

Những thách thức triển khai AI Agents

Việc triển khai AI Agents trong doanh nghiệp và các tổ chức mở ra nhiều cơ hội đổi mới, nhưng cũng đi kèm với không ít thách thức phức tạp. Về mặt kỹ thuật, các AI Agents có thể gặp rủi ro như vòng lặp phản hồi vô hạn, thiếu khả năng giải thích quyết định, và đòi hỏi tài nguyên tính toán lớn. Hạ tầng công nghệ như mô hình nền tảng, phần cứng, điện toán đám mây và khả năng tích hợp phần mềm cũng có thể trở thành điểm nghẽn. Trong quá trình triển khai, doanh nghiệp cần đối mặt với các vấn đề như định nghĩa quy trình, bảo mật dữ liệu, giám sát con người và đảm bảo đạo đức AI. Bên cạnh đó, thị trường AI đang bị chi phối bởi các nhà cung cấp lớn, thỏa thuận độc quyền và rào cản từ các nền tảng di động. Cuối cùng, vấn đề quản trị như khung pháp lý, tiêu chuẩn, cơ chế minh bạch và hợp tác liên ngành cũng là yếu tố sống còn để đảm bảo AI phát triển bền vững và có trách nhiệm.

I. Rủi ro và hạn chế kỹ thuật của AI Agents

1. Sự hợp tác có nghĩa là nhiều rủi ro hơn

Mô tả: Khi nhiều AI agents cùng tương tác hoặc cộng tác để hoàn thành một nhiệm vụ, rủi ro hệ thống trở nên phức tạp hơn. Nếu một tác nhân có hành vi sai lệch, nó có thể ảnh hưởng dây chuyền đến các tác nhân khác, gây ra hiệu ứng domino.

Ví dụ: Trong một hệ thống AI hỗ trợ vận hành chuỗi cung ứng, một AI phụ trách dự đoán nhu cầu có thể đưa ra dự báo sai, khiến AI điều phối kho bãi nhập hàng dư thừa, và AI phụ trách vận chuyển phải xử lý khối lượng vượt mức — dẫn đến tắc nghẽn toàn hệ thống.

2. Vòng phản hồi vô hạn

Mô tả: Khi AI agent học từ phản hồi của chính mình hoặc từ tác nhân khác mà không có cơ chế giới hạn, có thể tạo ra vòng lặp phản hồi vô tận. Điều này dẫn đến hành vi lệch lạc, dữ liệu huấn luyện sai lệch hoặc quyết định ngày càng sai.

Ví dụ: Một chatbot AI trả lời khách hàng và đồng thời học từ phản hồi... của chính nó (vì bị nhầm là từ người dùng). Sau một thời gian, nó bắt đầu lặp lại câu trả lời không phù hợp và mất kiểm soát nội dung.

3. Tài nguyên tính toán

Mô tả: Các AI agents, đặc biệt là những agent phức tạp dùng mô hình ngôn ngữ lớn (LLMs), đòi hỏi sức mạnh tính toán rất cao. Điều này có thể làm tăng chi phí hạ tầng, chậm hệ thống hoặc gây quá tải.

Ví dụ: Một công ty triển khai AI trợ lý nội bộ sử dụng GPT để tổng hợp báo cáo. Khi hàng trăm nhân viên truy cập đồng thời, hệ thống trở nên chậm và tiêu tốn quá nhiều GPU trên cloud, gây tốn kém tài chính.

4. Thiếu khả năng giải thích

Mô tả: Nhiều AI agents hoạt động như hộp đen (black box), đưa ra quyết định mà con người khó hiểu được logic phía sau. Điều này gây khó khăn trong việc đánh giá, kiểm tra và tin tưởng AI.

Ví dụ: Một AI agent được giao nhiệm vụ từ chối hoặc duyệt đơn vay vốn, nhưng không thể giải thích lý do cụ thể tại sao đơn A được duyệt còn đơn B thì không, dù hồ sơ tương tự nhau. Điều này dễ dẫn đến nghi ngờ và khiếu nại.

II. Mô hình nền tảng

Mô tả: AI Agents thường dựa trên các mô hình ngôn ngữ lớn hoặc mô hình thị giác sâu đã được huấn luyện trước. Tuy nhiên, việc phụ thuộc vào các mô hình nền này có thể gây ra các vấn đề về chi phí, quyền truy cập, và độ phù hợp với từng trường hợp cụ thể. Ngoài ra, những mô hình này cũng có thể mang theo những thiên kiến và hạn chế từ dữ liệu gốc.

Ví dụ: Một công ty tài chính sử dụng GPT để hỗ trợ phân tích báo cáo tài chính. Tuy nhiên, mô hình nền tảng không hiểu rõ các khái niệm chuyên sâu trong ngành tài chính Việt Nam, dẫn đến phân tích sai hoặc thiếu thông tin quan trọng.

