

Rủi ro và thách thức khi ứng dụng trí tuệ nhân tạo trong hoạt động xét xử ở Việt Nam

03/06/2026 15:27

(kiemsat.vn)

Nếu được áp dụng một cách thận trọng và hợp lý, AI có thể góp phần nâng cao chất lượng xét xử, hạn chế sai sót chủ quan, giảm tải công việc cho Thẩm phán, đồng thời thúc đẩy tính minh bạch, nhất quán và công bằng trong hệ thống tư pháp. Tuy nhiên, bên cạnh những lợi ích, việc triển khai AI cũng đặt ra hàng loạt thách thức về pháp lý, đạo đức, thể chế và nhân lực.

Việc ứng dụng trí tuệ nhân tạo trong hoạt động tư pháp là một xu thế tất yếu trong tiến trình chuyển đổi số quốc gia và cải cách nền công vụ hiện đại. Nếu được áp dụng một cách thận trọng và hợp lý, AI có thể góp phần nâng cao chất lượng xét xử, hạn chế sai sót chủ quan, giảm tải công việc cho Thẩm phán, đồng thời thúc đẩy tính minh bạch, nhất quán và công bằng trong hệ thống tư pháp. Tuy nhiên, bên cạnh những lợi ích, việc triển khai AI cũng đặt ra hàng loạt thách thức về pháp lý, đạo đức, thể chế và nhân lực.

Sự phát triển vượt bậc của trí tuệ nhân tạo (AI) đã mở ra nhiều triển vọng ứng dụng trong lĩnh vực tư pháp, nhất là khi các hệ thống pháp luật trên thế giới đang đứng trước yêu cầu chuyển đổi số và cải cách thể chế. AI được kỳ vọng góp phần hỗ trợ hoạt động xét xử, giảm tải công việc cho Thẩm phán, đồng thời nâng cao tính minh bạch và sự thống nhất trong áp dụng pháp luật. Tuy nhiên, việc tích hợp AI vào tư pháp không chỉ là một tiến bộ công nghệ, mà còn kéo theo những vấn đề phức tạp liên quan đến pháp lý, đạo đức và thể chế.

Đối với Việt Nam, vấn đề này càng có ý nghĩa khi hệ thống tư pháp đang trong quá trình cải cách theo hướng hiện đại hóa, chuyên nghiệp hóa và nâng cao chất lượng xét xử. Việc nghiên cứu và thử nghiệm các công nghệ số, trong đó có AI, đã được đặt ra như một xu thế tất yếu gắn với chiến lược chuyển đổi số quốc gia[1].

1. Nhận diện các rủi ro và thách thức trong ứng dụng AI tư pháp

- Nguy cơ bất bình đẳng trong tiếp cận công lý:

Một trong những rủi ro nghiêm trọng nhất của việc ứng dụng AI trong hoạt động xét xử là khả năng tái tạo và khuếch đại những định kiến xã hội vốn đã tồn tại trong xã hội[2]. AI chỉ có thể học từ dữ liệu mà con người cung cấp, trong khi dữ liệu lịch sử trong lĩnh vực pháp lý thường phản ánh những thiên kiến chủ quan, vô thức hoặc những bất cập xã hội kéo dài. Khi được huấn luyện trên những bộ

dữ liệu như vậy, AI không những không khắc phục được định kiến mà còn có thể “tự động hóa” và áp dụng định kiến đó một cách có hệ thống.

Trong bối cảnh Việt Nam, sự đa dạng về vùng miền, trình độ phát triển kinh tế - xã hội và khả năng tiếp cận dịch vụ pháp lý tạo ra những khác biệt đáng kể trong việc thực thi công lý. Người dân ở vùng sâu, vùng xa hoặc những nhóm yếu thế như dân tộc ít người và người nghèo thường gặp bất lợi ngay từ đầu khi tham gia tố tụng. Nếu các định kiến đối với những nhóm này được “mã hóa” vào dữ liệu huấn luyện, AI có thể vô tình tái tạo sự bất công, củng cố định kiến xã hội và khiến họ khó tiếp cận công lý hơn.

