Hội thảo trực tuyến về Trí tuệ nhân tạo (AI) 101

Ngày 22 tháng 10 năm 2025, 6:30-7:30 tối

Sở Công nghệ Thông tin Thành phố San José

Trình bày bởi

Nhóm Quyền riêng tư & AI
Khoa Công nghệ thông tin

Khoa Kỹ thuật Điện và Khoa học Máy tính
Đại học California, Berkeley

Chào mừng

- **Mục tiêu hôm nay:**
  - Giải thích cách thức hoạt động và rủi ro của AI
  - Chia sẻ cách Thành phố sử dụng và quản lý AI
  - Trình bày một trường hợp sử dụng AI
  - Lắng nghe câu hỏi và phản hồi của bạn

Trí tuệ nhân tạo (AI) là gì?

Khóa học cấp tốc + demo

Trí tuệ nhân tạo (AI) là gì?

1. Có rất nhiều loại AI khác nhau

AI dự đoán
"dự đoán điều gì đó trong tương lai"

Sự nhận thức Trí tuệ nhân tạo
"nhìn thấy điều gì đó trong thế giới thực"

Trí tuệ nhân tạo
"tạo ra cái gì đó"

ChatGPT

Google Translate

Nguồn: Trung tâm Bằng chứng Khoa học trong Các vấn đề Công cộng

Trí tuệ nhân tạo (AI) là gì?

1. Có rất nhiều loại AI khác nhau
2. AI là toán học, không phải phép thuật

AI hoạt động như thế nào?

Ví dụ: AI dự đoán

Khả năng San Jose Giants thắng trận đấu tuần tới là bao nhiêu?

Trận đấu tuần tới là một **trận đấu trên sân nhà.** → Trí tuệ nhân tạo → Có một **67%** cơ hội Giants sẽ thắng trận tiếp theo.

AI sử dụng **thuật toán** để tìm các mẫu trong **dữ liệu**

dữ liệu + thuật toán → mẫu

Vị trí	Kết quả
Trang chủ	Thắng
Xa	Thắng
Trang chủ	Mất
Xa	Mất
Xa	Mất
Trang chủ	Thắng

Bước 1. Có bao nhiêu trong số họ **trang chủ** Những trận đấu nào mà đội Giants thắng?

Thắng 2 trong 3 (67%) trận đấu trên sân nhà.

Bước 2. Có bao nhiêu trong số họ **xa** Những trận đấu nào mà đội Giants thắng?

Thắng 1 trong 3 (33%) trận đấu trên sân khách.

Trí tuệ nhân tạo

Trang chủ	67% cơ hội chiến thắng
Xa	Cơ hội chiến thắng 33%

AI sử dụng các mẫu dữ liệu trong quá khứ để đưa ra dự đoán

Khả năng San Jose Giants thắng trận đấu tuần tới là bao nhiêu?

Trận đấu tuần tới là một **trận đấu trên sân nhà.**

Trí tuệ nhân tạo

Trang chủ	67% cơ hội chiến thắng
Xa	Cơ hội chiến thắng 33%

Có một **67%** cơ hội Giants sẽ thắng trận tiếp theo.

AI sử dụng nhiều loại dữ liệu khác nhau để đưa ra các loại dự đoán khác nhau

Sự nhận thức Trí tuệ nhân tạo

"xác định"

A = A
a = A 4 ≠ A
A = A 4 ≠ A

A → Trí tuệ nhân tạo → California ABC1234 dmv.ca.gov

FASTRAK

Trí tuệ nhân tạo

"tạo ra"

Dữ liệu Internet

Của bạn là gì... → Trí tuệ nhân tạo → tên?

ChatGPT

Thành phố đang sử dụng AI như thế nào?

