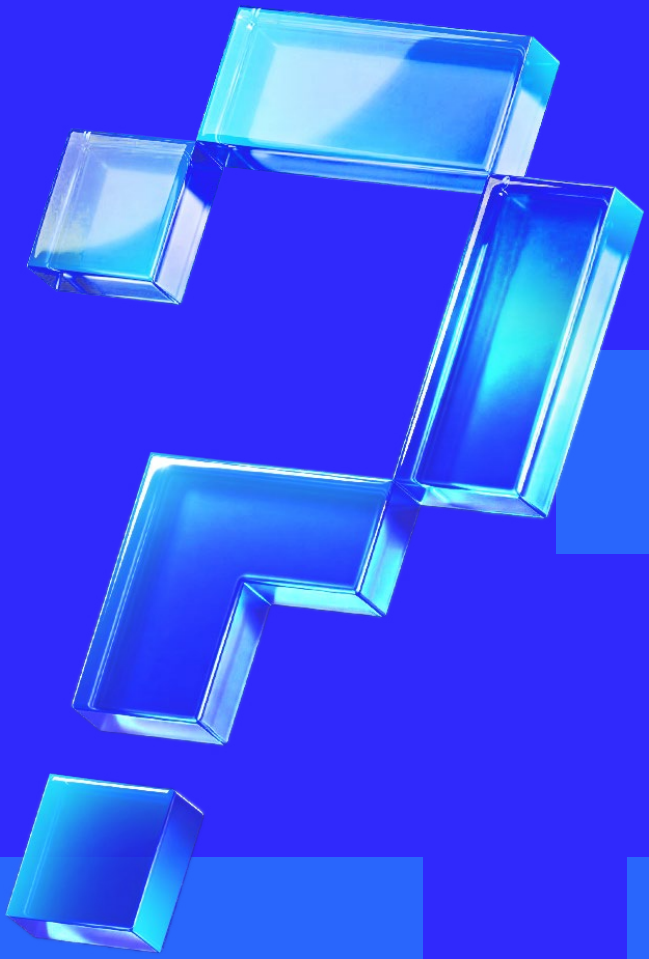
CỤC AN NINH MẠNG
VÀ PHÒNG, CHỐNG TỘI PHẠM
SỬ DỤNG CÔNG NGHỆ CAO

DTA

CẨM NANG

Kiểm chứng tin giả

tinAI



#TINAI
#KIEMCHUNGTINGIA

#KIEMTRUOCTINSAU
#CHIASETRACHNHIEM

MỤC LỤC

tr. 4 - 6 — LỜI MỞ ĐẦU

tr. 7 - 16 — CHƯƠNG I
THẾ GIỚI THÔNG TIN

tr. 8 - 39 — CHƯƠNG II
BẢN ĐỒ NHẬN DIỆN

tr. 40 - 53 — CHƯƠNG III
LA BÀN KIỂM CHỨNG

tr. 54 - 63 — CHƯƠNG IV
THẾ GIỚI THÔNG TIN THỜI AI

tr. 64 - 70 — CHƯƠNG V
ỨNG XỬ VỚI THÔNG TIN TRONG ĐỜI SỐNG SỐ

tr. 71 — LỜI KẾT

LỜI MỞ ĐẦU

#TINAI
#KIEMCHUNGTINGIA

#KIEMTRUOCTINSAU
#CHIASETRACHNHIEM

Từ lúc thức dậy cho đến khi đi ngủ, mỗi người đều tiếp xúc với một lượng thông tin khổng lồ. Chúng ta đọc tin tức trên báo điện tử, xem video trên mạng xã hội, nhận tin nhắn từ bạn bè, trao đổi qua các nhóm trò chuyện, tra cứu thông tin trên Internet hoặc đặt câu hỏi cho các hệ thống trí tuệ nhân tạo (AI).

Thông tin đã trở thành một phần không thể tách rời của cuộc sống hiện đại, là nền tảng của mọi quyết định.

Những quyết định hằng ngày, từ việc lựa chọn một sản phẩm tiêu dùng, chăm sóc sức khỏe, tìm kiếm việc làm đến đầu tư tài chính hay tham gia các hoạt động xã hội đều chịu ảnh hưởng bởi thông tin mà chúng ta tiếp nhận.

Sự phát triển của mạng xã hội và đặc biệt là trí tuệ nhân tạo đã tạo ra một môi trường thông tin hoàn toàn mới

Nếu trước đây, phần lớn thông tin được sản xuất bởi các cơ quan báo chí, tổ chức hoặc cơ quan nhà nước thì ngày nay bất kỳ ai cũng có thể tạo ra và lan truyền thông tin chỉ với một chiếc điện thoại thông minh.

Điều này mở ra nhiều cơ hội để tiếp cận tri thức và kết nối cộng đồng, nhưng cũng đặt ra không ít thách thức. Thông tin sai lệch, tin giả, nội dung bị cắt ghép, hình ảnh và video do AI tạo ra hay các hình thức giả mạo ngày càng xuất hiện với tần suất lớn và mức độ tinh vi hơn.

Trong bối cảnh đó, mỗi người dân không chỉ cần biết tiếp nhận thông tin, mà còn cần biết đánh giá, kiểm chứng và sử dụng thông tin một cách có trách nhiệm.

Trong kỷ nguyên số, khả năng tiếp cận thông tin không còn là lợi thế

Lợi thế thực sự là khả năng phân biệt đâu là thông tin đáng tin cậy.

[CHƯƠNG I]

01

THẾ GIỚI THÔNG TIN

Hiểu môi trường thông tin để trở thành người tiếp nhận một cách có chọn lọc

[1] THÔNG TIN ĐANG THAY ĐỔI NHƯ THẾ NÀO?

Trong nhiều thập kỷ trước, thông tin chủ yếu được truyền tải thông qua báo chí, phát thanh, truyền hình hoặc các tài liệu chính thức.

Quy trình sản xuất thông tin thường trải qua nhiều bước biên tập, kiểm chứng và chịu trách nhiệm bởi các tổ chức xuất bản.

Ngày nay, cùng với sự phát triển của Internet và các nền tảng số, quá trình tạo lập và lan truyền thông tin đã thay đổi đáng kể. Bất kỳ cá nhân nào cũng có thể trở thành người tạo ra nội dung, phát trực tiếp một sự kiện, chia sẻ quan điểm hoặc lan truyền thông tin đến hàng nghìn, thậm chí hàng triệu người chỉ trong vài giây.

Sự thay đổi của môi trường thông tin cũng đã được phản ánh qua những số liệu toàn cầu. Theo báo cáo Digital 2026 Global Overview của DataReportal, hơn 68,7% dân số thế giới đang sử dụng mạng xã hội. Số lượng người dùng đã tăng từ dưới 500 triệu người vào năm 2005 lên 2,27 tỷ người vào năm 2015 và đạt khoảng 5,66 tỷ người vào năm 2025, chủ yếu nhờ sự phổ cập của điện thoại thông minh và Internet.

Đối với nhiều người, mạng xã hội đã trở thành nguồn tiếp cận thông tin chính

Theo báo cáo thường niên do Viện nghiên cứu báo chí Reuters thuộc Đại học Oxford công bố ngày 16/6/2026, khoảng 54% số người được khảo sát cho biết họ đã tiếp cận tin tức thông qua mạng xã hội hoặc các nền tảng video. Điều đó có nghĩa là bên cạnh vai trò kết nối và giải trí, các nền tảng số ngày càng có ảnh hưởng lớn đến nhận thức và quan điểm của cộng đồng.

Sự thay đổi này mang lại nhiều lợi ích khi tri thức được chia sẻ nhanh hơn, người dân có nhiều cơ hội tiếp cận thông tin hơn và tiếng nói của cộng đồng được lắng nghe nhiều hơn. Tuy nhiên, tốc độ lan truyền nhanh chóng cũng khiến môi trường thông tin đối mặt với nhiều thách thức.

Những thông tin chưa được kiểm chứng có thể lan rộng trước khi xác thực và được đính chính.

Mối lo này không chỉ xuất hiện ở một quốc gia. Khảo sát cuối năm 2023 của UNESCO với hơn 8.000 người tham gia tại 16 quốc gia cho thấy, 85% người dân lo ngại về tác động của thông tin sai lệch trực tuyến, đặc biệt là trong bối cảnh các nền tảng mạng xã hội đã trở thành nguồn thông tin chính đối với phần lớn người dân. Trong khi đó, báo cáo Rủi ro Toàn cầu năm 2025 của Diễn đàn Kinh tế Thế giới (WEF) tiếp tục xếp thông tin sai lệch và thông tin xuyên tạc vào nhóm những rủi ro ngắn hạn nghiêm trọng nhất đối với thế giới.

Những thông tin chưa được kiểm chứng có thể tác động đến kinh tế, xã hội, an ninh và thậm chí ảnh hưởng đến các quyết định chính sách.

Tại Việt Nam, thống kê từ Bộ Công an chỉ tính riêng trong thời điểm từ 2022-2024, mỗi năm tin giả, tin sai sự thật tán phát từ **100.000 - 180.000 lượt tin, bài viết, hình ảnh, video trên các nền tảng mạng xã hội, diễn đàn.**

Một thực tế đáng chú ý khác là không phải ai tạo nội dung đều có thói quen kiểm chứng thông tin.

Theo báo cáo do UNESCO công bố năm 2024:

62%

Nhà sáng tạo nội dung số được khảo sát cho biết họ không thực hiện việc kiểm chứng thông tin một cách nghiêm ngặt trước khi chia sẻ tới công chúng

42%

Thừa nhận sử dụng số lượt thích, lượt xem hoặc mức độ phổ biến của nội dung như một tiêu chí để đánh giá độ tin cậy của thông tin.

Những con số này cho thấy trong môi trường mạng xã hội, sức hút và khả năng lan truyền đôi khi đang được đặt cao hơn tính xác thực của thông tin. Không ít người chia sẻ một nội dung vì thấy nó đang được quan tâm hoặc được nhiều người hưởng ứng, thay vì dành thời gian kiểm chứng độ chính xác của thông tin đó.

Những con số trên cho thấy chúng ta đang bước vào một thời kỳ mà thông tin được tạo ra nhiều hơn, lan truyền nhanh hơn và tác động mạnh hơn bao giờ hết.

Điều đó cũng đồng nghĩa với việc mỗi người cần trang bị những kỹ năng mới để tiếp cận và đánh giá thông tin một cách an toàn.

[II] CÁCH THÔNG TIN ĐƯỢC LAN TRUYỀN

Để biết vì sao một thông tin có thể lan truyền mạnh mẽ, trước hết cần hiểu "vòng đời" của thông tin trong môi trường số. Thông thường, một thông tin sẽ trải qua 4 giai đoạn:

■ TẠO RA

Thông tin có thể xuất phát từ cá nhân, tổ chức, cơ quan báo chí, tài liệu chính thức hoặc từ các công cụ AI đang ngày càng phổ biến hiện nay.

■ LAN TRUYỀN

Thông tin được chia sẻ qua mạng xã hội, ứng dụng nhắn tin, báo chí, diễn đàn hoặc các nền tảng trực tuyến. Trong giai đoạn này, thuật toán của các nền tảng số có thể ưu tiên hiển thị những nội dung thu hút nhiều sự chú ý, tạo cảm xúc mạnh hoặc đang được nhiều người quan tâm.

■ TIẾP NHẬN

Mỗi người tiếp cận thông tin theo những cách khác nhau. Kiến thức, kinh nghiệm, cảm xúc và quan điểm cá nhân đều ảnh hưởng đến cách chúng ta hiểu và đánh giá cùng một nội dung.

■ LAN TRUYỀN TIẾP

Khi nhấn nút "chia sẻ", "chuyển tiếp" hoặc kể lại cho người khác, mỗi người đều trở thành một mắt xích trong quá trình lan truyền thông tin. Một hành động tưởng như nhỏ cũng có thể góp phần giúp thông tin hữu ích được lan tỏa hoặc khiến thông tin sai lệch tiếp cận thêm nhiều người.