- Thiếu minh bạch và hạn chế khả năng giải trình:

Nguyên tắc tố tụng của Việt Nam nhấn mạnh tính minh bạch, công khai và giải trình trong mọi quyết định của Tòa án. Thẩm phán có nghĩa vụ lý giải các phán quyết dựa trên căn cứ pháp luật rõ ràng, có thể kiểm chứng và phản biện. Tuy nhiên, nhiều hệ thống AI hiện nay, đặc biệt là các mô hình học sâu (deep learning), hoạt động theo cơ chế “hộp đen”[3], nghĩa là quá trình ra quyết định quá phức tạp, khó truy vết và không dễ dàng giải thích[4]. Khi AI đưa ra một gợi ý hoặc kết luận mà không thể minh bạch hóa logic đằng sau, nguyên tắc giải trình sẽ bị thách thức nghiêm trọng[5].

Nếu không có cơ chế giải thích rõ ràng, nguy cơ xảy ra tình trạng người dân bị “xét xử bởi một thuật toán mà họ không hiểu và không thể phản biện” là hoàn toàn có thể. Điều này đi ngược lại nguyên tắc công bằng và quyền được xét xử minh bạch. Niềm tin của công chúng vào hệ thống tư pháp vốn phụ thuộc rất nhiều vào khả năng giải thích và tính hợp pháp của mỗi phán quyết. Việc mất đi yếu tố minh bạch có thể làm suy giảm sự chính danh của Tòa án trong mắt xã hội[6].

- Khoảng trống về trách nhiệm pháp lý:

Khi AI tham gia vào quá trình ra quyết định tư pháp, vấn đề đặt ra là ai sẽ chịu trách nhiệm nếu xảy ra sai sót. Truyền thống pháp lý Việt Nam quy định Thẩm phán là người chịu trách nhiệm cuối cùng cho quyết định của mình. Tuy nhiên, khi quyết định đó bị chi phối bởi một công cụ AI mà bản thân Thẩm phán không thể kiểm chứng toàn diện, ranh giới giữa trách nhiệm cá nhân và lỗi hệ thống trở nên mờ nhạt.

Pháp luật Việt Nam hiện chưa có quy định cụ thể về trách nhiệm pháp lý trong trường hợp sai sót do AI. Vậy trách nhiệm thuộc về nhà phát triển hệ thống, cơ quan tư pháp sử dụng công cụ, hay chính Thẩm phán áp dụng gợi ý từ AI? Ở một số khu vực như Liên minh Châu Âu, câu hỏi này được tiếp cận thông qua khái niệm trách nhiệm sản phẩm (product liability), trong đó nhà phát triển có thể bị truy cứu nếu sản phẩm AI gây ra thiệt hại[7]. Trong khi đó, Việt Nam vẫn đang thiếu một khung pháp lý tương tự, khiến cho việc xác định trách nhiệm khó khăn.

- Rủi ro mất an toàn dữ liệu và an ninh mạng:

Dữ liệu tư pháp luôn là loại dữ liệu đặc biệt, bao gồm thông tin cá nhân, hồ sơ vụ án và các tài liệu mật. Việc tích hợp AI vào hệ thống xét xử đồng nghĩa với việc xử lý khối lượng dữ liệu khổng lồ này. Các hệ thống GenAI được huấn luyện trên các bộ dữ liệu lớn, thường chứa dữ liệu cá nhân mà các nhà phát triển hoặc người dùng nhập vào[8]. Những dữ liệu này có thể bao gồm các thông tin cá nhân như chủng tộc, giới tính và tuổi tác, thậm chí là những thông tin mà người dùng cung cấp khi tương tác với GenAI qua các yêu cầu. Điều này đã tạo ra những mối lo ngại đáng kể về quyền riêng tư đối với cá nhân, ngành công nghiệp và Chính phủ. Trong bối cảnh hạ tầng công nghệ của ngành tư pháp Việt Nam còn hạn chế và chưa đồng bộ, nguy cơ rò rỉ hoặc bị tấn công mạng là rất lớn. Chỉ cần một lỗ hổng bảo mật nhỏ, hậu quả cũng có thể vô cùng nghiêm trọng, từ việc tiết lộ thông tin cá nhân của đương sự cho tới nguy cơ thao túng, bóp méo kết quả phân tích của AI, dẫn đến những phán quyết sai lệch.

- Ảnh hưởng đến năng lực nghề nghiệp của Thẩm phán:

AI được kỳ vọng sẽ giúp Thẩm phán tiết kiệm thời gian, giảm tải công việc và hạn chế sai sót. Tuy nhiên, nếu sự hỗ trợ này biến thành sự lệ thuộc, năng lực nghề nghiệp của Thẩm phán có thể bị suy giảm. Đặc biệt, đội ngũ Thẩm phán trẻ có nguy cơ bị ảnh hưởng đến khả năng phân tích độc lập, tư duy phản biện và sự linh hoạt trong xử lý những vụ việc phức tạp mà AI khó có thể nắm bắt.

