

University of Science - VNU-HCM

Faculty of Information Technology



Computational Thinking

BÀI BÁO CÁO NGẮN:

Responsible Artificial Intelligence

University of Science - VNU-HCM

Faculty of Information Technology



Computational Thinking

BÀI BÁO CÁO NGẮN:

Responsible Artificial Intelligence

No.	Full Name	Student ID
1	Nguyễn Kiều Anh Quân	24127512

Mục lục

1	Giới thiệu	2
2	Responsible AI là gì?	2
3	Các nguyên tắc cốt lõi	2
4	Lợi ích của Responsible AI	3
5	Thách thức trong triển khai	3
6	Gợi ý giải pháp	3
7	Liên hệ thực tiễn tại Việt Nam	3
8	Kết luận	3
9	Tài liệu tham khảo	4

Ghi chú sử dụng AI

- **DeepSeek.** DeepSeek, DeepSeek (), chat.deepseek.com, truy cập 8:30 14/10/2025, prompt: “chuyển nội dung sau qua ngôn ngữ latex cho overleaf” + toàn bộ nội dung báo cáo Responsible AI, dùng để hỗ trợ chuyển đổi định dạng văn bản sang LaTeX, AI tạo mã nguồn LaTeX hoàn chỉnh, sinh viên tự kiểm tra và điều chỉnh các lỗi định dạng, bổ sung gói lệnh và tùy chỉnh bô cục theo yêu cầu.
- **DeepSeek.** DeepSeek, DeepSeek (), chat.deepseek.com, truy cập 9 : 10 14/10/2025, prompt: “cho cái tài liệu tham khảo”, dùng để bổ sung phần tài liệu tham khảo học thuật, AI cung cấp danh sách tài liệu tham khảo mẫu, sinh viên tự lựa chọn và chỉnh sửa các trích dẫn phù hợp với nội dung báo cáo, kiểm tra tính chính xác của các nguồn tham khảo.

1 Giới thiệu

Trí tuệ nhân tạo (AI) đang thay đổi mạnh mẽ cách con người sống, làm việc và ra quyết định. Tuy nhiên, cùng với tiềm năng lớn, AI cũng mang theo nhiều rủi ro về thiên kiến, quyền riêng tư và đạo đức. Vì vậy, Responsible AI (AI có trách nhiệm) ra đời như một tiêu chuẩn cần thiết, nhằm đảm bảo AI được phát triển minh bạch, công bằng và có lợi cho xã hội.

Theo báo cáo từ Market.us, thị trường Responsible AI toàn cầu dự kiến đạt 47,1 tỷ USD vào năm 2034, từ mức 910 triệu USD năm 2024 – cho thấy nhu cầu cấp thiết về đạo đức trong công nghệ.

2 Responsible AI là gì?

Responsible AI là phương pháp phát triển và triển khai AI dựa trên các nguyên tắc đạo đức, pháp lý và giá trị xã hội. Mục tiêu không chỉ là hiệu quả kỹ thuật, mà còn là giảm rủi ro, tăng niềm tin và đảm bảo lợi ích cho mọi nhóm người.

3 Các nguyên tắc cốt lõi

- **Minh bạch:** AI phải giải thích được dữ liệu và cách ra quyết định.
- **Trách nhiệm giải trình:** Tổ chức phát triển AI phải chịu trách nhiệm về tác động xã hội.
- **Công bằng & không thiên vị:** Ngăn chặn AI phân biệt đối xử do dữ liệu lệch.
- **Bảo mật & quyền riêng tư:** Tuân thủ các quy định như GDPR, bảo vệ dữ liệu cá nhân.
- **Tăng khả năng giải thích (Explainability):** Giúp người dùng hiểu và tin tưởng AI.

4 Lợi ích của Responsible AI

- Bảo vệ người dùng khỏi quyết định sai lệch hoặc thiên vị.
- Tăng tính minh bạch và tin tưởng vào hệ thống AI.
- Hạn chế hậu quả tiêu cực từ tự động hóa.
- Tạo lợi thế cạnh tranh cho doanh nghiệp trong môi trường tuân thủ.

5 Thách thức trong triển khai

- Xung đột giữa sáng tạo và đạo đức.
- Đội ngũ phát triển thiếu đa dạng, dễ bỏ sót thiên kiến xã hội.
- AI phát triển quá nhanh, luật pháp khó theo kịp.
- Khó đo lường tính đạo đức và hiệu quả xã hội của AI.
- Hạn chế về nguồn lực, đặc biệt với doanh nghiệp nhỏ.

6 Gợi ý giải pháp

- **Doanh nghiệp:** Tích hợp đạo đức vào quy trình phát triển sản phẩm, đào tạo nhân sự.
- **Nhà nước:** Xây dựng khung pháp lý AI rõ ràng, hỗ trợ nghiên cứu và phổ biến đạo đức AI.
- **Người làm AI:** Tự nâng cao nhận thức, sử dụng công cụ kiểm bias và chia sẻ kiến thức.

7 Liên hệ thực tiễn tại Việt Nam

Tại Việt Nam, nhiều tổ chức như VNPT AI, Viettel AI, FPT.AI đã bắt đầu hướng đến việc phát triển các hệ thống AI có trách nhiệm. Tuy nhiên, vấn đề bảo mật dữ liệu cá nhân, thiên kiến trong dữ liệu tiếng Việt, và thiếu khung pháp lý cụ thể vẫn là rào cản cần sớm được tháo gỡ.

8 Kết luận

Responsible AI không chỉ là một lựa chọn, mà là nền tảng thiết yếu để phát triển công nghệ một cách bền vững, minh bạch và có kiểm soát. Trong bối cảnh chuyển đổi số mạnh mẽ, việc áp dụng AI có trách nhiệm chính là chìa khóa để xây dựng niềm tin xã hội và tạo ra những giá trị nhân văn đích thực từ công nghệ.

9 Tài liệu tham khảo

Tài liệu

- [1] Market.us. (2024). *Responsible AI Market Size, Share & Growth Report 2034*. Truy cập từ: <https://market.us/report/responsible-ai-market/>
- [2] Microsoft. (2023). *Responsible AI Standard, v2*. Truy cập từ: <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2023/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf>
- [3] Google. (2023). *Responsible AI Practices*. Truy cập từ: <https://ai.google/responsibility/responsible-ai-practices/>
- [4] European Commission. (2021). *Proposal for a Regulation on Artificial Intelligence*. Truy cập từ: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>
- [5] Bộ Thông tin và Truyền thông. (2021). *Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2030*. Số 127/QĐ-TTg.
- [6] Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 1(9), 389-399.
- [7] Floridi, L., & Cowls, J. (2019). *A Unified Framework of Five Principles for AI in Society*. Harvard Data Science Review, 1(1).
- [8] VNPT AI. (2025). *Responsible AI là gì? Nguyên tắc cốt lõi và lý do doanh nghiệp cần triển khai ngay từ hôm nay*. Truy cập từ: <https://vnptai.io/vi/blog/detail/responsible-ai>
- [9] Viettel AI. (2023). *Giải pháp AI có trách nhiệm*. Truy cập từ: <https://viettel.ai/>
- [10] FPT.AI. (2023). *Nguyên tắc đạo đức AI*. Truy cập từ: <https://fpt.ai/vi>