Mỗi cá nhân không chỉ là người nhận thông tin mà còn "phát đi thông tin". Vì vậy, mỗi lần chia sẻ đều đi kèm với trách nhiệm.

[III]

TẠI SAO CHÚNG TA DỄ TIN MỘT THÔNG TIN CHƯA KIỂM CHỨNG?

Con người không tiếp nhận thông tin một cách hoàn toàn lý trí. Mỗi ngày, chúng ta phải xử lý một lượng thông tin khổng lồ. Để tiết kiệm thời gian và năng lượng, não bộ thường sử dụng các **phương pháp phán đoán nhanh (heuristics)** hay các **"lối tắt nhận thức" (mental shortcuts)**.

Cơ chế này giúp đánh giá thông tin dựa trên kinh nghiệm, cảm xúc và các tín hiệu quen thuộc thay vì phân tích mọi thứ từ đầu. Trong đời sống hàng ngày, điều này khá hữu ích khi giúp chúng ta phản ứng kịp thời trước các tình huống khẩn cấp, giảm gánh nặng xử lý thông tin và đưa ra quyết định nhanh chóng.

Tuy nhiên, chính những cơ chế đó cũng có thể khiến chúng ta đưa ra những kết luận vội vàng hoặc tin vào những thông tin chưa được kiểm chứng, hình thành thiên kiến nhận thức.

Các **"lối tắt nhận thức"** thường gặp bao gồm:

TÍNH SẴN CÓ (AVAILABILITY HEURISTIC)

Xu hướng đánh giá mức độ phổ biến, khả năng xảy ra hoặc mức độ nghiêm trọng của một sự việc dựa trên những ví dụ hoặc hình ảnh dễ dàng xuất hiện trong tâm trí nhất, thay vì dữ liệu hoặc bằng chứng khách quan.

Ví dụ:
Tin giả thường sử dụng những hình ảnh gây sốc, video cắt ghép hoặc các câu chuyện giàu cảm xúc để tạo ấn tượng mạnh và khiến người xem dễ ghi nhớ. Khi những hình ảnh này liên tục xuất hiện trên mạng xã hội, nhiều người có thể cảm thấy hiện tượng đó đang xảy ra rất phổ biến hoặc có nguy cơ rất cao, dù trên thực tế chỉ là những trường hợp cá biệt hoặc đã bị thổi phồng.

TÍNH ĐẠI DIỆN (REPRESENTATIVENESS HEURISTIC)

Đưa ra phán đoán, đánh giá một sự việc, con người dựa trên mức độ tương đồng của nó với những khuôn mẫu đã có sẵn.

Ví dụ:
Một bài viết có giao diện giống báo điện tử, sử dụng logo của cơ quan nhà nước, trình bày chuyên nghiệp hoặc có người mặc trang phục của cán bộ có thể khiến người xem nhanh chóng cho rằng thông tin đó đáng tin cậy, dù nguồn tin hoặc danh tính của người nói chưa được xác minh.

HIỆU ỨNG MỎ NEO (ANCHORING)

Sai lệch nhận thức khiến con người phụ thuộc quá mức vào thông tin đầu tiên (mỏ neo) khi đánh giá hoặc đưa ra quyết định.

Ví dụ:
Những thông tin giật gân, sai lệch xuất hiện ban đầu trở thành điểm tham chiếu (dù thông tin đó không đầy đủ) và dùng nó làm thước đo cho toàn bộ sự việc, khiến cho não bộ khó chấp nhận sự thật dù đã được chính chủ đính chính sau đó.

SỰ QUEN THUỘC (FAMILIARITY HEURISTIC)

Xu hướng ưu tiên và dễ tin tưởng những thông tin mang lại cảm giác quen thuộc. Ngoài ra, việc lặp đi lặp lại một nội dung có thể tạo ra "Hiệu ứng ảo tưởng sự thật" (Illusory Truth Effect) khiến não bộ dễ nhầm lẫn giữa sự quen thuộc và sự thật.

Ví dụ:
Thông tin sai lệch liên tục xuất hiện trên mạng xã hội, trong các nhóm chat hoặc được nhiều tài khoản chia sẻ. Sau nhiều lần nhìn thấy, người đọc có thể dần cảm thấy thông tin đó đáng tin hơn, dù chưa từng được kiểm chứng.

Nhiều nghiên cứu đã chỉ ra rằng, cảm xúc có ảnh hưởng rất lớn đến cách con người tiếp nhận thông tin. Khi một nội dung khiến chúng ta sợ hãi, tức giận, phẫn nộ, thương cảm hoặc quá phấn khích

Não bộ thường ưu tiên phản ứng nhanh thay vì dành thời gian kiểm tra độ chính xác.

Một nghiên cứu của các nhà khoa học tại MIT năm 2021 cho thấy, những người tiếp nhận thông tin dựa nhiều vào cảm xúc có xu hướng tin vào tin giả nhiều hơn và khó phân biệt giữa tin giả với tin thật so với những người dành thời gian suy nghĩ và đánh giá một cách lý trí.

Trong môi trường số, thuật toán của các nền tảng trực tuyến thường ưu tiên hiển thị nội dung phù hợp với sở thích và hành vi của người dùng, dễ tạo ra **"bong bóng thông tin" (filter bubble)** và **"buồng vang" (echo chamber)**. Khi một quan điểm liên tục được lặp lại trong cùng một môi trường thông tin, chúng ta có thể cảm thấy nó đúng hoặc phổ biến hơn thực tế. Điều này khiến các thông tin sai lệch hoặc mang tính thao túng có điều kiện lan truyền và thuyết phục hơn.

Hiểu được cách não bộ xử lý thông tin không khiến chúng ta nghi ngờ mọi thứ mà để nhận ra rằng không ai hoàn toàn miễn nhiễm trước thông tin sai lệch.

Kiểm chứng thông tin không chỉ là việc xác minh nguồn tin mà còn là khả năng nhận biết những phản ứng và xu hướng nhận thức của chính mình trước khi quyết định tin, chia sẻ hoặc hành động.

[IV] TIN GIẢ KHÔNG PHẢI LÀ TẤT CẢ

Khi nhắc đến môi trường thông tin, nhiều người thường nghĩ ngay đến "tin giả". Thực tế, **tin giả chỉ là một trong nhiều dạng thông tin có thể gây hiểu nhầm.**

Trong đời sống hằng ngày, chúng ta còn có thể gặp thông tin sai do:

1. Nhầm lẫn hoặc thiếu cập nhật
2. Hình ảnh hoặc video sử dụng sai thời gian, địa điểm
3. Thông tin bị cắt ghép khỏi bối cảnh ban đầu
4. Nội dung mang tính suy đoán nhưng được trình bày như sự thật
5. Nội dung do AI tạo ra nhưng không được kiểm chứng
6. Thông tin giả mạo nhằm mục đích lừa đảo hoặc thao túng

Chính vì vậy, thay vì chỉ cố gắng phân biệt "tin giả" và "tin thật" thì điều quan trọng hơn là hình thành thói quen đánh giá mức độ tin cậy của mọi thông tin trước khi tin hoặc chia sẻ.

Đây cũng là lý do cuốn cẩm nang này giới thiệu về "Bản đồ nhận diện thông tin", được xem là một công cụ giúp người đọc nhận diện mức độ rủi ro của thông tin thay vì chỉ gắn nhãn "đúng" hoặc "sai".

[CHƯƠNG II]

02

BẢN ĐỒ NHẬN DIỆN

[I] VÙNG AN TOÀN

Thông tin có thể tiếp nhận nhưng không vì thế mà ngừng kiểm chứng.

01. HIỂU ĐÚNG VỀ "VÙNG AN TOÀN"

Vùng an toàn - nơi thông tin có mức rủi ro thấp trong tiếp nhận và lan truyền, do có nguồn gốc xác định, dữ kiện cụ thể, bối cảnh rõ ràng, căn cứ kiểm tra lại được và không thúc ép người tiếp nhận phải hành động ngay.

Trong đời sống hằng ngày, chúng ta thường có xu hướng chia thông tin thành hai nhóm: tin được và không tin được. Tuy nhiên, trên môi trường số, cách nhìn này không còn phù hợp.

Một thông tin được xếp vào vùng an toàn không có nghĩa là thông tin đó đúng tuyệt đối hay không bao giờ thay đổi.

Điều đó chỉ có nghĩa là tại thời điểm tiếp nhận, thông tin đã đáp ứng những điều kiện cơ bản để người đọc có thể tin cậy một cách có cơ sở.

Nói cách khác, "an toàn" không phải vì thông tin nghe hợp lý, được nhiều người chia sẻ hay đến từ một tài khoản nổi tiếng. Một thông tin được xem là an toàn khi người đọc có thể tra cứu nguồn gốc của nó và tự mình kiểm tra được tính xác thực.

Đây là điểm khác biệt quan trọng giữa niềm tin và kiểm chứng. Trong môi trường số, niềm tin thường được hình thành rất nhanh bởi cảm xúc, thói quen hoặc uy tín cá nhân. Trong khi đó, kiểm chứng dựa trên bằng chứng và khả năng xác minh.

Một thông tin chỉ thực sự có giá trị khi người tiếp nhận không cần "tin người đăng" mà có thể tin vào những căn cứ chứng minh cho thông tin đó.

Đừng tin vì người nói đáng tin. Hãy tin vì thông tin đó có thể kiểm chứng được.

02. DẤU HIỆU NHẬN BIẾT VÙNG AN TOÀN

Một thông tin thuộc vùng an toàn thường hội tụ đồng thời nhiều dấu hiệu sau đây.

Có chủ thể chịu trách nhiệm rõ ràng

Đây được xem là dấu hiệu quan trọng nhất. Người đọc cần phải xác định được:

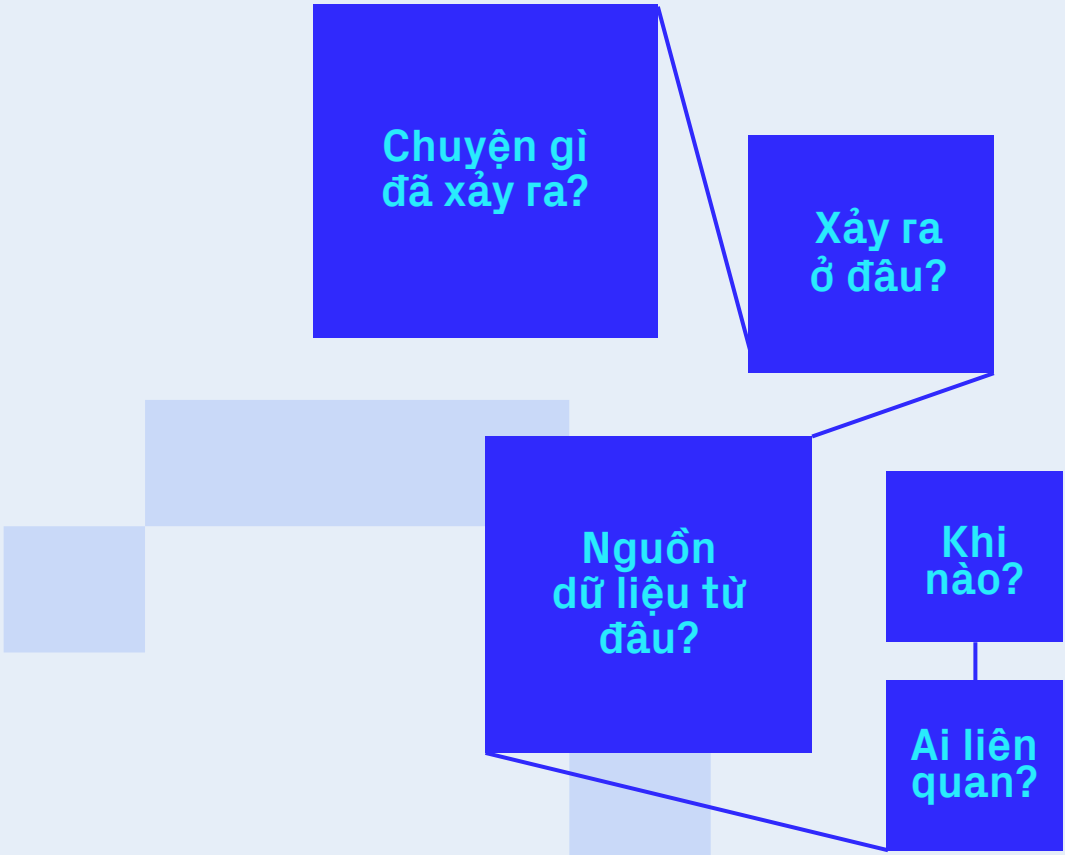
Một văn bản của cơ quan nhà nước luôn ghi rõ cơ quan ban hành, số văn bản, ngày ký và người ký. Một bài báo chính thống luôn ghi rõ cơ quan báo chí, tác giả và thời điểm đăng tải. Một thông báo của doanh nghiệp, trường học hoặc bệnh viện luôn được đăng trên các kênh chính thức của đơn vị.

