nhiều khu vực khác như Ethiopia, Trung Đông và châu Âu, hiện việc kiểm soát hiệu quả các nội dung kích động, đặc biệt tại các quốc gia sử dụng ngôn ngữ ít phổ biến hoặc đang thiếu đội ngũ kiểm duyệt địa phương.

Bên cạnh tác động chính trị và sắc tộc, sự phân cực xã hội trên môi trường số còn thể hiện ở xu hướng “buồng vang thông tin” (echo chamber), nơi người dùng chủ yếu tiếp xúc với các quan điểm giống mình do thuật toán đề xuất nội dung cá nhân hóa khiến tình trạng ngày càng trở nên phức tạp.

2.3. Lợi dụng thông tin sai lệch để trục lợi cá nhân

Trong thời đại số, misinformation không chỉ gây sai lệch nhận thức xã hội mà còn bị lợi dụng như một công cụ để trục lợi tài chính. Nhiều đối tượng đã sử dụng mạng xã hội, livestream, video ngắn hoặc công nghệ AI nhằm tạo dựng niềm tin giả, thao túng cảm xúc người xem để lừa đảo, kêu gọi quyên góp hoặc bán các sản phẩm không được kiểm chứng.

Lượng nội dung deepfake được chia sẻ trực tuyến đang bùng nổ. Sau khi ước tính có khoảng 500.000 deepfake được chia sẻ trên các nền tảng mạng xã hội vào năm 2023, con số này dự kiến sẽ tăng vọt lên 8 triệu vào năm 2025. Điều này phù hợp với tốc độ tăng trưởng mà trong đó số lượng video deepfake tăng 900% mỗi năm. Đây không phải là sự tăng trưởng tuyến tính; đó là sự lan truyền mạnh mẽ vượt xa hầu hết các mối đe dọa mạng khác. Sự gia tăng này được phản ánh trong các chỉ số phát hiện trên tất cả các ngành. Chỉ riêng trong năm 2023, số lượng các vụ việc deepfake được phát hiện đã tăng gấp 10 lần so với năm trước, cho thấy sự ứng dụng nhanh chóng và rộng rãi công nghệ này bởi các đối tượng xấu.

Một nghiên cứu tại Anh về nhận thức đối với deepfake cho thấy khoảng 15% người được khảo sát từng tiếp xúc với các nội dung deepfake có hại như lừa đảo tài chính, misinformation y tế hoặc deepfake mang tính thao túng dư luận; đồng thời hơn 90% người tham gia bày tỏ lo ngại về sự lan rộng của công nghệ này.

Thực tế cho thấy các đối tượng lừa đảo ngày càng tận dụng AI để giả mạo hình ảnh, giọng nói hoặc video của chuyên gia, bác sĩ, doanh nhân và người nổi tiếng

nhằm tăng độ tin cậy cho nội dung sai lệch. Một cuộc điều tra của tổ chức kiểm chứng Full Fact được The Guardian đăng tải năm 2025 phát hiện hàng trăm video deepfake trên TikTok và nhiều nền tảng mạng xã hội khác đã sử dụng hình ảnh và giọng nói giả mạo của các bác sĩ thật để quảng bá thực phẩm chức năng chưa được kiểm chứng. Các video này hướng người xem mua sản phẩm sức khỏe bằng các thông tin sai lệch về y tế và tồn tại trên nền tảng trong nhiều tuần trước khi bị gỡ bỏ.

UNESCO cũng cảnh báo rằng deepfake đang tạo ra một “khủng hoảng niềm tin” trong lĩnh vực y tế khi các video giả mạo bác sĩ được sử dụng để quảng bá thuốc và liệu pháp không có cơ sở khoa học. Theo UNESCO, điều này đã và đang gây thiệt hại tài chính và cũng đe dọa trực tiếp đến sức khỏe cộng đồng khi người dân bị dẫn dắt bởi các thông tin y khoa sai lệch.