Xét xử không chỉ là việc áp dụng pháp luật một cách máy móc, mà còn là quá trình đòi hỏi sự thấu cảm, tính nhân văn và sự linh hoạt để đưa ra các phán quyết phù hợp với bối cảnh xã hội cụ thể. Nếu quá phụ thuộc vào AI, Thẩm phán có thể bỏ qua những giá trị mềm này, biến xét xử thành một quy trình “khô cứng”. Hệ quả lâu dài là giảm khả năng phát triển án lệ, giảm tinh thần cải cách và làm yếu đi bản lĩnh nghề nghiệp của người “cầm cân nảy mực”.

- Hạn chế về hạ tầng, dữ liệu và nhân lực:

Triển khai AI trong ngành tư pháp chỉ có thể hiệu quả khi có một hệ thống dữ liệu chất lượng cao, đồng bộ và dễ truy cập. Thực tế tại Việt Nam, nhiều hồ sơ xét xử vẫn ở dạng giấy, lưu trữ phân tán tại các địa phương, chưa được số hóa toàn diện. Ngay cả khi dữ liệu được số hóa, tình trạng thiếu chuẩn hóa, không có cơ sở dữ liệu chung và liên thông giữa các cấp xét xử khiến AI khó phát huy hiệu quả.

Ngoài ra, nhân lực cũng là một thách thức lớn. Cán bộ tư pháp, kể cả Thẩm phán, Thư ký hay Kiểm sát viên, phần lớn chưa được đào tạo kỹ năng phân tích dữ liệu và vận hành hệ thống AI. Việc thiếu hụt đội ngũ chuyên gia liên ngành có hiểu biết sâu về cả công nghệ và pháp luật làm gia tăng rủi ro khi triển khai. Nếu không đầu tư nghiêm túc vào hạ tầng và nhân lực, việc ứng dụng AI rất dễ rơi vào tình trạng hình thức, thậm chí phản tác dụng.

- Nguy cơ xung đột với quyền con người và nguyên tắc pháp quyền:

Một trong những nguyên tắc cốt lõi của tố tụng hình sự là quyền được xét xử công bằng và nguyên tắc suy đoán vô tội. Tuy nhiên, việc sử dụng AI để “dự báo nguy cơ tái phạm” hoặc “chấm điểm bị cáo” có thể vô hình trung tạo ra định kiến ngay từ đầu, khiến nguyên tắc suy đoán vô tội bị suy yếu[9]. Người bị buộc tội có thể bị coi là “có khả năng phạm tội cao” chỉ dựa trên phân tích thuật toán, làm ảnh hưởng đến tâm lý Thẩm phán, Công tố viên hoặc thậm chí cả Luật sư bào chữa[10].

Hơn nữa, AI cần một lượng lớn dữ liệu cá nhân để hoạt động hiệu quả. Nếu không có cơ chế quản lý chặt chẽ, việc thu thập và xử lý dữ liệu này có thể xâm phạm nghiêm trọng quyền riêng tư của cá nhân.

2. Khoảng cách thể chế và thách thức đặc thù của Việt Nam

Tại Việt Nam, việc ứng dụng trí tuệ nhân tạo trong hoạt động của Tòa án nhân dân đang ngày càng được chú trọng trong bối cảnh chuyển đổi số quốc gia và yêu cầu cải cách tư pháp sâu rộng. Một số ứng dụng bước đầu, tiêu biểu là phần mềm Trợ lý ảo dành cho Thẩm phán triển khai từ năm 2022, đã cho thấy tiềm năng lớn của AI trong hoạt động tố tụng[11].