Ngược lại, nếu thông tin chỉ xuất hiện dưới dạng ảnh chụp màn hình, tin nhắn chuyển tiếp, bài đăng cắt ghép không đầu không đuôi, ảnh không rõ nguồn, video được chia sẻ nhiều lần,... thì người đọc chưa thể xác định ai là người chịu trách nhiệm về nội dung.

> Nếu không biết ai chịu trách nhiệm về thông tin, bạn cũng đừng vội chịu trách nhiệm khi chia sẻ nó.

Có dữ kiện cụ thể

Thông tin đáng tin luôn trả lời được những câu hỏi cơ bản:



Một thông tin càng liên quan đến sức khỏe, thiên tai, tài chính hoặc sự an toàn của người dân thì càng cần có những dữ kiện cụ thể. Ngược lại, những tiêu đề như "sắp có biến lớn", "một loại bệnh rất nguy hiểm đang bùng phát", "có người vừa bị lừa" được đưa ra mà không cung cấp đủ căn cứ để người đọc hiểu đúng sự việc sẽ dễ gây hiểu lầm.

Một thông tin thiếu dữ kiện cụ thể sẽ mang tới nguy cơ bị hiểu theo nhiều cách khác nhau.

Có đầy đủ bối cảnh

Đây là một yếu tố vô cùng quan trọng nhưng lại thường bị bỏ qua nhất. Trong thời đại bùng nổ thông tin trên mạng xã hội, người dùng thường tiếp xúc với những mẩu nội dung rất ngắn như 1 đoạn video 20 giây, 1 bức ảnh chụp vội hay chỉ là 1 câu nói được cắt ra từ bài phát biểu dài. Điều này khiến người xem lầm tưởng rằng mình đã hiểu toàn bộ câu chuyện.

Thực tế, thiếu bối cảnh là một trong những nguyên nhân phổ biến nhất dẫn đến hiểu sai thông tin.

Ví dụ:
Một đoạn video quay cảnh dòng lũ lớn nhưng không ghi rõ thời gian và nơi xảy ra, xuất hiện cùng thông tin về cơn bão đang đổ bộ sẽ dễ dàng gây hiểu sai đây là điều đang xảy ra do cơn bão. Thông tin không thay đổi nhưng bối cảnh đã làm thay đổi cách chúng ta hiểu thông tin.

Một thông tin đúng nhưng thiếu bối cảnh vẫn có thể dẫn đến kết luận sai.

Có thể kiểm tra nguồn

Một thông tin đáng tin không yêu cầu người đọc phải "tin" người đăng. Thay vào đó, nó cho phép bất kỳ ai cũng có thể tự kiểm tra.

Khả năng kiểm chứng chính là yếu tố phân biệt thông tin đáng tin với lời đồn

Không thao túng cảm xúc người đọc

Các thông tin chính thống khi được đưa ra có thể cảnh báo nguy cơ, nhưng sẽ không tìm cách khiến người đọc hoảng sợ hoặc hành động thiếu suy nghĩ hay mang tính kích động, cực đoan

Ngược lại, những tiêu đề như

[khẩn cấp!]

[đọc ngay nếu không sẽ hối hận]

[hãy chia sẻ ngay lập tức]

không làm thông tin đáng tin hơn. Đó chỉ là hành động tạo áp lực tâm lý để người đọc phản ứng trước khi suy nghĩ.

Một thông tin thực sự đáng tin không cần ép buộc người đọc phải tin ngay

03. VÌ SAO VẪN CẦN THẬN TRỌNG VỚI VÙNG AN TOÀN?

Không ít người cho rằng khi một thông tin được đăng trên báo chí hoặc bởi cơ quan nhà nước thì có thể chia sẻ ngay mà không cần đọc kỹ. Đây là một quan niệm chưa đầy đủ. Bởi, có nhiều loại thông tin luôn thay đổi theo thời gian.

Ví dụ như dự báo thời tiết được cập nhật liên tục khi bão thay đổi hướng đi. Hay khuyến cáo y tế có thể điều chỉnh khi có thêm bằng chứng khoa học. Lịch cắt điện, lịch học hoặc phương án phân luồng giao thông có thể thay đổi theo tình hình thực tế. Trong khi đó, các chính sách mới thường trải qua nhiều lần sửa đổi trước khi chính thức có hiệu lực.

Nếu người đọc chỉ nhìn thấy một phiên bản cũ của thông tin rồi tiếp tục chia sẻ, nội dung từng đúng vẫn có thể trở thành thông tin sai lệch trong bối cảnh hiện tại. Điều này cho thấy, độ tin cậy của thông tin luôn gắn với thời điểm và bối cảnh của nó.

Thông tin thuộc vùng an toàn có thể được tiếp nhận và sử dụng nhưng người đọc vẫn nên duy trì thói quen kiểm tra ba yếu tố trước khi chia sẻ hoặc ra quyết định:

- **Thông tin còn mới hay đã cũ?**
- **Có áp dụng đúng cho trường hợp của mình không?**
- **Có bản cập nhật nào mới hơn không?**

Việc dành thêm vài phút để xác nhận sẽ giúp hạn chế rất nhiều việc tiếp nhận sai lệch thông tin và góp phần xây dựng môi trường thông tin đáng tin cậy hơn.

CHECKLIST NHẬN DIỆN

Trước khi tin hoặc chia sẻ, hãy tự hỏi:

- Tôi biết rõ ai là người công bố thông tin này.
- Tôi có thể tìm được nguồn gốc ban đầu.
- Thông tin nêu rõ thời gian, địa điểm và bối cảnh.
- Tôi có thể kiểm chứng bằng ít nhất một nguồn độc lập khác.
- Nội dung không thúc ép tôi phải hành động ngay.

Nếu hầu hết các câu trả lời là "Có", thông tin có khả năng thuộc vùng an toàn. Tuy nhiên, hãy nhớ rằng "an toàn" không đồng nghĩa với "không cần kiểm tra", đặc biệt đối với những thông tin ảnh hưởng trực tiếp đến sức khỏe, tài chính hoặc quyền lợi của bản thân.

Hãy hình thành thói quen tìm nguồn trước khi tìm kết luận. Đây là nền tảng của mọi kỹ năng kiểm chứng trong thời đại số.

[II] VÙNG NGUY HIỂM

Khi thông tin khiến bạn muốn hành động ngay, hãy dừng lại để kiểm chứng.

01. HIỂU VỀ "VÙNG NGUY HIỂM"

Vùng nguy hiểm - nơi thông tin có mức rủi ro cao trong tiếp nhận và lan truyền, do thiếu nguồn gốc xác định, thiếu dữ kiện kiểm chứng hoặc thiếu căn cứ rõ ràng, nhưng lại tác động mạnh đến cảm xúc và thúc ép người tiếp nhận phải tin, chia sẻ hoặc hành động nhanh.

Trong môi trường số, thông tin sai lệch không phải lúc nào cũng được tạo ra với mục đích khiến bạn tin vào một điều không có thật. Nhiều trường hợp, điều người tạo ra thông tin hướng tới không phải niềm tin mà là hành động.

Một bài đăng yêu cầu chuyển tiền ngay để "giữ chỗ" cho người thân đang cấp cứu; một tin nhắn cảnh báo tài khoản ngân hàng sắp bị khóa kèm đường dẫn xác thực; hay một thông tin lan truyền về việc sắp xảy ra thiên tai khiến người dân đổ xô mua tích trữ hàng hóa... Điểm chung của những nội dung này là chúng không cho người đọc đủ thời gian để suy nghĩ. Đó chính là bản chất của vùng nguy hiểm.

Thông tin thuộc vùng này có thể đúng, có thể sai hoặc chỉ đúng một phần. Điều khiến nó trở nên nguy hiểm không nằm ở nội dung mà ở cách tác động đến người tiếp nhận. Thông tin này đánh vào cảm xúc, tạo áp lực về thời gian và thúc đẩy hành động trước khi quá trình kiểm chứng diễn ra.

Thông tin càng thúc ép bạn phải hành động ngay thì càng cần được kiểm chứng trước.

02. KHI NÀO MỘT THÔNG TIN ĐƯỢC COI LÀ "NGUY HIỂM"?

Không phải mọi thông tin sai đều nguy hiểm như nhau. Một thông tin được xếp vào "vùng nguy hiểm" khi việc tin hoặc hành động theo thông tin đó có thể gây hậu quả ngay lập tức cho bản thân hoặc cộng đồng, dù sau đó thông tin được đính chính.

Những nhóm thông tin cần đặc biệt cảnh giác bao gồm:

- **THÔNG TIN LIÊN QUAN ĐẾN TIỀN BẠC VÀ TÀI SẢN:**
  Đầu tư, chuyển tiền, vay vốn, khuyến mại bất thường, trúng thưởng, yêu cầu cung cấp thông tin ngân hàng hoặc mã OTP.

- **THÔNG TIN VỀ SỨC KHỎE VÀ TÍNH MẠNG:**
  Hướng dẫn chữa bệnh, cảnh báo dịch bệnh, thực phẩm, thuốc, thiên tai, tai nạn hoặc các tình huống khẩn cấp.

- **THÔNG TIN VỀ AN NINH, TRẬT TỰ VÀ PHÁP LUẬT:**
  Thông báo giả mạo của cơ quan nhà nước, cảnh báo bắt cóc, khủng bố, truy nã, thay đổi chính sách hoặc quy định chưa được xác nhận.

- **THÔNG TIN LIÊN QUAN ĐẾN DANH DỰ VÀ UY TÍN CỦA CÁ NHÂN, TỔ CHỨC:**
  Cáo buộc, tố giác, hình ảnh hoặc video chưa được kiểm chứng có thể gây tổn hại nghiêm trọng nếu lan truyền.

- **THÔNG TIN ĐƯỢC TẠO HOẶC CHỈNH SỬA BẰNG AI (DEEPFAKE):**
  Hình ảnh, video, giọng nói hoặc tài liệu giả có khả năng khiến người xem tin là thật và đưa ra quyết định sai lầm.

Điểm chung của các thông tin này là chỉ cần một lần chia sẻ hoặc một hành động thiếu kiểm chứng cũng có thể gây thiệt hại lớn.

Trên môi trường số, tốc độ lan truyền luôn nhanh hơn tốc độ kiểm chứng. Chỉ trong vài phút, một thông tin sai có thể được chia sẻ hàng nghìn lần và tạo ra những hậu quả khó khắc phục.

Một quyết định vội vàng có thể dẫn đến mất tiền do chuyển khoản hoặc cung cấp thông tin cho đối tượng lừa đảo, đánh cắp tài khoản, dữ liệu cá nhân hoặc danh tính số, làm tổn hại danh dự, uy tín của cá nhân, tổ chức do chia sẻ thông tin chưa được xác minh hay vô tình tiếp tay cho chiến dịch phát tán tin giả, thao túng dư luận hoặc lừa đảo trên diện rộng. Nhiều hậu quả không thể khắc phục hoàn toàn, ngay cả khi thông tin sau đó được đính chính.

Một trong những minh chứng đau lòng là những vụ việc thương tâm xảy ra khi tin đồn "vỡ đập" được lan truyền vào cuối năm 2025. Không chỉ dừng ở mức độ gây hoang mang dư luận mà hậu quả của những thông tin thất thiệt, bịa đặt xuất hiện trên mạng xã hội đã dẫn đến những cái chết oan uổng cho người dân vùng lũ.