2. Phần cứng

Mô tả: Các tác nhân AI cần phần cứng mạnh mẽ, như GPU, TPU, RAM lớn hoặc thiết bị cảm biến chuyên biệt (trong robotics). Thiếu phần cứng phù hợp có thể làm chậm hệ thống, giảm hiệu suất, hoặc không thể chạy tác vụ thời gian thực.

Ví dụ: Một công ty bán lẻ triển khai robot AI để kiểm kê hàng hóa trong kho, nhưng robot không có đủ cảm biến hoặc camera chất lượng cao để đọc mã vạch trong điều kiện ánh sáng yếu — dẫn đến sai sót và tốn thời gian kiểm tra lại thủ công.

3. Hạ tầng đám mây

Mô tả: Hầu hết AI Agents hiện nay vận hành trên nền tảng điện toán đám mây. Tuy nhiên, hiệu suất của AI phụ thuộc nhiều vào khả năng mở rộng, bảo mật, và độ ổn định của cloud provider. Bất kỳ sự cố nào về mạng, bảo mật, hay tài nguyên hạn chế cũng ảnh hưởng đến toàn hệ thống.

Ví dụ: Một AI chăm sóc khách hàng đang vận hành qua dịch vụ đám mây bị gián đoạn do sự cố từ nhà cung cấp (ví dụ AWS bị downtime). Kết quả là hệ thống phản hồi khách hàng bị tê liệt trong nhiều giờ, ảnh hưởng đến trải nghiệm và uy tín thương hiệu.

4. Tích hợp phần mềm

Mô tả: Để AI agents hoạt động hiệu quả, chúng cần tích hợp liền mạch với các hệ thống hiện tại như CRM, ERP, HRM, v.v. Tuy nhiên, sự khác biệt về kiến trúc hệ thống, giao thức truyền thông hoặc độ tương thích có thể tạo ra rào cản lớn cho việc tích hợp.

Ví dụ: Một công ty muốn tích hợp AI để tự động tạo và gửi hóa đơn thông qua hệ thống kế toán cũ. Tuy nhiên, hệ thống hiện tại không hỗ trợ API hoặc không tương thích với mô hình AI hiện có, buộc phải xây dựng cầu nối trung gian tốn kém và mất thời gian.

III. Các vấn đề thực hiện

Triển khai AI Agents trong môi trường doanh nghiệp không chỉ đơn thuần là việc đưa một công nghệ mới vào hệ thống hiện có — đó là một quá trình phức tạp đòi hỏi sự chuẩn bị kỹ lưỡng về quy trình, hạ tầng và con người. Những thách thức này bắt đầu từ việc **xác định rõ quy trình** mà AI sẽ tham gia, cho đến việc **tích hợp** AI với các hệ thống hiện hành. **Bảo mật dữ liệu** là mối quan tâm hàng đầu, đặc biệt khi AI xử lý thông tin nhạy cảm. Đồng thời, AI cần có **sự giám sát của con người** để đảm bảo hoạt động đúng định hướng và phù hợp với giá trị đạo đức. Ngoài ra, việc duy trì tính **công bằng, không dự trữ dữ liệu** và đảm bảo **khả năng mở rộng, bảo trì lâu dài** cũng là những yếu tố sống còn cho sự thành công của AI trong thực tế.

1. Định nghĩa quy trình

Mô tả: Trước khi triển khai AI agent, doanh nghiệp cần xác định rõ quy trình nào sẽ được tự động hóa hoặc hỗ trợ. Thiếu định nghĩa quy trình rõ ràng sẽ khiến AI hoạt động không hiệu quả, chồng chéo với công việc con người hoặc gây gián đoạn vận hành.

Ví dụ: Một công ty muốn dùng AI để xử lý yêu cầu hỗ trợ khách hàng, nhưng không xác định rõ khi nào AI nên chuyển tiếp cho nhân viên thật. Hậu quả là AI giữ khách hàng quá lâu hoặc chuyển không đúng lúc, gây bất mãn.

2. Tích hợp

Mô tả: AI agents cần được tích hợp liền mạch vào hệ thống hiện có (CRM, ERP, website, v.v.). Quá trình này thường gặp khó khăn do không tương thích, thiếu API, hoặc kiến trúc phần mềm cũ.

Ví dụ: Một bệnh viện triển khai AI hỗ trợ đọc kết quả chụp X-quang, nhưng hệ thống AI không kết nối được với phần mềm lưu trữ hồ sơ bệnh án (EMR), khiến nhân viên phải nhập tay kết quả → chậm trễ và sai sót.