Trí tuệ nhân tạo trong thành phố

Dịch thuật AI vượt qua rào cản ngôn ngữ

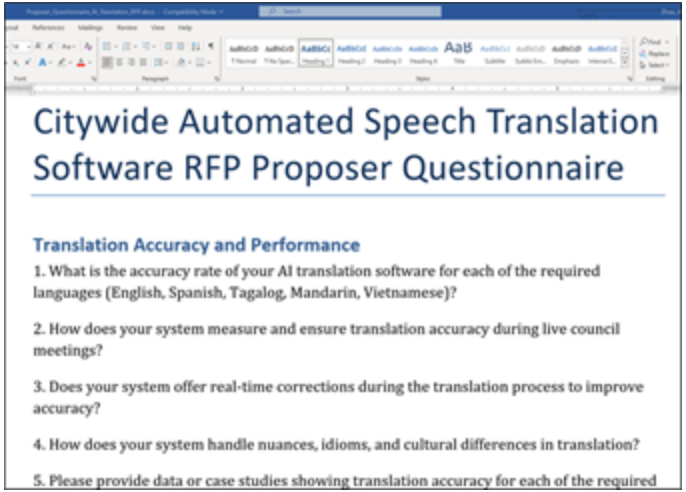
Thành phố sử dụng Wordly để phiên dịch hai chiều cho hơn 50 ngôn ngữ. Kể từ khi triển khai, hơn **5.500 cư dân** đã truy cập các cuộc họp của Hội đồng thành phố bằng tiếng Tây Ban Nha.

Tối ưu hóa giao thông tại các giao lộ bằng AI

Tuyến xe buýt số 66 và 68 ở trung tâm thành phố San José sử dụng tín hiệu giao thông AI để di chuyển tới **Nhanh hơn 20%** và chi tiêu **giảm 50%** thời gian ở đèn đỏ.

Trí tuệ nhân tạo trong thành phố

Soạn thảo Hồ sơ mời thầu



Citywide Automated Speech Translation Software RFP Proposer Questionnaire

Translation Accuracy and Performance

1. What is the accuracy rate of your AI translation software for each of the required languages (English, Spanish, Tagalog, Mandarin, Vietnamese)?

2. How does your system measure and ensure translation accuracy during live council meetings?

3. Does your system offer real-time corrections during the translation process to improve accuracy?

4. How does your system handle nuances, idioms, and cultural differences in translation?

5. Please provide data or case studies showing translation accuracy for each of the required

Sự va chạm: Tiết kiệm ~8 giờ cho mỗi lần mua sắm

- Yêu cầu ChatGPT tạo bản nháp trong vài phút để xác định những câu hỏi thực sự quan trọng.
- Khởi động quá trình phát triển chào hàng và tiết kiệm chuyên môn của nhân viên để tinh chỉnh các yêu cầu kinh doanh quan trọng

Thành phố bảo vệ quyền riêng tư của tôi và đảm bảo việc sử dụng AI an toàn như thế nào?

Chính sách AI của Thành phố quản lý việc sử dụng hệ thống AI.

- Thông qua tháng 6 năm 2024
  - Áp dụng cho tất cả AI được Thành phố sử dụng hoặc thay mặt Thành phố
  - Nguyên tắc chỉ đạo cho AI có trách nhiệm
  - Yêu cầu đánh giá AI đối với hoạt động mua sắm AI

City of San José | City Administrative Policy Manual

Artificial Intelligence (AI) Policy 1.7.12

PURPOSE

This policy establishes a governance structure that allows the City of San José (hereafter referred to as "City") to utilize Artificial Intelligence (AI) and AI systems (systems) while providing the necessary safeguards for purposeful and responsible use.