Bi kịch xảy ra hoàn toàn chỉ vì những tin đồn thất thiệt. Tin đồn có thể lan rất nhanh nhưng chỉ một quyết định dựa trên tin chưa được kiểm chứng cũng có thể để lại hậu quả không thể đảo ngược.

CÂU CHUYỆN THƯƠNG TÂM TẠI XÃ HÒA MỸ, ĐẮK LẮK (THÁNG 11/2025):

Khi nghe tin đồn thất thiệt "vỡ đập", cả xóm Mới (thôn Vạn Lộc) đã nháo nhào tháo chạy trong đêm. Trên đường di tản, vợ chồng anh P.V.T (51 tuổi) và chị P.T.N.T (40 tuổi) chở theo con trai 10 tuổi bằng xe máy qua cầu. Mặt cầu bị sụt lún do nước lớn kéo tuột xe xuống hố sâu, người chồng chỉ kịp đẩy con trai lên vùng an toàn trước khi cả hai vợ chồng bị dòng nước lũ cuốn trôi tử vong, bỏ lại các con mồ côi.

TAI NẠN TẠI XÃ PHÚ HÒA 1, TỈNH ĐẮK LẮK (THÁNG 11/2025):

Chị N.T.D.P nghe tin đồn vỡ đập đã vội vã dùng xe máy chở hai con nhỏ đi lánh nạn. Khi băng qua đoạn đường ngập sâu, xe bị tắt máy và ba mẹ con bị dòng nước dữ cuốn trôi. Người dân kịp thời cứu được chị P. và con lớn, riêng cháu nhỏ 6 tuổi bị tuột khỏi tay mẹ và không qua khỏi.

03. VÌ SAO CON NGƯỜI DỄ BỊ CUỐN VÀO VÙNG NGUY HIỂM?

Não bộ con người được thiết kế để phản ứng nhanh trước các mối đe dọa. Khi đối mặt với một thông tin gây sốc hoặc khẩn cấp, chúng ta thường ưu tiên phản ứng để bảo vệ bản thân hơn là dành thời gian phân tích. Các đối tượng phát tán tin giả hoặc lừa đảo thường khai thác chính cơ chế này. Thay vì đưa ra lập luận thuyết phục, họ tạo ra cảm giác rằng "nếu chậm một phút, bạn sẽ mất cơ hội hoặc gặp hậu quả nghiêm trọng".

Bốn cảm xúc thường bị lợi dụng nhất là:

SỢ HÃI

[Tài khoản của bạn sắp bị khóa] [Con bạn đang gặp nạn] [Dịch bệnh đang bùng phát tại khu vực bạn sinh sống]

MONG CẦU

[Đầu tư lợi nhuận cao] [Trúng thưởng] [Việc nhẹ lương cao] [Tình yêu đôi lứa]

SỰ PHẪN NỘ

Thông tin kích động, chia rẽ hoặc tạo cảm giác bất công nhằm lôi kéo chia sẻ.

LÒNG TRẮC ẨN

Kêu gọi quyên góp khẩn cấp, cứu trợ, tìm người thân... nhưng không có thông tin xác thực.

Đây là những cảm xúc hoàn toàn bình thường. Điều quan trọng là nhận ra khi cảm xúc đang dẫn dắt quyết định của mình.

04. NHỮNG DẤU HIỆU CẢNH BÁO

Tạo áp lực về thời gian

Thông tin liên tục nhấn mạnh: **“Chỉ còn hôm nay”, “Làm ngay”, “Khẩn cấp”, “Nếu không sẽ mất cơ hội”**… Áp lực thời gian khiến người đọc bỏ qua bước kiểm tra thông tin.

↳ Sự thật không sợ chờ thêm vài phút để được kiểm chứng.

Thúc ép chia sẻ hoặc thực hiện hành động

Một dấu hiệu rất phổ biến là những lời kêu gọi như **“Chia sẻ ngay”, “Chuyển tiền ngay”, “Bấm vào đường dẫn”, “Cung cấp mã OTP”,**… Thông tin đáng tin cậy thường không yêu cầu người đọc hành động ngay lập tức nếu chưa có đủ căn cứ.

Tác động mạnh vào cảm xúc

Khi đọc một thông tin, hãy tự hỏi: Điều đầu tiên mình cảm thấy là gì? Nếu đó là sự hoảng sợ, tức giận, thương cảm hoặc quá phấn khích, hãy coi đây là tín hiệu cần kiểm chứng thêm. Bởi, cảm xúc không phải bằng chứng.

Thiếu căn cứ nhưng khẳng định chắc chắn

Thông tin có các cụm từ như **"100% đúng"**, **"Nguồn nội bộ"**, **"Không tiện công khai"**, **"Đã được xác nhận"**... nhưng lại không dẫn nguồn, không có tài liệu chứng minh hoặc chỉ dẫn đến các nguồn không đáng tin cậy. Một thông tin đáng tin thường đưa ra bằng chứng trước, kết luận sau, thay vì yêu cầu người đọc tin ngay.

Không rõ nguồn gốc hoặc nguồn không đáng tin cậy

Thông tin không cho biết ai là người đăng đầu tiên, được lấy từ cơ quan hay tổ chức nào, thời điểm xảy ra sự việc, địa điểm cụ thể, không có tài liệu hoặc nguồn chính thức để đối chiếu. Hoặc chỉ ghi rất chung chung như **"Theo nguồn tin…"**, **"Nghe nói…"**, **"Có người gửi/tiết lộ…"**

↳ Nguồn càng mơ hồ, rủi ro càng cao.

Hình ảnh, video hoặc âm thanh quá "thật" nhưng khó kiểm chứng

Sự phát triển của AI khiến nhiều hình ảnh, video và giọng nói giả có chất lượng rất cao. Hãy cảnh giác khi video không rõ thời gian, địa điểm, hình ảnh bị cắt ghép hoặc thiếu ngữ cảnh, chỉ có một đoạn video ngắn nhưng đã kết luận một sự việc lớn, người nổi tiếng hoặc lãnh đạo phát biểu những điều bất thường, âm thanh hoặc hình ảnh có dấu hiệu bị chỉnh sửa. **Có hình ảnh không đồng nghĩa với có sự thật.**

Chỉ đưa ra một góc nhìn

Thông tin chỉ phản ánh một phía, không có ý kiến từ các bên liên quan hoặc không dẫn nguồn độc lập để đối chiếu. Các bài đăng thường sử dụng những câu như “Sự thật không ai dám nói”, “đây là toàn bộ sự thật”, “đừng tin báo chí”... **Thông tin đáng tin cậy thường cho phép người đọc kiểm chứng từ nhiều nguồn khác nhau.**

Lan truyền quá nhanh nhưng không thấy xác nhận chính thức

Một nội dung xuất hiện đồng loạt trên nhiều nhóm mạng xã hội, nhiều tài khoản cá nhân nhưng lại không thấy cơ quan chức năng xác nhận, báo chí chính thống đăng tải hay đơn vị liên quan lên tiếng. Sự lan truyền rộng không phải là bằng chứng cho thấy thông tin là đúng. Nhiều chiến dịch phát tán tin giả được thiết kế để tạo cảm giác “ai cũng đang chia sẻ”, từ đó khiến người đọc dễ tin theo.

05. CHÚNG TA NÊN LÀM GÌ?

Khi nhận thấy một thông tin có dấu hiệu thuộc vùng nguy hiểm, điều đầu tiên cần làm không phải là kiểm chứng ngay mà nên ngắt chuỗi phản ứng cảm xúc. Hãy tạo cho mình một khoảng dừng ngắn trước khi thực hiện bất kỳ hành động nào.

Trong khoảng thời gian đó, hãy tự đặt ra câu hỏi:

- **Ai sẽ được hưởng lợi nếu mình tin vào thông tin này?**
- **Nếu đây là thông tin sai, hậu quả của việc mình chia sẻ sẽ là gì?**
- **Có nguồn chính thức nào xác nhận chưa?**
- **Mình có thực sự cần phải hành động ngay lúc này không?**

Khoảng dừng ngắn đó thường đủ để tránh được nhiều quyết định sai lầm.

Đừng đánh giá một thông tin chỉ bằng nội dung của nó. Hãy đánh giá bằng cách nó khiến bạn phản ứng. Nếu thông tin thúc ép bạn phải hành động trước khi kịp suy nghĩ, đó chính là tín hiệu mạnh nhất của "Vùng nguy hiểm".

CHECKLIST NHẬN DIỆN

- ✔ Thông tin khiến tôi muốn phản ứng ngay.
- ✔ Nội dung tạo cảm giác khẩn cấp.
- ✔ Không có đủ bằng chứng nhưng khẳng định rất chắc chắn.
- ✔ Kêu gọi chia sẻ hoặc hành động ngay.
- ✔ Chủ yếu đánh vào cảm xúc thay vì cung cấp thông tin.

↳ Nếu có từ ba dấu hiệu trở lên, hãy coi đây là thông tin có mức độ rủi ro cao và cần được kiểm chứng trước khi tiếp tục.

Quy tắc 10 giây: Nếu một thông tin khiến bạn muốn chuyển tiền, nhấp vào liên kết, cung cấp thông tin cá nhân, chia sẻ ngay hoặc thay đổi một quyết định quan trọng, hãy dành ít nhất 10 giây để dừng lại và kiểm chứng. Chỉ một khoảng dừng ngắn cũng có thể giúp bạn tránh được những thiệt hại rất lớn.

[III] VÙNG BẪY

Điều nguy hiểm nhất không phải tin giả mà là những thông tin trông giống sự thật.

01. BẢN CHẤT CỦA VÙNG BẪY

Vùng bẫy - nơi thông tin dễ gây nhầm lẫn trong quá trình tiếp nhận và lan truyền, trong đó hình ảnh, video, giọng nói, số liệu, trích dẫn, danh nghĩa cơ quan, tổ chức hoặc một phần sự thật được sử dụng như bằng chứng ban đầu, nhưng nguồn gốc, bối cảnh, tính toàn vẹn hoặc tính xác thực của các yếu tố đó chưa được xác lập.

Nếu vùng nguy hiểm khiến bạn phản ứng bằng cảm xúc thì vùng bẫy lại khai thác một điểm yếu khác: niềm tin vào những gì chúng ta nhìn thấy và nghe thấy. Con người có xu hướng tin vào những thông tin có vẻ "có bằng chứng" như một bức ảnh/video, một đoạn ghi âm, một tài liệu có dấu đỏ hay lời phát biểu của người nổi tiếng. Những yếu tố này tạo cảm giác rằng "mình đã tận mắt chứng kiến", từ đó làm giảm sự cảnh giác. Chính vì vậy, vùng bẫy là nhóm thông tin nguy hiểm nhất. Chúng không yêu cầu người đọc phải tin một điều vô lý mà khiến người đọc tin vào một điều có vẻ rất hợp lý.

Trong thời đại AI, ranh giới giữa thật và giả ngày càng khó phân biệt. Hình ảnh có thể được tạo ra hoàn toàn bằng AI. Giọng nói có thể được sao chép chỉ từ vài giây ghi âm. Video có thể ghép khuôn mặt và cử động môi để tạo nên những phát biểu chưa từng tồn tại.

↳ Nhìn thấy không đồng nghĩa với sự thật. Nghe thấy không đồng nghĩa với bằng chứng.

02. NHỮNG "CHIẾC BẪY" PHỔ BIẾN

Bẫy hình ảnh/video

Một bức ảnh/video có thể là thật nhưng bị gắn sai thời gian, sai địa điểm hoặc sai sự kiện. Đây là hình thức thường gặp khi hình ảnh từ một thiên tai hoặc vụ việc cũ được chia sẻ như thể vừa mới xảy ra.

Người xem cần phải đặt ra câu hỏi để kiểm tra:

Ảnh/video xuất hiện lần đầu khi nào?

Có phải đã từng được dùng trong sự kiện khác?

Có dấu hiệu chỉnh sửa hoặc cắt ghép không?