3. Bảo mật dữ liệu

Mô tả: AI agents cần truy cập và xử lý dữ liệu nhạy cảm như thông tin cá nhân, tài chính, y tế,... Nếu không được bảo vệ tốt, đây sẽ là lỗ hổng nghiêm trọng về bảo mật.

Ví dụ: Một AI hỗ trợ kế toán nội bộ bị khai thác qua API không được bảo mật, khiến dữ liệu lương và hợp đồng nhân sự bị rò rỉ ra bên ngoài.

4. Giám sát của con người

Mô tả: AI cần được giám sát bởi con người để đảm bảo hoạt động đúng, tránh lỗi hệ thống hoặc đưa ra quyết định không phù hợp. Thiếu giám sát có thể gây hậu quả nghiêm trọng, nhất là trong các ngành nhạy cảm.

Ví dụ: Một AI phỏng vấn tuyển dụng tự động loại bỏ nhiều ứng viên chỉ vì giọng nói địa phương, nhưng không có người kiểm tra lại các tiêu chí đánh giá → dẫn đến mất ứng viên chất lượng và phản ứng tiêu cực.

5. Mối quan tâm về đạo đức và thiên vị

Mô tả: AI có thể vô tình tái tạo hoặc khuếch đại các định kiến từ dữ liệu huấn luyện. Nếu không được kiểm soát, AI có thể đưa ra các quyết định phân biệt đối xử hoặc phi đạo đức.

Ví dụ: Một AI phân tích hồ sơ vay vốn học tập lại có xu hướng ưu tiên người từ khu vực đô thị hơn nông thôn, vì mô hình học từ dữ liệu lịch sử thiếu công bằng → gây bất bình đẳng xã hội.

6. Khả năng mở rộng và bảo trì

Mô tả: AI agent ban đầu có thể chạy tốt, nhưng khi số lượng người dùng hoặc dữ liệu tăng lên, hệ thống có thể bị quá tải nếu không được thiết kế để mở rộng. Đồng thời, AI cần được cập nhật và bảo trì định kỳ.

Ví dụ: Một cửa hàng online dùng AI để gợi ý sản phẩm. Trong mùa sale, lượng truy cập tăng gấp 10 lần khiến AI đưa ra đề xuất sai hoặc đung hệ thống → làm mất doanh thu và trải nghiệm khách hàng.

III. Các nhà cung cấp chi phối thị trường

Thị trường hiện tại đang bị chi phối bởi một số rào cản lớn có thể cản trở khả năng phát triển và triển khai của các doanh nghiệp. Những thách thức này bao gồm sự thống trị của các nhà cung cấp lớn (dominant providers), các thỏa thuận độc quyền (exclusivity agreements), và sự kiểm soát chặt chẽ từ các nền tảng di động như hệ điều hành và chợ ứng dụng (mobile OS and app stores). Những yếu tố này không chỉ làm hạn chế sự lựa chọn công nghệ, mà còn ảnh hưởng đến khả năng tiếp cận, phân phối và sáng tạo trong không gian AI. Do đó, các doanh nghiệp muốn ứng dụng AI Agents cần hiểu rõ môi trường thị trường để xây dựng chiến lược thích nghi phù hợp và giảm thiểu rủi ro phụ thuộc.

1. Các nhà cung cấp chi phối thị trường

Mô tả: Thị trường AI hiện đang bị thống trị bởi một số ít công ty lớn như OpenAI, Google, Microsoft, Anthropic... Họ kiểm soát các mô hình AI nền tảng, hạ tầng tính toán và API, khiến doanh nghiệp nhỏ phụ thuộc và thiếu lựa chọn.

Ví dụ: Một công ty khởi nghiệp muốn xây dựng AI agent nội bộ nhưng buộc phải phụ thuộc vào GPT của OpenAI vì không đủ nguồn lực huấn luyện mô hình riêng. Khi OpenAI tăng giá API hoặc thay đổi chính sách, công ty này không có phương án thay thế.

2. Thỏa thuận độc quyền

Mô tả: Một số công ty công nghệ lớn ký thỏa thuận độc quyền với các nhà cung cấp AI, giới hạn quyền truy cập hoặc tích hợp vào nền tảng của đối thủ, khiến thị trường bị khóa và cản trở đổi mới.

Ví dụ: Microsoft ký thỏa thuận độc quyền tích hợp GPT vào các sản phẩm Microsoft 365. Một đối thủ trong lĩnh vực phần mềm văn phòng không thể tích hợp cùng mức độ hoặc chức năng tương đương, dẫn đến mất lợi thế cạnh tranh.