Trong lĩnh vực tài chính, các vụ lừa đảo đầu tư bằng deepfake cũng gia tăng mạnh. Financial News London cho biết các đối tượng lừa đảo đã sử dụng video AI giả mạo các chuyên gia tài chính nổi tiếng để quảng bá các chương trình đầu tư giả trên Facebook và Instagram. Báo cáo của Signicat được trích dẫn trong bài viết cho thấy số vụ gian lận tài chính liên quan đến AI đã tăng tới 2.137% trong vòng ba năm gần đây.

Khoảng 48% người Mỹ cho rằng các cuộc gọi rác đang trở nên tồi tệ hơn, gấp khoảng ba lần so với số người cho rằng tình trạng này đã được cải thiện. Xu hướng này xuất hiện ở nhiều quốc gia tham gia khảo sát. Người tiêu dùng tại sáu thị trường cho biết trung bình mỗi tuần họ nhận được 7,4 cuộc gọi không mong muốn. Người trả lời khảo sát ở Pháp báo cáo số lượng cuộc gọi không mong muốn cao nhất, trong khi người trả lời khảo sát ở Anh báo cáo thiệt hại tài chính cao nhất trong số các nạn nhân bị lừa đảo. Khoảng 38% người dùng cho biết họ có khả năng chuyển sang nhà cung cấp khác nếu cảm thấy nhà mạng hiện tại không thể bảo vệ họ. Nhiều người dùng (72%) hiện đang kêu gọi chính phủ ban hành các quy định chặt chẽ hơn, yêu cầu các công ty viễn thông phải ứng phó với các vụ lừa đảo do trí tuệ nhân tạo gây ra. Một số người được hỏi cũng tin rằng các nhà mạng nên cùng chịu trách nhiệm về những thiệt hại do lừa đảo phát sinh trên mạng lưới của họ. Khoảng 67% người dùng ủng hộ một mức độ trách nhiệm tài chính nhất định, trong khi 55% ủng hộ các biện pháp bảo vệ không chịu trách nhiệm tương tự như những biện pháp được các công ty thẻ tín dụng sử dụng..

Có thể nhận định rằng AI đang làm gia tăng đáng kể quy mô và mức độ tinh vi của các hành vi trục lợi trên không gian mạng. Theo Cybernews, trong khoảng 346 sự cố AI được ghi nhận năm 2025, phát hiện tới 179 vụ liên quan đến deepfake và 81% các vụ lừa đảo AI sử dụng công nghệ deepfake để giả mạo hình ảnh, giọng nói hoặc video.

Những thực trạng trên cho thấy misinformation trong môi trường số không còn đơn thuần là vấn đề sai lệch thông tin mà đã trở thành công cụ phục vụ hành vi

lừa đảo, thao túng và trục lợi cá nhân. Điều này đặt ra yêu cầu cấp thiết phải tăng cường kiểm soát nội dung số, nâng cao kỹ năng nhận diện thông tin giả và hoàn thiện cơ chế pháp lý nhằm bảo vệ người dùng trên không gian mạng.

3. Giải pháp và trách nhiệm trong chuyển đổi số

Trong bối cảnh chuyển đổi số, quyền tự do ngôn luận cần được thực hiện song song với trách nhiệm pháp lý và trách nhiệm xã hội. Để xây dựng môi trường số an toàn, minh bạch và phát triển bền vững, cần có sự phối hợp của nhiều chủ thể.

3.1. Nâng cao trách nhiệm của người dùng

Trong môi trường số, mỗi người dùng không chỉ tiếp nhận mà còn trực tiếp lan truyền thông tin, vì vậy cần nâng cao trách nhiệm khi sử dụng mạng xã hội. Theo UNESCO, kỹ năng truyền thông và thông tin (Media and Information Literacy – MIL) giúp người dùng nhận diện, kiểm chứng và sử dụng thông tin một cách có trách nhiệm nhằm hạn chế misinformation và tin giả.