Video thường có sức thuyết phục cao vì ghi lại hình ảnh và diễn biến của sự việc. Tuy nhiên, chỉ cần cắt bỏ một đoạn, thay đổi thứ tự hoặc thêm phụ đề sai, nội dung có thể dẫn người xem đến một kết luận hoàn toàn khác.

Đừng chỉ xem lướt qua, hãy tìm bản đầy đủ hoặc nguồn đăng tải ban đầu

Bẫy âm thanh

Các công cụ AI hiện nay có thể tạo giọng nói giống người thật với độ chính xác rất cao. Điều này làm gia tăng nguy cơ giả mạo người thân, lãnh đạo doanh nghiệp hoặc cơ quan chức năng để lừa chuyển tiền hoặc lấy thông tin cá nhân.

Nếu nhận được cuộc gọi bất thường yêu cầu chuyển tiền hoặc cung cấp dữ liệu nhạy cảm, hãy xác minh lại qua một kênh liên lạc khác

Bẫy ngữ cảnh

Đây là dạng bẫy khó nhận biết nhất. Không phải mọi thông tin sai lệch đều là bịa đặt. Nhiều nội dung sử dụng dữ kiện có thật nhưng bị tách khỏi bối cảnh ban đầu, lược bỏ các chi tiết quan trọng hoặc chỉ chọn những thông tin có lợi cho một góc nhìn nhất định.

Khi thiếu ngữ cảnh, người đọc rất dễ đi đến kết luận sai dù từng chi tiết riêng lẻ đều có thể đúng. Một nửa sự thật đôi khi nguy hiểm hơn một lời nói dối hoàn toàn vì nó khiến người đọc ít nghi ngờ hơn.

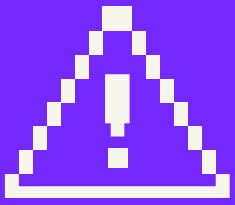
Thông tin đúng nhưng thiếu ngữ cảnh vẫn có thể dẫn đến nhận thức sai.

Bẫy AI

AI có thể tạo văn bản, hình ảnh, video và âm thanh với chất lượng ngày càng cao. Tuy nhiên, AI không "biết" điều gì là đúng hay sai, nó tạo ra nội dung dựa trên dữ liệu đã học. Vì vậy, AI có thể tạo ra những thông tin nghe rất thuyết phục nhưng hoàn toàn không có thật hoặc pha trộn giữa đúng và sai.

Khi sử dụng AI để tìm kiếm hoặc tóm tắt thông tin, hãy luôn coi kết quả là điểm khởi đầu để kiểm chứng chứ không phải là kết luận cuối cùng.

NGUYÊN TẮC "3 KHÔNG" KHI GẶP VÙNG BẪY



- ✖ **KHÔNG** vội tin chỉ vì có hình ảnh, video hoặc giọng nói.
- ✖ **KHÔNG** chia sẻ khi chưa xác minh được nguồn gốc và bối cảnh.
- ✖ **KHÔNG** hành động chỉ dựa trên một nguồn thông tin duy nhất.

[CHƯƠNG III]

03

LA BÀN KIỂM CHỨNG

Từ nhận diện ⟶ hành động

Biết cách nhận diện một thông tin thuộc vùng an toàn, vùng nguy hiểm hay vùng bẫy mới chỉ là bước đầu. Trong thực tế, các thông tin sai lệch thường không xuất hiện dưới dạng rõ ràng mà được "ngụy trang" rất khéo léo. Chúng có thể sử dụng hình ảnh thật, số liệu thật hoặc một phần sự thật để tạo cảm giác đáng tin cậy.

Chính vì vậy, thay vì cố gắng ghi nhớ thật nhiều dấu hiệu, người đọc cần hình thành một quy trình tư duy thống nhất khi gặp bất kỳ thông tin nào. Mục tiêu không phải là trở thành chuyên gia kiểm chứng mà là biết dừng lại đúng lúc, đặt đúng câu hỏi và tìm đúng nguồn trước khi tin hoặc chia sẻ.

Không phải ai cũng có thể phát hiện ngay tin giả nhưng ai cũng có thể tránh trở thành người lan truyền tin giả nếu biết dừng lại để kiểm chứng.

[I] "BẪY TƯ DUY" KHI TIẾP NHẬN THÔNG TIN

Con người không chỉ bị đánh lừa bởi công nghệ hay tin giả mà còn dễ bị đánh lừa bởi chính cách bộ não xử lý thông tin. Trong tâm lý học có một khái niệm gọi là **thiên kiến nhận thức (Cognitive Biases)**. Đó là một quá trình suy nghĩ có hệ thống do bộ não con người có xu hướng đơn giản hóa việc xử lý thông tin thông qua bộ lọc kinh nghiệm và sở thích cá nhân. Chúng giúp chúng ta đưa ra quyết định nhanh hơn trong cuộc sống hằng ngày nhưng cũng khiến chúng ta dễ tin vào thông tin sai lệch nếu không nhận ra.

Việc xác định những sai lầm phổ biến khi tiếp cận thông tin sẽ giúp chúng ta tránh đưa ra kết luận sai và có phương pháp kiểm chứng phù hợp.

01. CHỈ TÌM THÔNG TIN MÌNH MUỐN TIN

Đây gọi là thiên kiến xác nhận (confirmation bias)

Con người thường thích tìm các thông tin củng cố niềm tin sẵn có, đồng thời bỏ qua những bằng chứng trái ngược.

Mỗi ngày chúng ta phải tiếp nhận một lượng thông tin khổng lồ. Thay vì đánh giá mọi thông tin từ đầu, não bộ thường sử dụng những niềm tin và kinh nghiệm sẵn có như một "lối tắt nhận thức" để đưa ra kết luận nhanh hơn. Khi một thông tin phù hợp với điều chúng ta vốn đã nghĩ, não bộ có xu hướng cảm thấy nó hợp lý, quen thuộc và đáng tin, từ đó giảm động lực kiểm chứng thêm.

Ngược lại, những thông tin trái với niềm tin sẵn có thường tạo ra cảm giác khó chịu hoặc mâu thuẫn trong suy nghĩ. Để duy trì sự nhất quán về nhận thức, con người dễ nghi ngờ, tìm điểm sai hoặc thậm chí bác bỏ những thông tin đó trước khi đánh giá một cách khách quan.

Ví dụ:
Một người tin rằng một phương pháp chữa bệnh "thần kỳ" hiệu quả sẽ chỉ đọc và chia sẻ các bài viết ủng hộ, trong khi bỏ qua các nghiên cứu khoa học cho thấy phương pháp đó không có hiệu quả hoặc có thể gây hại.

Hãy chủ động tìm cả những bằng chứng phản biện, không chỉ những gì đồng tình với quan điểm của mình.

02. ĐỒNG NHẤT "NỔI TIẾNG" VỚI "ĐÁNG TIN"

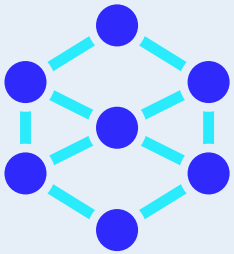
Khái niệm **"Hiệu ứng hào quang" (Halo Effect)** là xu hướng đánh giá một người hoặc một tổ chức đáng tin cậy trong mọi lĩnh vực chỉ vì họ nổi bật hoặc có uy tín ở một lĩnh vực khác. Điều này khiến chúng ta dễ tin vào phát biểu của người nổi tiếng, người có nhiều người theo dõi hoặc một thương hiệu quen thuộc.

Thế nhưng cần phải phân định rõ ràng rằng: Một người nổi tiếng có thể có sức ảnh hưởng lớn trong lĩnh vực giải trí, thể thao hoặc nghệ thuật nhưng điều đó không đồng nghĩa với việc họ là nguồn thông tin đáng tin cậy về y tế, tài chính hay pháp luật.

Hãy đánh giá thông tin dựa trên chuyên môn và bằng chứng, không chỉ dựa trên danh tiếng của người phát ngôn.

03. TIN THEO "HIỆU ỨNG ĐÁM ĐÔNG"

Hiệu ứng đám đông là xu hướng tin hoặc làm theo một điều chỉ vì thấy nhiều người khác cũng tin hoặc làm như vậy. Khi đứng trước những thông tin chưa rõ đúng sai, con người thường xem phản ứng của người khác như một tín hiệu để xác định điều gì là đúng hoặc đáng tin. Xét về mặt tâm lý, chúng ta thường cảm thấy yên tâm hơn khi hành động giống số đông.



Tuy nhiên, số lượng người tin vào một thông tin không phải là bằng chứng cho tính chính xác của thông tin đó.

Nhiều người có thể cùng chia sẻ một thông tin vì tò mò, vì cảm xúc, muốn cảnh báo người khác hoặc đơn giản là vì họ cũng chưa kiểm chứng nguồn tin ban đầu.

Trong môi trường số, hiệu ứng đám đông còn được khuếch đại bởi các chỉ số hiển thị như số lượt thích, chia sẻ, bình luận, lượt xem hoặc việc một chủ đề liên tục xuất hiện trên bảng xu hướng. Những tín hiệu này dễ khiến người dùng vô thức cho rằng nội dung đang được nhiều người quan tâm thì cũng phải đúng hoặc đáng tin cậy, trong khi thực tế chúng chỉ phản ánh mức độ lan truyền, không phản ánh tính chính xác.

Một thông tin không trở thành sự thật chỉ vì có nhiều người tin vào nó.

04. CHO RẰNG "CHƯA ĐƯỢC CHỨNG MINH LÀ SAI" CÓ NGHĨA "LÀ ĐÚNG"

Một lỗi tư duy phổ biến khi tiếp nhận thông tin là cho rằng nếu chưa có bằng chứng bác bỏ thì thông tin đó có thể xem là đúng. Xét về logic, đây được gọi là **ngụy biện từ sự thiếu hiểu biết (Appeal to Ignorance)**.

Ngụy biện này xảy ra khi một kết luận được xem là đúng chỉ vì chưa ai chứng minh được nó sai, hoặc ngược lại, cho rằng một điều là sai chỉ vì chưa có bằng chứng chứng minh nó đúng. Tuy nhiên, việc thiếu bằng chứng không phải là bằng chứng cho bất kỳ kết luận nào.

Khi tiếp nhận thông tin trên không gian mạng, kiểu lập luận này xuất hiện rất thường xuyên. Ví dụ như:

[Nếu tin này là giả thì đã có người đính chính rồi]

[Chưa có cơ quan nào bác bỏ nên chắc là thật]

[Chưa ai chứng minh sản phẩm này không hiệu quả thì cứ tin là có tác dụng]

Trên thực tế, con người thường cảm thấy không thoải mái trước sự không chắc chắn. Khi chưa có đủ thông tin, não bộ có xu hướng vội vàng lấp đầy khoảng trống bằng một kết luận có vẻ hợp lý, thay vì chấp nhận rằng "chưa đủ dữ liệu để kết luận". Trong hoạt động kiểm chứng thông tin, điều quan trọng không phải là tìm xem đã có ai phủ nhận hay chưa mà là tìm xem có những bằng chứng đáng tin cậy nào.

Kiểm chứng là tìm kiếm bằng chứng, không phải tìm kiếm sự đồng tình.

[II] QUY TRÌNH “5D” TRÁNH BẪY THÔNG TIN

BƯỚC 1 DỪNG LẠI

Khi một thông tin khiến bạn bất ngờ, lo lắng, tức giận hoặc quá phấn khích, đừng phản ứng ngay. Hãy dành một khoảng thời gian ngắn để bình tĩnh trước khi bấm chia sẻ, bình luận, chuyển tiền hoặc làm theo hướng dẫn. Khoảng dừng ngắn này là “hàng rào” đầu tiên giúp ngăn cảm xúc chi phối quyết định.

Dừng lại không có nghĩa là bỏ qua thông tin. Dừng lại là giữ lại một khoảng thời gian ngắn để không bị kéo đi bởi cảm xúc, tốc độ lan truyền hoặc lời kêu gọi hành động. Có những việc chỉ cần chậm lại vài giây, hậu quả đã khác.

Nếu một thông tin muốn bạn hành động ngay, đó cũng là lúc bạn nên dừng lại.