Tuy nhiên, để công nghệ AI thâm nhập sâu rộng vào hoạt động xét xử và phát huy được những lợi ích của nó trong hoạt động tố tụng[12], Việt Nam phải đối mặt với những thách thức sau:

Một trong những thách thức cốt lõi khi ứng dụng trí tuệ nhân tạo vào hoạt động tư pháp tại Việt Nam xuất phát từ sự khác biệt căn bản giữa hệ thống pháp luật trong nước và logic hoạt động của các mô hình AI pháp lý hiện nay. Hầu hết các hệ thống AI tiên tiến trong lĩnh vực tư pháp được thiết kế và huấn luyện dựa trên cơ sở dữ liệu, nguyên tắc và phương pháp lập luận của hệ thống Common law[13], nơi án lệ đóng vai trò là nguồn luật trung tâm và có giá trị ràng buộc pháp lý. Các mô hình này được tối ưu để phân tích, so sánh và dự đoán dựa trên các phán quyết đã có, đồng thời đề xuất lập luận pháp lý dựa trên tiền lệ[14]. Trong khi đó, Việt Nam cơ bản vận hành theo hệ thống Civil law[15], trong đó luật thành văn do cơ quan lập pháp ban hành mới là nguồn luật tối cao, còn án lệ chỉ giữ vai trò tham khảo. Điều này tạo ra một khoảng cách lớn giữa “nguyên liệu” mà AI thường xử lý và cách thức áp dụng luật tại Việt Nam.

Khoảng cách này dẫn đến một loạt hệ quả là việc “nhập khẩu” các công cụ AI được thiết kế cho hệ thống pháp luật Common law có thể gây ra sự lệch pha trong lập luận pháp lý; AI quen “tìm kiếm tiền lệ” để suy ra nguyên tắc. Sự khác biệt này khiến những gợi ý từ AI có nguy cơ thiếu tính phù hợp với cách tư duy pháp lý của Việt Nam, thậm chí làm Thẩm phán rơi vào tình trạng “lập luận vay mượn” không tương thích với hệ thống pháp luật thành văn.

Một khía cạnh khác là văn hóa và quy trình tố tụng. Tại Việt Nam, Tòa án vận hành trong một môi trường có nhiều đặc thù: Vai trò trung tâm của Thẩm phán, sự tham gia của Kiểm sát viên, vai trò của Luật sư và cách thức thu thập, đánh giá chứng cứ vốn gắn liền với quy định thành văn cụ thể. Các công cụ AI quốc tế, nếu không được điều chỉnh sẽ khó thích ứng với những đặc thù này. Ví dụ, một hệ thống được thiết kế để đưa ra khuyến nghị dựa trên án lệ có thể bỏ qua các yếu tố quan trọng như nghị định, thông tư hoặc văn bản hướng dẫn thi hành luật - vốn là cơ sở quan trọng trong xét xử tại Việt Nam. Điều này không chỉ làm giảm giá trị ứng dụng của AI mà còn tiềm ẩn nguy cơ gây “nhiều” hoạt động tố tụng, buộc Thẩm phán phải mất thêm thời gian kiểm chứng và hiệu chỉnh.

Khoảng cách thể chế còn thể hiện ở mức độ ổn định và tính dự báo của pháp luật. Hệ thống Common law có đặc điểm là phát triển dần dần thông qua án lệ, nhờ đó dữ liệu pháp lý thường có tính liên tục và được cập nhật theo từng vụ việc. Trong khi đó, hệ thống pháp luật Việt Nam thường xuyên thay đổi và cập nhật thông qua các văn bản pháp luật mới, khiến dữ liệu pháp lý không ổn định. Việc này gây khó khăn lớn cho AI, bởi hệ thống cần thường xuyên được “tái huấn luyện” để theo kịp sự thay đổi, nhưng lại thiếu nguồn dữ liệu chuẩn hóa và liên thông. Nếu không được cập nhật kịp thời, AI có thể đưa ra các gợi ý dựa trên văn bản đã hết hiệu lực, làm sai lệch phán quyết và đe dọa tính hợp pháp của quyết định tư pháp.

Một yếu tố khác không thể bỏ qua là khung thể chế và mức độ sẵn sàng pháp lý cho việc ứng dụng AI. Ở nhiều quốc gia theo Common law, AI được tích hợp vào tư pháp trong một bối cảnh pháp luật đã phát triển mạnh về tính minh bạch dữ liệu, quyền tiếp cận án lệ công khai và cơ chế kiểm soát độc lập. Trong khi đó, Việt Nam mới ở giai đoạn đầu của tiến trình số hóa tư pháp và bước đầu công khai bản án theo Nghị quyết số 03/2017/NQ-HĐTP ngày 16/3/2017 của Hội đồng Thẩm phán Tòa án nhân dân tối cao, dữ liệu còn phân tán và chưa có khung pháp lý về trách nhiệm AI trong hoạt động công vụ. Điều này dẫn đến sự bất cân xứng giữa công cụ AI và môi trường pháp lý tiếp nhận, làm gia tăng rủi ro.