Hiện nay việc chia sẻ thông tin sai lệch do thiếu kỹ năng kiểm chứng đã và đang diễn ra nhiều hơn. Khảo sát của UNESCO năm 2024 cho thấy khoảng 60% nhà sáng tạo nội dung không kiểm tra độ chính xác thông tin trước khi đăng tải. Vì vậy, người dùng cần: Kiểm tra nguồn tin trước khi chia sẻ; Đối chiếu với báo chí chính thống hoặc cơ quan chức năng; Không lan truyền nội dung giật gân, chưa xác thực.

Ví dụ, trong đại dịch Covid-19, nhiều thông tin sai lệch về vaccine lan truyền trên mạng xã hội đã khiến một bộ phận người dân hoang mang và từ chối tiêm chủng. WHO gọi đây là “infodemic” – tức “đại dịch thông tin”.

Ngoài ra, người dùng cũng cần nâng cao kỹ năng nhận diện deepfake và nội dung do AI tạo ra. Năm 2025, nhiều video deepfake giả mạo đóng thành bác sĩ đã xuất hiện trên TikTok để quảng bá thực phẩm chức năng không được kiểm chứng, gây ảnh hưởng đến sức khỏe cộng đồng.

Bên cạnh đó, cũng nên xây dựng văn hóa tranh luận văn minh trên mạng xã hội cũng rất cần thiết. Tự do ngôn luận không đồng nghĩa với quyền xúc phạm, công kích hoặc kích động thù hằn đối với người khác. Vì vậy, mỗi cá nhân cần sử dụng mạng xã hội có trách nhiệm nhằm góp phần xây dựng môi trường số an toàn và lành mạnh.

3.2. Tăng cường trách nhiệm của các nền tảng số

Các nền tảng mạng xã hội như Facebook, TikTok, YouTube hay X (Twitter) giữ vai trò trung tâm trong việc lan truyền thông tin trên không gian mạng, vì vậy cần có trách nhiệm kiểm soát nội dung sai lệch, kích động thù hằn và vi phạm pháp luật. UNESCO cho rằng các nền tảng số hiện nay đang góp phần làm gia tăng

misinformation và hate speech do thuật toán ưu tiên nội dung có khả năng tạo tương tác cao hơn độ tin cậy thông tin.

Theo khảo sát toàn cầu của UNESCO/Ipsos năm 2023, hơn 85% người được hỏi lo ngại về tác động của misinformation trực tuyến và 87% cho rằng thông tin sai lệch đã gây ảnh hưởng tiêu cực đến chính trị và xã hội tại quốc gia của họ. Ngoài ra, UNESCO cũng cho biết khoảng 2/3 nhà sáng tạo nội dung số không thực hiện kiểm chứng thông tin trước khi đăng tải.

Tháng 8 năm 2023, Luật Dịch vụ Kỹ thuật số (DSA) chính thức có hiệu lực pháp lý, điều chỉnh hoạt động của các trung gian và nền tảng trực tuyến. Mục tiêu chính của luật này là ngăn chặn các hoạt động bất hợp pháp và gây hại trên mạng cũng như sự lan truyền thông tin sai lệch. Luật đảm bảo an toàn cho người dùng, bảo vệ các quyền cơ bản và tạo ra một môi trường nền tảng trực tuyến công bằng và cởi mở. Các nền tảng trực tuyến lớn và các công cụ tìm kiếm trực tuyến lớn phải chia sẻ báo cáo đánh giá rủi ro hàng năm về nội dung bất hợp pháp được phát tán thông qua dịch vụ của họ. Họ cũng phải điều chỉnh các biện pháp giảm thiểu rủi ro..

Một ví dụ điển hình là trong xung đột Gaza năm 2023, NewsGuard ghi nhận ít nhất 14 thông tin sai lệch liên quan đến chiến sự đã thu hút hơn 22 triệu lượt xem chỉ trong vài ngày trên X, TikTok và Instagram. Sau đó, Liên minh châu Âu đã mở điều tra đối với X về việc lan truyền disinformation và nội dung cực đoan.