BƯỚC 2 ĐỌC TOÀN BỘ

Đừng chỉ đọc tiêu đề hoặc xem vài giây đầu của video. Hãy đọc toàn bộ nội dung để hiểu: **Sự việc xảy ra là gì? Có đầy đủ thời gian, địa điểm và bối cảnh không? Có dẫn nguồn hay bằng chứng không?**

Nhiều thông tin sai lệch được tạo ra bằng cách cắt xén nội dung hoặc sử dụng tiêu đề gây hiểu nhầm.

Tiêu đề được tạo ra để thu hút sự chú ý không phải lúc nào cũng phản ánh đầy đủ nội dung.

BƯỚC 3 ĐỐI CHIẾU NGUỒN

Đối chiếu không có nghĩa là tìm thật nhiều bài giống nhau. Nhiều bài giống nhau đôi khi chỉ là nhiều bản sao chép lại từ cùng một nguồn chưa rõ. Đối chiếu đúng là tìm xem có nguồn độc lập và đáng tin nào xác nhận các dữ kiện chính hay không. Hãy tự đặt câu hỏi:

- Thông tin xuất phát từ đâu?
- Có phải nguồn gốc ban đầu không?
- Có cơ quan, tổ chức hoặc chuyên gia có thẩm quyền xác nhận không?
- Có ít nhất hai nguồn độc lập cùng đưa tin không?

Nếu không thể xác định được nguồn gốc, hãy coi đây là thông tin chưa đủ căn cứ.

Nguồn đầu tiên luôn quan trọng hơn người chia sẻ cuối cùng.

BƯỚC 4 ĐÁNH GIÁ MỤC ĐÍCH

Không chỉ xem thông tin nói gì, hãy nghĩ đến điều nó muốn bạn làm. Liệu thông tin đó muốn bạn chia sẻ hay bấm vào liên kết, muốn bạn chuyển tiền hay đang muốn bạn phải tin ngay.

Một thông tin đáng tin cậy thường cho người đọc đủ dữ kiện để tự đánh giá. Ngược lại, thông tin có dấu hiệu thao túng thường tìm cách kéo người đọc vào một hành động cụ thể trước khi họ kịp kiểm tra. Nó có thể dùng các từ như:

[khẩn cấp] [chỉ còn hôm nay]

[ai không chia sẻ là vô tâm]

[hãy chuyển ngay] [bấm vào đây để xác nhận]

Càng thúc ép hành động nhanh, càng cần đọc chậm lại.

Hiểu được mục đích của thông tin sẽ giúp bạn nhận ra nhiều dấu hiệu thao túng.

BƯỚC 5 ĐƯA RA QUYẾT ĐỊNH

Sau khi hoàn thành 4 bước trên, hãy lựa chọn cách ứng xử phù hợp.

- **Nếu thông tin có đủ căn cứ:**
  Có thể tiếp nhận và chia sẻ một cách có trách nhiệm.
- **Nếu thông tin còn nghi vấn:**
  Chưa chia sẻ, tiếp tục kiểm chứng.
- **Nếu thông tin có dấu hiệu giả mạo hoặc lừa đảo:**
  Không tương tác, báo cáo cho nền tảng hoặc cơ quan có thẩm quyền khi cần thiết.

Không chia sẻ cũng là một hành động có trách nhiệm khi thông tin chưa được kiểm chứng.

[III] BIẾN KIỂM CHỨNG THÀNH PHẢN XẠ

Một lần kiểm chứng có thể giúp tránh một tin sai. Nhưng thói quen kiểm chứng mới giúp người đọc có nền tảng an toàn hơn trong môi trường thông tin hiện nay.

Ban đầu, việc dừng lại có thể gây cảm giác chậm trễ. Việc đọc toàn bộ có thể mất thêm vài phút. Việc đối chiếu nguồn có thể chưa quen. Nhưng khi lặp lại đủ nhiều, những bước đó sẽ dần trở thành phản xạ.

Khi thấy chữ "GẤP", người đọc biết tự chậm lại. Khi thấy tin không nguồn, người đọc biết tự dừng. Khi thấy ảnh gây sốc, người đọc biết tự hỏi ảnh từ đâu. Khi thấy video quá thuyết phục, người đọc biết tìm bản đầy đủ. Khi thấy người quen gửi tin, người đọc vẫn biết kiểm tra, đối chiếu thêm một nguồn.

Đó là lúc kiểm chứng không còn là một thao tác tình huống, mà trở thành một cách ứng xử trong môi trường số. La bàn kiểm chứng vì vậy không chỉ giúp xử lý một tin cụ thể. Nó giúp hình thành một thói quen: không để cảm xúc quyết định thay mình, không để tốc độ kéo mình đi, không để hình thức làm mình quên nguồn gốc.

Nếu bản đồ giúp người đọc nhận ra mình đang đứng ở đâu, thì la bàn giúp người đọc chọn cách đi tiếp.

THỰC HÀNH NHẬN DIỆN

Dưới đây là một số tình huống thường gặp. Hãy thử xác định thông tin đang thuộc vùng nào trước khi xem gợi ý.

TÌNH HUỐNG 1

BẠN NHẬN THẤY ĐIỀU GÌ?

- CÓ NGUỒN CHÍNH THỨC
- CÓ ĐƯỜNG DẪN XÁC MINH
- NỘI DUNG TẠO ÁP LỰC VỀ THỜI GIAN
- YÊU CẦU THỰC HIỆN HÀNH ĐỘNG NGAY
- CÓ DẤU HIỆU GIẢ MẠO

ĐÁP ÁN

Thông tin này thuộc vùng nguy hiểm

Lý do:

- TẠO CẢM GIÁC KHẨN CẤP.
- YÊU CẦU NGƯỜI ĐỌC HÀNH ĐỘNG NGAY.
- KHÔNG DẪN NGUỒN CHÍNH THỨC.
- CÓ NGUY CƠ DẪN TỚI TRANG WEB GIẢ MẠO ĐỂ ĐÁNH CẮP THÔNG TIN.

[BÀI HỌC RÚT RA]

Các ngân hàng và cơ quan quản lý thường thông báo qua kênh chính thức và không yêu cầu khách hàng xác thực thông tin qua các liên kết không rõ nguồn gốc.

TÌNH HUỐNG 2

Một video ghi lại cảnh nước lũ cuốn trôi nhiều phương tiện được chia sẻ với chú thích:

Đây là tình hình tại tỉnh X sáng nay

BẠN NÊN LÀM GÌ?

- KIỂM TRA THỜI GIAN ĐĂNG TẢI.
- TÌM NGUỒN ĐĂNG ĐẦU TIÊN.
- TÌM KIẾM HÌNH ẢNH HOẶC VIDEO TƯƠNG TỰ
- ĐỐI CHIẾU VỚI THÔNG TIN TỪ BÁO CHÍ HOẶC CƠ QUAN CHỨC NĂNG.

ĐÁP ÁN

Video có thể là thật nhưng được sử dụng sai thời gian hoặc sai địa điểm. Đây là vùng bẫy.

[BÀI HỌC RÚT RA]

Điều dễ khiến người xem nhầm lẫn là hình ảnh chân thực tạo cảm giác sự việc đang diễn ra.

TÌNH HUỐNG 3

Bạn nhận được cuộc gọi từ một người có giọng nói rất giống người thân, yêu cầu chuyển tiền gấp vì đang gặp sự cố

BẠN SẼ LÀM GÌ?

- GIỮ BÌNH TĨNH
- KHÔNG CHUYỂN TIỀN NGAY
- GỌI LẠI BẰNG SỐ ĐIỆN THOẠI QUEN THUỘC HOẶC LIÊN HỆ QUA MỘT KÊNH KHÁC ĐỂ XÁC MINH

[PHÂN TÍCH]

Đây là một trong những hình thức lừa đảo sử dụng AI tạo giọng nói (voice cloning). Công nghệ hiện nay có thể tái tạo giọng nói rất giống người thật chỉ từ một đoạn ghi âm ngắn. Vì vậy, giọng nói không còn là bằng chứng xác thực tuyệt đối.

[CHƯƠNG IV]

04

THẾ GIỚI THÔNG TIN THỜI AI

Từ “thấy mới tin”

→ “thấy cũng phải kiểm tra”

Trước đây, chúng ta vẫn luôn tin rằng **“Thấy tận mắt, nghe tận tai”** là cách đáng tin cậy nhất để xác nhận một thông tin. Một bức ảnh, đoạn video hay một bản ghi âm từng được xem là những bằng chứng có giá trị. Tuy nhiên, sự phát triển của trí tuệ nhân tạo (AI) đang làm thay đổi nguyên tắc đó.

Ngay cả những gì mắt thấy, tai nghe cũng có thể được tạo ra hoặc thay đổi bằng công nghệ. Nhiều nội dung do AI tạo ra gần như không thể phân biệt bằng mắt thường hoặc chỉ qua một lần xem.

Điều này không có nghĩa chúng ta phải nghi ngờ tất cả những gì nhìn thấy trên Internet. Nhưng nó nhắc nhở rằng:

Hình ảnh, âm thanh và video không còn là bằng chứng tuyệt đối.

Chúng cần được đặt trong bối cảnh, đối chiếu với các nguồn khác và kiểm chứng trước khi tin tưởng hoặc chia sẻ

01. AI LÀM THAY ĐỔI CÁCH THÔNG TIN ĐƯỢC TẠO RA

Trước đây, việc tạo ra một bài viết, bức ảnh hay video thường đòi hỏi kiến thức chuyên môn, kỹ năng và nhiều thời gian. Ngày nay, AI có thể hỗ trợ tạo ra hàng nghìn nội dung chỉ trong vài phút. Không chỉ những tổ chức chuyên nghiệp mà gần như bất kỳ cá nhân nào cũng có thể tạo ra lượng lớn nội dung với chất lượng ngày càng cao. Điều này mang lại nhiều lợi ích trong học tập, nghiên cứu, truyền thông và sáng tạo.

Tuy nhiên, nó cũng mang đến nhiều rủi ro khi lượng thông tin trên Internet tăng lên nhanh chóng, có những nội dung chính xác nhưng cũng có không ít nội dung sai lệch, gây hiểu nhầm hoặc được tạo ra nhằm thao túng nhận thức, lừa đảo hoặc trục lợi.

Thực tế này đặt ra yêu cầu mới đối với mỗi người dân: không chỉ biết tiếp nhận thông tin mà còn phải biết đánh giá chất lượng và độ tin cậy của thông tin trong thời đại AI.

AI có thể tạo ra nội dung rất giống thật, nhưng "giống thật" không đồng nghĩa với "đúng sự thật"

02. AI LÀM THAY ĐỔI CÁCH THÔNG TIN ĐƯỢC LAN TRUYỀN

Không chỉ thay đổi cách tạo ra thông tin, AI và các thuật toán của nền tảng số còn làm thay đổi cách thông tin được phân phối đến người dùng.

Thay vì tất cả mọi người cùng nhìn thấy một nội dung giống nhau, các nền tảng ngày nay liên tục phân tích hành vi, sở thích, lịch sử tìm kiếm và tương tác của từng cá nhân để ưu tiên hiển thị những nội dung mà họ có nhiều khả năng quan tâm. Việc cá nhân hóa này giúp người dùng tiếp cận thông tin phù hợp hơn với nhu cầu của mình nhưng cũng có những hệ quả nhất định.

Theo thời gian, mỗi người có thể dần sống trong một "bong bóng thông tin" (Filter Bubble), phần lớn nội dung chúng ta nhìn thấy đều tương đồng với sở thích, quan điểm và hành vi trước đó.

Khi ít tiếp xúc với các góc nhìn khác, người dùng có thể hình thành cảm giác rằng những gì mình đang thấy phản ánh đầy đủ thực tế.