The key objectives of the AI Policy are to:

- Provide guidance that is clear, easy to follow, and supports decision-making for the staff, interns, consultants, contractors, partners, and volunteers who may be purchasing, configuring, developing, using, maintaining, or leveraging AI to provide services to the City;
- Ensure that the use of AI systems adheres to the Guiding Principles with regard to how systems are purchased, configured, developed, operated, or maintained;
- Define roles and responsibilities related to the usage of AI;
- Establish and maintain processes to assess and manage risks presented by AI;
- Align the governance of AI with existing data governance, security, and privacy measures in accordance with the City's Information and Systems Security Policy and City Council's Digital Privacy Policy;
- Define prohibited uses of AI systems;
- Establish "sunset" procedures to safely retire systems that no longer meet the needs of the City; and
- Define how AI may be used for legitimate purposes in accordance with applicable local, state, and federal laws, and existing agency policies.

AI systems and the data contained therein will be purchased, configured, developed, operated, and maintained using the City's AI Handbook, which will be managed by the Chief Information Officer (CIO).

SCOPE

This policy applies to:

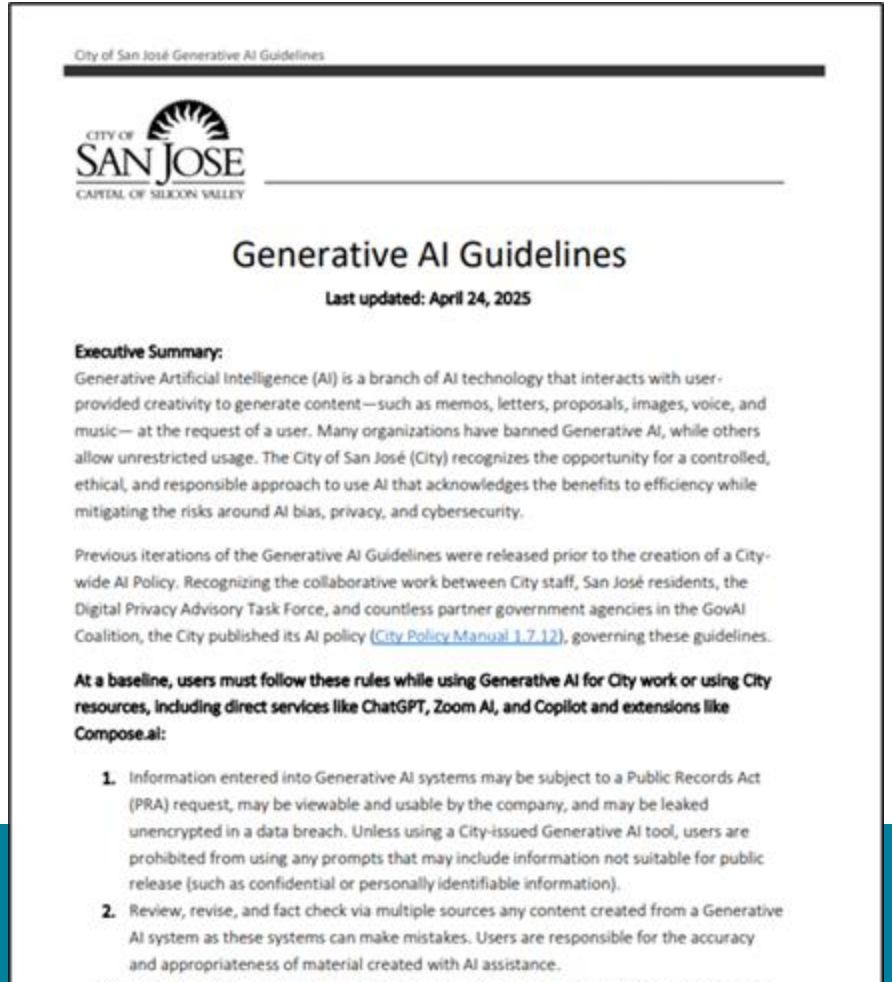
1. All systems deployed by the City; and
2. Staff, interns, consultants, contractors, partners, and volunteers who may be purchasing, configuring, developing, using, or maintaining the AI or who may be leveraging systems to provide services to the City (collectively referred to as "users").