Cuối cùng, sự khác biệt về bối cảnh xã hội và văn hóa pháp lý cũng làm cho việc ứng dụng AI trong xét xử tại Việt Nam trở nên đặc thù. Văn hóa pháp lý ở Việt Nam coi trọng vai trò của Nhà nước trong định hướng pháp luật và nhấn mạnh yếu tố nhân văn, đạo đức trong hoạt động xét xử. Nếu không có sự điều chỉnh phù hợp, nguy cơ là AI sẽ đưa ra những khuyến nghị thiếu tính nhân văn, làm suy giảm bản chất “cầm cân nảy mực” của nghề Thẩm phán.

3. Khuyến nghị chính sách cho việc ứng dụng AI trong tư pháp Việt Nam

Thách thức lớn nhất không chỉ đến từ yếu tố kỹ thuật hay dữ liệu, mà còn từ khoảng cách thể chế và sự khác biệt về phương pháp lập luận pháp lý so với bối cảnh phát triển của các công cụ AI hiện nay. Điều này đòi hỏi Việt Nam phải có một lộ trình chính sách chủ động, toàn diện và mang tính đặc thù, thay vì áp dụng nguyên xi các mô hình “nhập khẩu”. Trong quá trình đó, mục tiêu không phải là

thay thế con người bằng máy móc, mà là tận dụng sức mạnh công nghệ để hỗ trợ, nâng cao năng lực và phẩm chất nghề nghiệp của đội ngũ Thẩm phán - những người giữ vai trò trung tâm trong hoạt động xét xử.

Về định hướng chính sách, trước hết, cần xây dựng một khung pháp lý đầy đủ và linh hoạt nhằm điều chỉnh việc ứng dụng AI trong tố tụng và xét xử. Khung pháp lý này phải quy định rõ ràng trách nhiệm pháp lý khi xảy ra sai sót, thiết lập tiêu chuẩn về chất lượng dữ liệu, cơ chế giải trình cũng như các nguyên tắc bảo vệ quyền con người và quyền riêng tư. Việc luật hóa vai trò công cụ hỗ trợ (không thay thế) của AI trong quá trình ra quyết định tư pháp là cần thiết để bảo đảm tính chính danh và tính pháp quyền.

Tiếp theo, cần đẩy mạnh chuẩn hóa và số hóa dữ liệu tư pháp. AI chỉ có thể phát huy hiệu quả khi được vận hành trên cơ sở dữ liệu chất lượng cao, đầy đủ và đáng tin cậy. Điều này đòi hỏi một chiến lược tổng thể nhằm xây dựng cơ sở dữ liệu pháp lý thống nhất, liên thông giữa các cấp xét xử, đồng thời bảo đảm tính chính xác và khách quan thông qua cơ chế kiểm định chặt chẽ. Việc đầu tư cho dữ liệu không chỉ là một yêu cầu kỹ thuật, mà còn là điều kiện tiên quyết để ngăn ngừa nguy cơ thiên lệch thuật toán và sai lệch kết luận.

Song song với dữ liệu, việc đầu tư phát triển hạ tầng công nghệ và nhân lực chuyên sâu là yếu tố then chốt. Việt Nam cần hình thành đội ngũ chuyên gia liên ngành có khả năng kết nối giữa công nghệ AI và khoa học pháp lý, đồng thời tổ chức các chương trình đào tạo thường xuyên cho Thẩm phán, Kiểm sát viên và Thư ký Tòa án về kỹ năng sử dụng, giám sát và phản biện kết quả từ AI.

Một giải pháp quan trọng khác là triển khai mô hình thử nghiệm có kiểm soát (sandbox pháp lý[16]) trước khi áp dụng rộng rãi. Các ứng dụng AI cần được thí điểm trong một phạm vi hẹp, có giám sát chặt chẽ từ nhiều bên, để đánh giá hiệu quả, nhận diện rủi ro và điều chỉnh kịp thời. Mô hình này vừa bảo đảm tính an toàn, vừa tạo không gian linh hoạt cho đổi mới sáng tạo, tránh tình trạng áp dụng ồ ạt dẫn đến sai sót hệ thống.