Những thực tế này cho thấy các nền tảng số không thể chỉ đóng vai trò “trung gian truyền tải thông tin” mà cần chủ động kiểm soát nội dung, tăng cường minh bạch thuật toán và phối hợp với cơ quan chức năng nhằm xây dựng môi trường số an toàn, lành mạnh và hạn chế lạm dụng quyền tự do ngôn luận trên không gian mạng.

3.3. Hoàn thiện hệ thống pháp luật và quản lý nhà nước

Bên cạnh trách nhiệm của người dùng và các nền tảng số, Nhà nước giữ vai trò quan trọng trong việc xây dựng khung pháp lý nhằm bảo đảm cân bằng giữa quyền tự do ngôn luận và bảo vệ trật tự, an toàn xã hội trên không gian mạng.

Trước sự gia tăng của misinformation và deepfake, nhiều quốc gia đã ban hành các quy định kiểm soát nội dung số và yêu cầu nền tảng mạng xã hội chịu trách nhiệm đối với thông tin lan truyền trên hệ thống của mình. Liên minh châu Âu đã triển khai Digital Services Act (DSA), yêu cầu các nền tảng như Facebook, TikTok và YouTube phải minh bạch thuật toán, xử lý nội dung độc hại và hạn chế misinformation. Đức cũng áp dụng luật Network Enforcement Act (NetzDG), cho phép phạt tới 50 triệu euro đối với các nền tảng không gỡ bỏ nội dung vi phạm trong thời gian quy định.

Tại Việt Nam, Nhà nước đã ban hành nhiều quy định nhằm xử lý hành vi cung cấp hoặc lan truyền thông tin sai sự thật trên môi trường mạng, trong đó có Nghị định 15/2020/NĐ-CP về xử phạt vi phạm hành chính trong lĩnh vực bưu chính, viễn thông và an ninh mạng. Đồng thời, các cơ quan chức năng cũng tăng cường phối hợp với nền tảng số để xử lý tài khoản phát tán thông tin xuyên tạc, lừa đảo hoặc kích động thù hận.

Ngoài xử lý vi phạm, nhiều tổ chức quốc tế nhấn mạnh cần đẩy mạnh giáo dục kỹ năng số và truyền thông tin cho người dân. UNESCO cho rằng “Media and Information Literacy” (MIL) là kỹ năng thiết yếu trong thời đại số nhằm giúp người dùng nhận diện, đánh giá và kiểm chứng thông tin một cách có trách nhiệm. Theo UNESCO, misinformation đang gia tăng nhanh do sự phát triển của AI và mạng xã hội, vì vậy các quốc gia cần đầu tư mạnh hơn vào giáo dục kỹ năng số cho công dân.

Bên cạnh đó, UNICEF cũng khuyến nghị các quốc gia cần tăng cường chương trình giáo dục truyền thông số cho thanh thiếu niên nhằm nâng cao tư duy phản biện và hạn chế tác động của misinformation trên internet.

Như vậy, để bảo vệ quyền tự do ngôn luận trong môi trường số một cách bền vững, Nhà nước không chỉ cần hoàn thiện pháp luật và tăng cường quản lý không gian mạng mà còn phải thúc đẩy giáo dục kỹ năng số, nâng cao nhận thức cộng đồng và phối hợp với các nền tảng công nghệ trong việc xây dựng môi trường thông tin an toàn, minh bạch và lành mạnh.

4. Tổng kết

Quyền tự do ngôn luận là một trong những quyền cơ bản của con người và đóng vai trò quan trọng trong xã hội hiện đại cũng như tiến trình chuyển đổi số. Tuy nhiên, trong môi trường số, quyền này không phải là quyền tuyệt đối mà cần được thực hiện trong khuôn khổ pháp luật và gắn liền với trách nhiệm xã hội.

Thực tiễn cho thấy nếu thiếu cơ chế kiểm soát phù hợp, thông tin sai lệch, ngôn từ thù hận và các hành vi lợi dụng tự do ngôn luận có thể gây ảnh hưởng nghiêm trọng đến an ninh xã hội, quyền con người và niềm tin cộng đồng. Vì vậy việc xây dựng nhận thức và có ranh giới rõ ràng sẽ tạo ra môi trường tốt đẹp hơn.