GUIDING PRINCIPLES FOR RESPONSIBLE AI SYSTEMS

These principles describe the City's values with regard to how AI systems are purchased, configured, developed, used, or maintained.

1. **Effectiveness:** Systems are reliable, meet their objectives, and deliver precise and dependable outcomes for the utility and contexts in which they are deployed;
2. **Transparency:** The purpose and use of systems is proactively communicated and disclosed

Hướng dẫn về AI tạo sinh



City of San José Generative AI Guidelines

CITY OF SAN JOSE
CAPITAL OF SILICON VALLEY

Generative AI Guidelines

Last updated: April 24, 2025

Executive Summary:
Generative Artificial Intelligence (AI) is a branch of AI technology that interacts with user-provided creativity to generate content—such as memos, letters, proposals, images, voice, and music— at the request of a user. Many organizations have banned Generative AI, while others allow unrestricted usage. The City of San José (City) recognizes the opportunity for a controlled, ethical, and responsible approach to use AI that acknowledges the benefits to efficiency while mitigating the risks around AI bias, privacy, and cybersecurity.

Previous iterations of the Generative AI Guidelines were released prior to the creation of a City-wide AI Policy. Recognizing the collaborative work between City staff, San José residents, the Digital Privacy Advisory Task Force, and countless partner government agencies in the GovAI Coalition, the City published its AI policy (City Policy Manual 1.7.12), governing these guidelines.

At a baseline, users must follow these rules while using Generative AI for City work or using City resources, including direct services like ChatGPT, Zoom AI, and Copilot and extensions like Compose.ai:

1. Information entered into Generative AI systems may be subject to a Public Records Act (PRA) request, may be viewable and usable by the company, and may be leaked unencrypted in a data breach. Unless using a City-issued Generative AI tool, users are prohibited from using any prompts that may include information not suitable for public release (such as confidential or personally identifiable information).
2. Review, revise, and fact check via multiple sources any content created from a Generative AI system as these systems can make mistakes. Users are responsible for the accuracy and appropriateness of material created with AI assistance.

1. Đánh giá và kiểm chứng kết quả từ Trí tuệ nhân tạo (AI)
2. Tắt mô hình AI khỏi quá trình đào tạo các cuộc hội thoại của bạn
3. Trích dẫn bất kỳ hình ảnh, video hoặc âm thanh nào bạn tạo ra bằng GenAI là do AI tạo ra

Chương trình Đào tạo Nhân viên

Nhân viên được đào tạo về cách sử dụng AI một cách an toàn và có trách nhiệm cho công việc của mình.

Chương trình Nâng cao Kỹ năng AI

- Khởi xướng năm 2024
- Đào tạo nhân viên Thành phố xây dựng và sử dụng các công cụ AI tùy chỉnh để nâng cao năng suất làm việc và cung cấp dịch vụ.
- Nhấn mạnh những điều nên làm và không nên làm + sử dụng có trách nhiệm
- Hợp tác với Đại học Bang San José

Bản demo: Sử dụng ChatGPT để viết thư cho Hội đồng Thành phố

Những rủi ro tiềm ẩn của AI là gì và chúng ta có thể giảm thiểu chúng như thế nào?

Điều quan trọng là phải suy nghĩ về thời điểm và cách thức bạn có thể gặp phải những rủi ro này.

Khi **Bạn** sử dụng AI

Khi **người khác** sử dụng AI khi tương tác với bạn

Ứng dụng AI rộng rãi hơn trong xã hội

Những rủi ro tiềm ẩn khi sử dụng AI

Rủi ro: Ngôn ngữ hoặc hành vi có hại

They Asked an A.I. Chatbot Questions. The Answers Sent Them Spiraling.