Ngoài ra, việc bảo đảm tính minh bạch và cơ chế giám sát độc lập đối với thuật toán cũng có ý nghĩa then chốt. Các hệ thống AI được sử dụng trong xét xử phải có khả năng giải trình, đồng thời được công khai một phần cơ chế hoạt động để các bên liên quan có thể tham gia giám sát. Đây là điều kiện để tạo dựng niềm tin của công chúng và duy trì tính chính danh của hệ thống tư pháp trong kỷ nguyên số.

Cuối cùng, dù công nghệ có phát triển đến đâu, con người vẫn phải giữ vai trò trung tâm trong xét xử. Quyết định cuối cùng của một phiên tòa phải thuộc về Thẩm phán - những người có đạo đức nghề nghiệp, tư duy pháp lý độc lập và khả năng thấu cảm với hoàn cảnh xã hội cụ thể. Việc áp dụng AI chỉ thực sự có ý

nghĩa khi nó nâng cao năng lực của con người, mà không làm xói mòn bản lĩnh nghề nghiệp và giá trị nhân văn của người “cầm cân nảy mực”.p

Nguyễn Văn Quân

Tài liệu tham khảo:

1. Chien, C, et al. (forthcoming), How LLMs Can Help Address the Access to Justice Gap through the Courts, Loyola of Los Angeles Law Review. Xem: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4683309, truy cập ngày 18/7/2025.
2. Commission of Ontario, 2025, AI and Risk Assessment in Bail, Sentencing, and Recidivism, Law Commission of Ontario. Canada. Xem: <https://www.lco-cdo.org/wp-content/uploads/2025/04/LCO-AI-in-Criminal-Justice-Paper-3-AI-and-Risk-Assessment.pdf>, truy cập ngày 15/8/2025.
3. Đào Trí Úc, “Basic Information of Legal Research - A Case Study of Vietnam”, Project of Doing Legal Research in Asian Countries: China, India, Malaysia, Philippines, Thailand and Vietnam, Conducted by the Institute of Developing Economies (IDE-JETRO), Japan, 2003.
4. Drew Simshaw, Access to AI Justice: Avoiding an Inequitable Two-Tiered System of Legal Services, 24 Yale J.L. & Tech 150 (2022).
5. Gupta, M, et al. (2023), From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy, Xem: <https://doi.org/10.1109/ACCESS.2023.3300381>.
6. Jennifer Cobbe, Legal Singularity and the Reflexivity of Law, 1st edition, Is Law Computable?, Critical Perspectives on Law and Artificial Intelligence, pages 107–134. Hart Publishing, 2020, tr.107-134.
7. Josef Valvoda, Alec Thompson, Ryan Cotterell, Simone Teufel; The Ethics of Automating Legal Actors, Transactions of the Association for Computational Linguistics 2024; 12 700–720.
8. Lương Văn Tuấn, Xây dựng Tòa án điện tử trong xu thế chuyển đổi số quốc gia, Tạp chí Nghề Luật, số 5/2024, tr.3-8.
9. Nguyễn Trí Tuệ, Ứng dụng công nghệ thông tin trong nâng cao hiệu quả hoạt động tố tụng tại Tòa án, Tạp chí Luật học, số 4/2019, tr.39-50.
10. Nguyễn Văn Quân, Vai trò của trí tuệ nhân tạo trong việc nâng cao chất lượng xét xử, Tạp chí Kiểm sát số 11/2024, tr.61-64.

11. Sachoulidou, A. Going beyond the “common suspects”: to be presumed innocent in the era of algorithms, big data and artificial intelligence. *Artif Intell Law* (2023), Xem: <https://link.springer.com/article/10.1007/s10506-023-09347-w>, truy cập ngày 18/7/2025.

12. Socol de la Osa DU, Remolina N, Artificial intelligence at the bench: Legal and ethical challenges of informing - or misinforming - judicial decision-making through generative AI, *Data & Policy*. 2024;6:e59. doi:10.1017/dap.2024.53.

13. von Eschenbach, W.J. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philos. Technol.* 34, 1607–1622 (2021).

14. Vũ Thị Lan Anh, Mô hình khung thể chế thử nghiệm có kiểm soát (sandbox) ở Việt Nam trong bối cảnh hiện nay, *Tạp chí Luật học*, số 4 (299), 2025.

[1] Lương Văn Tuấn, Xây dựng Tòa án điện tử trong xu thế chuyển đổi số quốc gia, *Tạp chí Nghề Luật*, số 5/2024, tr.3-8.