3. Hệ điều hành và cửa hàng ứng dụng di động

Mô tả: Các nền tảng như iOS (Apple) và Android (Google) kiểm soát chặt chẽ hệ sinh thái ứng dụng, bao gồm quyền truy cập vào phần cứng, dữ liệu người dùng, và cả phân phối ứng dụng. Việc đưa AI agent lên điện thoại gặp rào cản chính sách, giới hạn kỹ thuật hoặc chia sẻ doanh thu.

Ví dụ: Một startup phát triển AI agent tương tác bằng giọng nói cho thiết bị di động, nhưng Apple không cho phép quyền truy cập toàn phần vào micro hoặc Siri. Ngoài ra, ứng dụng phải chia 30% doanh thu cho Apple nếu bán qua App Store.

IV. Quản trị AI Agents

Quản trị(Governance) trong lĩnh vực AI đề cập đến việc xây dựng các cơ chế, chính sách, luật pháp và quy trình nhằm đảm bảo rằng các hệ thống AI – bao gồm cả AI Agents – được phát triển, triển khai và vận hành một cách **an toàn, minh bạch, có trách nhiệm và tuân thủ đạo đức**.

Khi AI Agents ngày càng trở nên tự chủ và mạnh mẽ, việc quản trị chúng không còn là một lựa chọn, mà là điều bắt buộc để ngăn ngừa rủi ro và bảo vệ lợi ích xã hội, doanh nghiệp và cá nhân.

1. Quy định pháp lý

Mô tả: Các luật lệ và quy định từ chính phủ hoặc cơ quan quốc tế nhằm kiểm soát cách AI được thiết kế, triển khai và sử dụng.

Ví dụ:

- Liên minh Châu Âu ban hành **AI Act**, yêu cầu các hệ thống AI có mức rủi ro cao phải được kiểm tra nghiêm ngặt về an toàn, độ tin cậy và quyền riêng tư.
- Ở Việt Nam, nếu AI Agent xử lý dữ liệu cá nhân nhạy cảm mà không tuân thủ **Luật An ninh mạng** hoặc **Luật Bảo vệ dữ liệu cá nhân (dự thảo)**, doanh nghiệp có thể bị xử phạt.

2. Tiêu chuẩn kỹ thuật

Mô tả: Các hướng dẫn và tiêu chuẩn chung để thiết kế, phát triển và đánh giá AI một cách nhất quán, minh bạch và có thể kiểm chứng.

Ví dụ:

- ISO/IEC 42001** là tiêu chuẩn quốc tế mới cho hệ thống quản lý AI.
- Doanh nghiệp xây dựng AI Agent theo chuẩn **IEEE P7001** để đảm bảo tính minh bạch trong thuật toán ra quyết định.

3. Cơ chế trách nhiệm

Mô tả: Xác định rõ **ai chịu trách nhiệm khi AI Agent gây ra sự cố** hoặc hành vi sai lệch. Điều này giúp doanh nghiệp không "đổ lỗi cho AI" mà phải gắn trách nhiệm với con người hoặc tổ chức cụ thể.

Ví dụ:

Nếu một AI Agent từ chối đơn bảo hiểm sai do lỗi thuật toán, công ty bảo hiểm cần có bộ phận chịu trách nhiệm xử lý khiếu nại, sửa lỗi hệ thống, và bồi thường nếu cần.

4. Yêu cầu minh bạch

Mô tả: Các AI Agent cần minh bạch trong hoạt động của mình: chúng đưa ra quyết định như thế nào, sử dụng dữ liệu gì, và người dùng có thể hiểu (hoặc kiểm tra) quá trình đó.

Ví dụ:

Một AI Agent gợi ý sản phẩm trên sàn thương mại điện tử cần nêu rõ lý do: "Sản phẩm này được đề xuất vì bạn đã mua các sản phẩm tương tự trong 30 ngày qua".

5. Hợp tác liên ngành

Mô tả: Các tổ chức, doanh nghiệp, cơ quan chính phủ và giới học thuật cần phối hợp cùng nhau để đảm bảo AI phát triển một cách cân bằng, bền vững, và toàn diện.

Ví dụ:

- Một AI Agent hỗ trợ chẩn đoán y tế cần được xây dựng với sự hợp tác giữa công ty công nghệ, bệnh viện, bộ y tế và các tổ chức đạo đức để đảm bảo tính chính xác, an toàn và nhân đạo.
- Trong lĩnh vực tài chính, ngân hàng cần hợp tác với các chuyên gia pháp lý và công nghệ để triển khai AI tuân thủ các quy định chống rửa tiền (AML) và biết khách hàng (KYC).

Tác giả: Đỗ Ngọc Tú
Công Ty Phần Mềm VHTSoft

Phiên bản #3

Được tạo 9 tháng 4 2025 03:11:14 bởi Đỗ Ngọc Tú

Được cập nhật 14 tháng 4 2025 03:38:52 bởi Đỗ Ngọc Tú