Tài liệu tham khảo

1. Reuters Institute for the Study of Journalism (2025), Digital News Report 2025, University of Oxford. Truy cập tại: <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary>
2. UNESCO (2023), Online Disinformation: UNESCO unveils action plan to regulate social media platforms. Truy cập

- tại:<https://www.unesco.org/en/articles/online-disinformation-unesco-unveils-action-plan-regulate-social-media-platforms>
3. UNESCO (2024), Deepfakes and Crisis: Knowing What's Real in the Age of AI. Truy cập tại: <https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing>
 4. UNESCO (2020), Media and Information Literacy Framework and Recent Initiatives. Truy cập tại:<https://www.unesco.gov.ph/wp-content/uploads/2020/03/UNESCO-Media-and-Information-Literacy-Framework-and-Recent-Initiatives.pdf>
 5. World Economic Forum (2025), Global Risks Report 2025. Truy cập tại:<https://www.weforum.org/press/2025/01/global-risks-report-2025-conflict-environment-and-disinformation-top-threats/>
 6. European Commission, Strengthening online platforms' responsibility under the Digital Services Act. Truy cập tại:https://commission.europa.eu/topics/countering-information-manipulation/strengthening-online-platforms-responsibility_en
 7. TIME (2018), UN Says Facebook Played a 'Determining Role' in Myanmar Genocide. Truy cập tại: <https://time.com/5197039/un-facebook-myanmar-rohingya-violence/>
 8. TIME (2018), Inside Facebook's Myanmar Operation: The Social Network's Hate Speech Problem. Truy cập tại: <https://time.com/5383780/myanmar-facebook-propaganda-rohingya/>
 9. Amnesty International (2022), Facebook's systems promoted violence against Rohingya. Truy cập tại:<https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/>
 10. The Guardian (2025), AI deepfakes of real doctors spreading health misinformation on social media. Truy cập tại:<https://www.theguardian.com/society/2025/dec/05/ai-deepfakes-of-real-doctors-spreading-health-misinformation-on-social-media>
 11. The Guardian (2024), Online influencers need urgent fact-checking training, warns UNESCO. Truy cập tại:<https://www.theguardian.com/media/2024/nov/26/online-influencers-need-urgent-fact-checking-training-warns-unesco>
 12. UNICEF (2021), Digital Misinformation/Disinformation and Children. Truy cập tại:<https://www.unicef.org/innocenti/media/856/file/UNICEF-Global-Insight-Digital-Mis-Disinformation-and-Children-2021.pdf>
 13. DeepStrike (2025), Deepfake Statistics 2025. Truy cập tại: <https://deepstrike.io/blog/deepfake-statistics-2025>
 14. TechRadar (2026), Deepfake worries hit a new high. Truy cập tại:<https://www.techradar.com/pro/security/deepfake-worries-hit-a-new-high-as-one-in-four-americans-say-they-have-received-a-deepfake-voice-call-in-the-past-12-months-experts-blame-the-weaponization-of-ai>

- 15.Cybernews (2025), 346 AI incidents in 2025: from deepfakes and fraud to dangerous advice. Truy cập tại:<https://cybernews.com/ai-news/346-ai-incidents-in-2025-from-deepfakes-and-fraud-to-dangerous-advice/>
- 16.Chính phủ Việt Nam (2020), Nghị định số 15/2020/NĐ-CP về xử phạt vi phạm hành chính trong lĩnh vực bưu chính, viễn thông, công nghệ thông tin và giao dịch điện tử.
- 17.Bộ Khoa học và Công nghệ Việt Nam, Cẩm nang phòng chống tin giả trên không gian mạng.
- 18.World Health Organization (WHO), Infodemic Management. Truy cập tại: https://www.who.int/health-topics/infodemic#tab=tab_1

TRẦN NGỌC SƠN