[2] Drew Simshaw, Access to AI Justice: Avoiding an Inequitable Two-Tiered System of Legal Services, 24 *Yale J.L. & Tech* 150 (2022).

[3] Thuật ngữ dùng để chỉ các hệ thống trí tuệ nhân tạo mà quá trình ra quyết định hoặc hoạt động bên trong không rõ ràng, không dễ giải thích hoặc không thể hiểu được đối với con người, ngay cả với những người phát triển hệ thống đó. Ví dụ: Các mô hình học sâu như mạng nơ-ron nhân tạo với nhiều lớp, xử lý hàng triệu tham số và việc hiểu chi tiết cách mỗi tham số ảnh hưởng đến kết quả là rất khó.

[4] Chien, C, et al. (forthcoming) How LLMs Can Help Address the Access to Justice Gap through the Courts. *Loyola of Los Angeles Law Review*. Xem: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4683309, truy cập ngày 18/7/2025.

[5] Von Eschenbach, W.J. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philos. Technol.* 34, 1607–1622 (2021).

[6] Socol de la Osa DU, Remolina N, Artificial intelligence at the bench: Legal and ethical challenges of informing - or misinforming - judicial decision-making through generative AI, *Data & Policy*. 2024;6:e59. doi:10.1017/dap.2024.53.

[7] Socol de la Osa DU, Remolina N, *tlđđ* (footnote 6).

[8] Gupta, M, et al. (2023) From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE* Xem: <https://doi.org/10.1109/ACCESS.2023.3300381>.

[9] Sachoulidou, A. Going beyond the “common suspects”: To be presumed innocent in the era of algorithms, big data and artificial intelligence. *Artif Intell Law* (2023). Xem: <https://link.springer.com/article/10.1007/s10506-023-09347-w>, truy cập ngày 18/7/2025.

[10] Commission of Ontario, 2025, AI and Risk Assessment in Bail, Sentencing, and Recidivism, Law Commission of Ontario. Canada. Xem: <https://www.lco-cdo.org/wp-content/uploads/2025/04/LCO-AI-in-Criminal-Justice-Paper-3-AI-and-Risk-Assessment.pdf>, truy cập ngày 15/8/2025.

[11] Xem: Nguyễn Trí Tuệ, Ứng dụng công nghệ thông tin trong nâng cao hiệu quả hoạt động tố tụng tại Tòa án, *Tạp chí Luật học*, số 4/2019, tr.39-50; Lê Thuỳ Khanh, Mô hình Tòa án điện tử một số nước trên thế giới và kinh nghiệm cho Việt Nam, *Tạp chí Tòa án nhân dân*, số 20, tháng 10/2020, tr.36-40, 48; Vũ Thúy Hòa, Bàn về việc nâng cao hiệu quả của việc xây dựng và triển khai Tòa án điện tử ở Việt Nam, *Tạp chí Tòa án nhân dân*, số 11, tháng 6/2023, tr.15-26; Nguyễn Hoà Bình, Xây dựng Tòa án điện tử - một nhiệm vụ quan trọng của chiến lược cải cách tư pháp, *Tạp chí Tòa án nhân dân*, số 2, tháng 1/2022.

[12] Nguyễn Văn Quân, Vai trò của trí tuệ nhân tạo trong việc nâng cao chất lượng xét xử, *Tạp chí Kiểm sát* số 11/2024, tr.61-64.

[13] Jennifer Cobbe, *Legal Singularity and the Reflexivity of Law*, 1st edition, *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*, pages 107–134. Hart Publishing, 2020, tr.107-134.

[14] Josef Valvoda, Alec Thompson, Ryan Cotterell, Simone Teufel; *The Ethics of Automating Legal Actors*. *Transactions of the Association for Computational Linguistics* 2024; 12 700 - 720.

[15] Đào Trí Úc, “Basic Information of Legal Research - A Case Study of Vietnam”, *Project of Doing Legal Research in Asian Countries: China, India, Malaysia, Philippines, Thailand and Vietnam*, Conducted by the Institute of Developing Economies (IDE-JETRO), Japan, 2003, tr.206.

[16] Vũ Thị Lan Anh, Mô hình khung thể chế thử nghiệm có kiểm soát (sandbox) ở Việt Nam trong bối cảnh hiện nay, *Tạp chí Luật học*, số 4 (299), 2025, tr.3-7.