Nếu tình trạng này kéo dài, người dùng dễ rơi vào "buồng vang" (Echo Chamber). Một môi trường thông tin mà các quan điểm và niềm tin tương đồng liên tục được lặp lại và củng cố bởi những người có cùng cách nhìn. Trong buồng vang, các thông tin hoặc ý kiến trái chiều ít có cơ hội xuất hiện hoặc dễ bị bác bỏ khiến người dùng ngày càng khó tiếp cận những góc nhìn khác. Sự lặp lại liên tục này có thể làm một thông tin trở nên quen thuộc và có vẻ đáng tin hơn, ngay cả khi thông tin đó chưa được kiểm chứng hoặc thiếu căn cứ.

Trong môi trường số, mức độ lan truyền không phản ánh mức độ chính xác.

Một nội dung được xuất hiện nhiều lần trên bảng tin, được nhiều người chia sẻ hoặc bình luận không đồng nghĩa với việc nội dung đó là sự thật.

03. AI LÀM THAY ĐỔI CÁCH THÔNG TIN ĐƯỢC TẠO RA

Deepfake: Khi "mắt thấy, tai nghe" chưa chắc là sự thật

Deepfake là công nghệ sử dụng AI để tạo hoặc chỉnh sửa hình ảnh, âm thanh và video với mức độ chân thực rất cao. Không giống các kỹ thuật chỉnh sửa truyền thống, deepfake có thể tái tạo rất chính xác khuôn mặt, biểu cảm, cử động môi và giọng nói của một người chỉ từ một số lượng nhỏ hình ảnh hoặc vài giây ghi âm.

Deepfake mang lại nhiều lợi ích trong điện ảnh, giáo dục và sáng tạo nội dung nhưng cũng có thể bị lợi dụng cho nhiều mục đích xấu như giả mạo danh tính, lừa đảo tài chính, bôi nhọ danh dự hoặc phát tán thông tin sai lệch.

Trong thực tế, AI có thể tạo ra nhiều dạng nội dung giả mạo khác nhau:

- **Tái tạo giọng nói của một người từ một đoạn ghi âm rất ngắn.**
- **Ghép khuôn mặt của một người vào video của người khác.**
- **Tạo video người nổi tiếng hoặc người có uy tín phát biểu những nội dung họ chưa từng nói.**
- **Tạo hình ảnh về các sự kiện chưa từng xảy ra nhưng có chất lượng rất giống ảnh chụp thật.**
- **Tạo các cuộc họp trực tuyến giả với hình ảnh và giọng nói của nhiều người tham gia.**

Sự phát triển của công nghệ khiến những nội dung này ngày càng khó phát hiện bằng mắt thường, đặc biệt khi được chia sẻ trên mạng xã hội với chất lượng hình ảnh thấp hoặc thời lượng ngắn.

Ví dụ thực tế: **Cuộc gọi giả giọng Tổng thống Mỹ Joe Biden (2024)**

Đầu năm 2024, cử tri tại bang New Hampshire nhận được cuộc gọi tự động giả mạo Joe Biden để khuyên họ không nên đi bỏ phiếu cho ông trong cuộc bầu cử sơ bộ. Dựa trên phân tích của các chuyên gia, giọng nói được tạo ra bằng công nghệ deepfake.

Cuộc gọi đã phản ánh nguy cơ mà AI có thể mang lại, đặc biệt là ở những sự kiện quan trọng như cho cuộc bầu cử Mỹ năm 2024, làm dấy lên lo ngại gieo rắc tin vịt hoặc lan tỏa thông tin không đúng sự thật.

Trong kỷ nguyên AI, "thấy tận mắt, nghe tận tai" không còn đủ để xác minh danh tính. Hãy xác minh bằng nhiều kênh độc lập trước khi đưa ra quyết định quan trọng.

AI Hallucination: Khi AI "bịa" thông tin

Một trong những hiểu lầm phổ biến nhất hiện nay là cho rằng AI "biết mọi thứ". Thực tế, các hệ thống AI không hoạt động như một cơ sở dữ liệu lưu trữ sự thật mà tạo ra câu trả lời bằng cách "đoán ý" dựa trên dữ liệu đã được huấn luyện.

Vì vậy, trong một số trường hợp, AI có thể tạo ra những thông tin không chính xác hoặc hoàn toàn không có thật nhưng vẫn trình bày rất logic, mạch lạc và tự tin. Hiện tượng này được gọi là **AI Hallucination (ảo giác của AI)**.

Hiện tượng AI Hallucination có thể xuất hiện dưới nhiều dạng khác nhau như trích dẫn một văn bản pháp luật không tồn tại, gán phát biểu cho một chuyên gia chưa từng nói, đưa ra số liệu thống kê không có nguồn, tạo danh mục tài liệu tham khảo giả nhưng trình bày giống tài liệu thật hay kết hợp các thông tin đúng và sai trong cùng một câu trả lời. Điều này đặc biệt nguy hiểm khi người dùng sử dụng AI cho các lĩnh vực như pháp luật, y tế, tài chính hoặc nghiên cứu mà không kiểm tra lại nguồn gốc thông tin.

> AI có thể viết rất hay nhưng viết hay không đồng nghĩa với viết đúng. Mọi thông tin quan trọng cần được đối chiếu với nguồn chính thức hoặc các nguồn đáng tin cậy khác.

Giả mạo bằng AI: Khi lừa đảo trở nên tinh vi hơn

Sự phát triển của AI không chỉ tạo ra các nội dung giả mạo như hình ảnh hay video mà còn làm thay đổi cách thức các đối tượng lừa đảo tiếp cận nạn nhân. Nếu trước đây, email hay tin nhắn giả thường có nhiều lỗi chính tả, câu văn lủng củng, dễ nhận biết thì ngày nay AI có thể giúp tạo ra những nội dung rất tự nhiên, đúng ngữ cảnh và được cá nhân hóa theo từng người nhận.

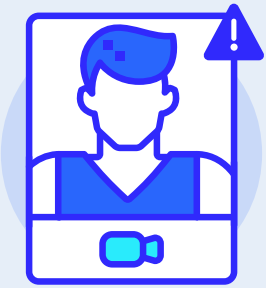
Ví dụ thực tế: **Bị lừa 25 triệu USD sau cuộc gọi của giám đốc tài chính giả**

Tháng 2 năm 2024, cảnh sát Hong Kong công bố một vụ lừa đảo đặc biệt nghiêm trọng. Một nhân viên tài chính của một doanh nghiệp đa quốc gia đã tham gia cuộc họp trực tuyến với người mà anh tin là Giám đốc Tài chính cùng nhiều đồng nghiệp. Trong cuộc họp, những người này yêu cầu anh thực hiện một số giao dịch chuyển tiền khẩn cấp.

Tin rằng cuộc họp là thật, nhân viên này đã thực hiện nhiều lệnh chuyển khoản với tổng giá trị khoảng 200 triệu đô la Hong Kong (tương đương khoảng 25 triệu USD).

Sau quá trình điều tra, cảnh sát xác định toàn bộ những người xuất hiện trong cuộc họp đều được tạo bằng công nghệ deepfake, từ hình ảnh khuôn mặt đến giọng nói. Đây được xem là một trong những vụ lừa đảo sử dụng deepfake lớn nhất được công bố trên thế giới.

Những nội dung này thường được kết hợp với các kỹ thuật thao túng tâm lý như tạo cảm giác khẩn cấp, đánh vào lòng tin hoặc lòng tham để thúc đẩy người dùng hành động nhanh mà không kịp kiểm chứng.



Vụ việc cho thấy ngay cả khi nhìn thấy khuôn mặt quen thuộc và nghe đúng giọng nói, chúng ta vẫn không thể khẳng định danh tính của người đang giao tiếp.

Đối với các yêu cầu liên quan đến chuyển tiền, cung cấp thông tin cá nhân hoặc phê duyệt giao dịch quan trọng, cần xác minh lại thông qua một kênh liên lạc độc lập như gọi điện trực tiếp đến số điện thoại đã được xác thực hoặc gặp trực tiếp nếu có thể.

04. NHÃN AI VÀ CÔNG CỤ PHÁT HIỆN: CẦN THIẾT NHƯNG CHƯA ĐỦ

Nhiều nền tảng đã bắt đầu yêu cầu người sáng tạo gắn nhãn nội dung AI, nhất là những nội dung có hình ảnh, âm thanh hoặc video trông chân thực. YouTube yêu cầu nhà sáng tạo khai báo khi nội dung thực tế được tạo ra hoặc chỉnh sửa đáng kể bằng AI hoặc phương tiện tổng hợp, để nền tảng có thể gắn nhãn cho người xem. TikTok cũng yêu cầu người sáng tạo gắn nhãn tất cả nội dung AI có hình ảnh, âm thanh hoặc video chân thực theo hướng dẫn cộng đồng của nền tảng.

Việc gắn nhãn là cần thiết, vì nó giúp người xem có thêm bối cảnh. Tuy nhiên, nhãn không thể thay thế kiểm chứng. Có nhãn AI không có nghĩa là nội dung an toàn. Một video được gắn nhãn AI vẫn có thể được dùng để dẫn dụ đầu tư, kích động cảm xúc hoặc làm người xem hiểu sai. Ngược lại, không có nhãn AI cũng không có nghĩa là nội dung thật, vì người đăng có thể không gắn nhãn hoặc nội dung đã bị đăng lại qua nhiều kênh khác nhau.

Các công cụ phát hiện AI cũng có vai trò hỗ trợ, nhưng không nên được xem là "máy phán quyết". Con người cần biết sử dụng công cụ nhưng không được giao toàn bộ phán đoán cho công cụ. Một kết quả "có thể là AI" không đủ để kết luận nội dung sai. Một kết quả "không phát hiện AI" cũng không đủ để kết luận nội dung thật.

Kiểm chứng vẫn phải quay lại những câu hỏi cơ bản: nguồn ở đâu, bản gốc là gì, bối cảnh ra sao, ai chịu trách nhiệm, có nguồn độc lập xác nhận không và nội dung đang muốn người đọc làm gì.

05. LÀM THẾ NÀO ĐỂ GIẢM THIỂU RỦI RO?

Không có phương pháp nào giúp phát hiện tất cả các nội dung giả mạo do AI tạo ra chính xác 100%. Tuy nhiên, mỗi người có thể giảm đáng kể nguy cơ bị lừa bằng cách hình thành một số nguyên tắc đơn giản:

- **Không đưa ra quyết định quan trọng chỉ dựa trên hình ảnh, video hoặc giọng nói.**
- **Luôn kiểm tra nguồn gốc của thông tin trước khi chia sẻ.**
- **Đối chiếu với các nguồn chính thức.**
- **Không chuyển tiền hoặc cung cấp thông tin cá nhân khi chưa xác minh danh tính của người yêu cầu qua một kênh độc lập.**
- **Khi sử dụng AI để học tập hoặc làm việc, luôn kiểm tra lại các số liệu, trích dẫn và văn bản pháp luật từ nguồn chính thức.**

Công cụ chỉ hỗ trợ tìm dấu vết. Người đọc vẫn cần đối chiếu với nguồn chính thống và bối cảnh cụ thể trước khi kết luận.

AI không làm cho thông tin trở nên kém tin cậy hơn nhưng nó làm cho việc tạo ra thông tin giả trở nên dễ dàng hơn bao giờ hết. Trong kỷ nguyên AI, phản xạ kiểm chứng chính là "lá chắn" quan trọng nhất giúp mỗi người bảo vệ bản thân và cộng đồng.

[CHƯƠNG V]

05

ỨNG XỬ VỚI THÔNG TIN TRONG ĐỜI SỐNG SỐ

Tin giả, tin sai sự thật hoặc thông tin gây hiểu nhầm không chỉ lan truyền từ người tạo ra, nó lan tiếp qua những lần người khác tin, chia sẻ, bình luận, nhấn thích, kể lại hoặc làm theo. Vì vậy, mỗi người không chỉ đứng ngoài quan sát dòng chảy thông tin. Mỗi người đều có thể làm tin đúng đi xa hơn hoặc làm tin sai dừng lại sớm hơn.

01. NGƯỜI TIẾP NHẬN: ĐỪNG ĐỂ CẢM XÚC QUYẾT ĐỊNH THAY MÌNH

Người tiếp nhận thông tin có quyền dừng lại trước khi tin hoặc hành động. Khi đọc một tin khiến bản thân thấy sợ, thương, giận hoặc háo hức, hãy tự hỏi:

[mình đang tin vì có căn cứ, hay vì tin này chạm đúng cảm xúc của mình?]

[mình biết nguồn gốc của tin này chưa?]

[nếu chia sẻ mà tin sai, ai có thể bị ảnh hưởng?]

Trong những tình huống liên quan đến sức khỏe, tài sản, an toàn, danh dự, dữ liệu cá nhân hoặc hành động khẩn cấp, chỉ cần một câu hỏi thêm đã có thể tránh được nhiều rủi ro không đáng có.

02. NGƯỜI CHIA SẺ TRONG NHÓM: KIỂM CHỨNG TRƯỚC KHI CHUYỂN TIẾP

Nhóm gia đình, nhóm cư dân, nhóm lớp, nhóm phụ huynh hoặc nhóm cơ quan là nơi thông tin lan rất nhanh. Người gửi tin thường có thiện chí. Nếu bị phản bác gay gắt, họ dễ cảm thấy mất mặt và cuộc trao đổi có thể biến thành tranh cãi. Vì vậy, cách nhắc kiểm chứng cũng cần nhẹ nhàng.

Thay vì nói "tin này sai rồi, đừng gửi nữa", có thể nói: "Tin này nghe đáng lo, mình thử tìm nguồn chính thức trước khi chia sẻ rộng nhé."

Thay vì phủ nhận ngay một lời kêu gọi giúp đỡ, có thể nói: "Việc hỗ trợ là rất đáng quý, mình kiểm tra tài khoản tiếp nhận trên kênh chính thức trước đã."

Cách nói như vậy giữ được thiện chí của người gửi, đồng thời tạo được điểm dừng cho cả nhóm.

03. QUẢN TRỊ VIÊN NHÓM: TẠO ĐIỂM DỪNG CHO THÔNG TIN CHƯA RÕ

Quản trị viên nhóm không chỉ là người xóa bài hoặc khóa bình luận. Trong nhiều trường hợp, họ là người giúp nhóm giữ bình tĩnh. Khi một tin khẩn lan trong nhóm, quản trị viên có thể ghim một thông báo ngắn: **"Nội dung này đang được kiểm tra, đề nghị mọi người chưa chuyển tiếp thêm và chưa làm theo khi chưa có nguồn xác nhận."**

Với các nhóm đông người, một vài quy tắc nhỏ có thể giúp giảm rủi ro:

Tin khẩn phải có nguồn rõ

Lời kêu gọi chuyển tiền phải có kênh xác nhận

Không đăng thông tin cá nhân hoặc lời buộc tội khi chưa đủ bối cảnh

Nội dung nghi lừa đảo hoặc giả mạo cần được báo cáo thay vì tranh luận kéo dài.

04. NGƯỜI CÓ ẢNH HƯỞNG: THẬN TRỌNG VỚI SỨC LAN TỎA

Người có ảnh hưởng không chỉ phát đi nội dung; họ còn phát đi cảm xúc và định hướng phản ứng của cộng đồng. Một nhận định vội, một tiêu đề giật gân, một lời kêu gọi phẫn nộ có thể khiến hàng nghìn người tin, chia sẻ hoặc công kích trước khi sự việc được làm rõ.

Với người có ảnh hưởng, trách nhiệm đầu tiên là không biến suy đoán thành khẳng định chắc chắn. Khi sự việc chưa rõ, nên dùng cách nói thận trọng: "thông tin này cần được kiểm chứng thêm", "hiện chưa có nguồn xác nhận", "không nên vội kết luận từ một đoạn video ngắn". Không nên dùng phẫn nộ cộng đồng để tăng tương tác khi sự việc chưa đủ bối cảnh.

Nếu đã chia sẻ sai, xóa bài thường chưa đủ. Người đã tiếp nhận thông tin sai có thể không biết bài đã bị xóa.

Vì vậy, người có ảnh hưởng cần đính chính công khai với mức độ tương xứng với phạm vi thông tin đã lan truyền. Đính chính không làm giảm uy tín; đính chính đúng cách cho thấy trách nhiệm.

05. TỔ CHỨC HOẶC CÁ NHÂN BỊ MẠO DANH, CÔNG KÍCH: PHẢN HỒI RÕ, CHỈ DẪN ĐÚNG

Khi bị mạo danh hoặc công kích bởi thông tin thiếu căn cứ, phản ứng đầu tiên thường là bức xúc. Nhưng phản hồi hiệu quả là cần giúp công chúng phân biệt rõ điều gì là thật, điều gì là giả, kênh nào là kênh chính thức và người dân nên làm gì ngay.

Một thông báo đính chính tốt không chỉ nói "đây là tin giả". Nó nên chỉ rõ tài khoản/website/văn bản nào là giả mạo, kênh chính thức nào cần theo dõi, người dân không nên bấm vào đường link nào, không nên chuyển tiền vào tài khoản nào, và có thể gửi thông tin nghi vấn đến đâu để được hỗ trợ.

Nếu có bằng chứng về nội dung giả mạo, nên lưu lại đường link, ảnh chụp màn hình, thời điểm phát hiện và phạm vi lan truyền để phục vụ việc báo cáo hoặc xử lý tiếp theo.

06. TỪ GIA ĐÌNH ĐẾN NHÀ TRƯỜNG: ĐƯA KIỂM CHỨNG VÀO THÓI QUEN HẰNG NGÀY

Khi những quy tắc nhỏ được nhắc đi nhắc lại, kiểm chứng không còn là việc xa lạ. Nó trở thành một cách chăm sóc nhau trong môi trường số: nhắc nhau chậm lại, giúp nhau kiểm nguồn, cùng nhau đính chính khi sai và không để lòng tốt bị lợi dụng.

Thói quen kiểm chứng không hình thành chỉ bằng một lần đọc cẩm nang. Nó cần được nhắc lại trong đời sống.

Gia đình có thể thống nhất rằng mọi yêu cầu chuyển tiền gấp đều phải xác nhận qua kênh thứ hai.

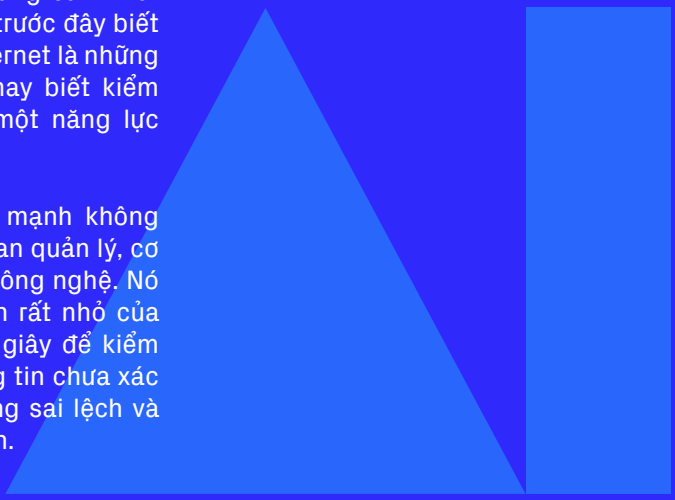
Lớp học có thể tập thói quen hỏi nguồn khi làm bài hoặc chia sẻ thông tin.

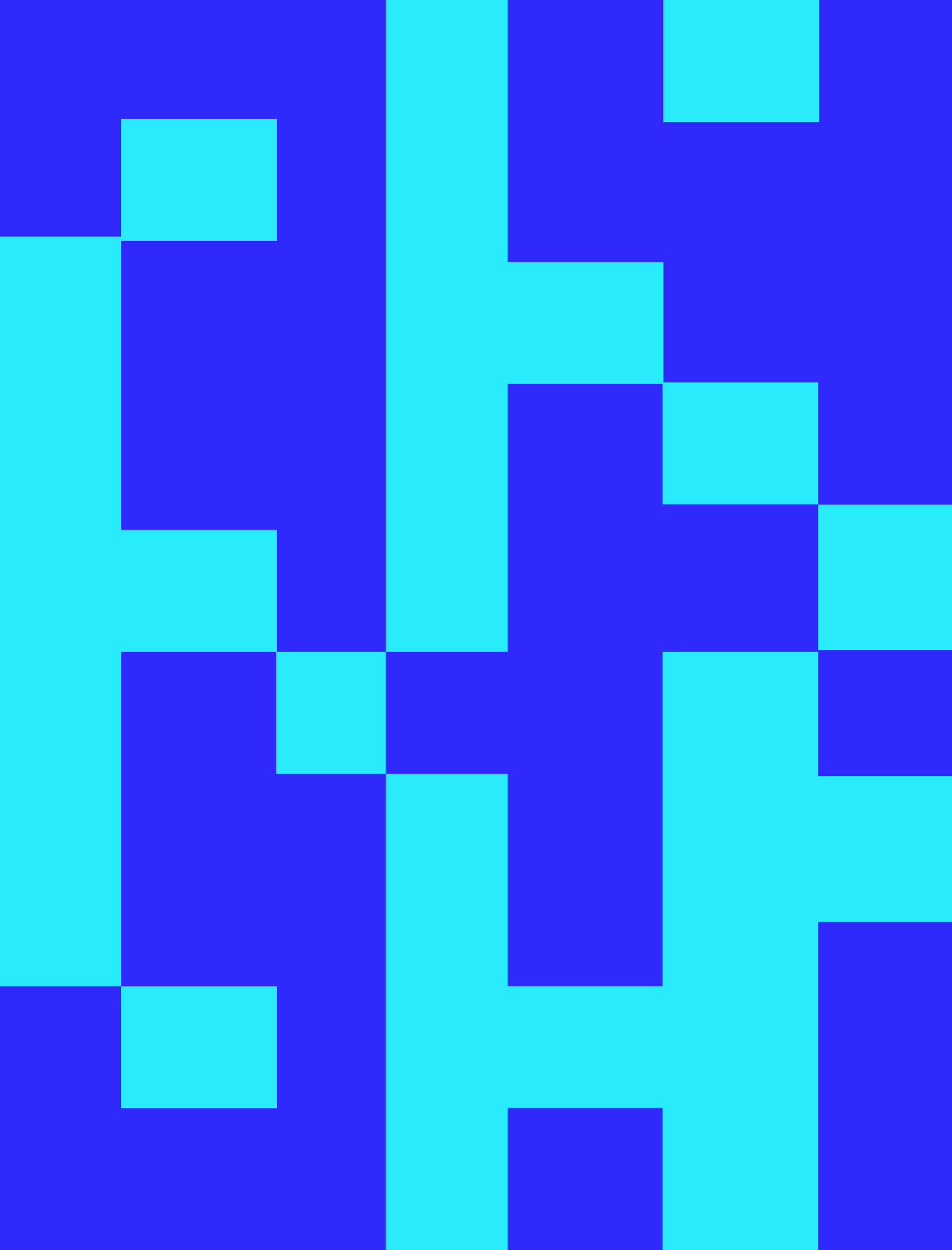
Nhóm cư dân có thể thống nhất không chuyển tiếp tin khẩn nếu không có nguồn rõ.

LỜI KẾT

Mỗi thời đại đều có những kỹ năng cần thiết để thích nghi và phát triển. Nếu trước đây biết đọc, biết viết và biết sử dụng Internet là những năng lực quan trọng thì ngày nay biết kiểm chứng thông tin đã trở thành một năng lực thiết yếu của công dân số.

Một môi trường thông tin lành mạnh không chỉ được xây dựng bởi các cơ quan quản lý, cơ quan báo chí hay các nền tảng công nghệ. Nó được tạo nên từ những lựa chọn rất nhỏ của mỗi chúng ta: dừng lại một vài giây để kiểm chứng, từ chối chia sẻ một thông tin chưa xác thực, chủ động báo cáo nội dung sai lệch và lan tỏa những thông tin đúng đắn.



Mỗi lần kiểm chứng trước khi chia sẻ không chỉ là bảo vệ chính mình mà còn góp phần bảo vệ gia đình, cộng đồng và xây dựng một không gian mạng an toàn, văn minh, đáng tin cậy hơn.

Hãy bắt đầu từ hôm nay
Kiểm trước tin sau