Generative A.I. chatbots are going down conspiratorial rabbit holes and endorsing wild, mystical belief systems. For some people, conversations with the technology can deeply distort reality.

https://www.nytimes.com/2025/06/13/technology/chatgpt-ai-chatbots-conspiracies.html

Rủi ro: Bản dịch không chính xác, thông tin sai lệch

JAMA Internal Medicine

Home | JAMA Internal Medicine | Vol. 179, No. 4

Research Letter FREE

Assessing the Use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions

Elaine C. Khoong, MD, MS[1]; Eric Steinbrook, BA[2]; Cortlyn Brown, MD[3]; et al

https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2725080

Các công ty có thể thay đổi cách xây dựng các chatbot này

Kiểm tra lại bản dịch với người biết cả hai ngôn ngữ.

Rủi ro tiềm ẩn khi người khác sử dụng AI để đưa ra quyết định ảnh hưởng đến bạn

Rủi ro: Dự đoán có thể lặp lại các mô hình phân biệt đối xử trong lịch sử

Want a job? Employers say: Talk to the computer

1.https://apnews.com/article/job-seekers-artificial-intelligence-hiring-decisions-784ccec7ac6175aadc573465ab604fb9

AI Detection Tools Falsely Accuse International Students of Cheating

Stanford study found AI detectors are biased against non-native English speakers

By Tara García Mathewson

August 14, 2023 08:00 ET

2.https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating

Những người xây dựng và sử dụng AI có thể kiểm tra AI và kiểm tra các mô hình phân biệt đối xử tiềm ẩn *trước* sử dụng công nghệ

Rủi ro tiềm ẩn về mặt xã hội

Rủi ro: Chi phí môi trường đằng sau việc xây dựng và sử dụng AI

As data centers for AI strain the power grid, bills rise for everyday customers

Explained: Generative AI's environmental impact

Rapid development and deployment of powerful generative AI models comes with environmental consequences, including increased electricity demand and water consumption.

https://www.smithsonianmag.com/science-nature/with-ai-on-the-rise-what-will-be-the-environmental-impacts-of-data-centers-180987379/

https://apnews.com/article/virginia-election-data-centers-prince-william-229cb44d34ccf4bd1cc4e9f0d0131649

Đặt câu hỏi khi nào là phù hợp để sử dụng AI tạo sinh, giảm số lượng lời nhắc và viết lời nhắc rõ ràng một cách chiến lược.

Làm thế nào chúng ta có thể giảm thiểu rủi ro AI?

Những thay đổi về cách mọi người **xây dựng** Trí tuệ nhân tạo

ví dụ, thiết kế chatbot để hỗ trợ các tương tác lành mạnh hơn

Những thay đổi về cách chúng tôi **sử dụng** Trí tuệ nhân tạo

ví dụ, kiểm tra lại kết quả đầu ra của AI

Những thay đổi đối với ***liệu*** chúng ta xây dựng hoặc sử dụng AI ngay từ đầu

Ví dụ, đối với một số vấn đề, quyết định rằng AI không phải là giải pháp tốt nhất

Nguồn biểu tượng: Flaticon.com

Bản tóm tắt

- **Trí tuệ nhân tạo (AI) là gì?**
  - Có rất nhiều loại AI khác nhau
  - AI sử dụng các mẫu trong dữ liệu để dự đoán, quan sát thế giới và tạo ra
- **Thành phố đang sử dụng AI như thế nào?**
  - Ví dụ: Biên dịch, Vận tải công cộng, Soạn thảo hợp đồng mua sắm
- **Có những rủi ro tiềm ẩn nhưng có cách để giảm thiểu chúng.**
  - Các rủi ro bao gồm ngôn ngữ gây khó chịu, bản dịch không chính xác, phân biệt đối xử và tác động đến môi trường.
  - Chúng ta có thể giảm thiểu những rủi ro này bằng cách thay đổi AI mà chúng ta xây dựng và sử dụng, cũng như cách chúng ta xây dựng và sử dụng nó.

Tiếng nói của bạn rất quan trọng đối với cách Thành phố sử dụng AI

Cơ hội tham gia:

- Các cuộc họp của hội đồng thành phố
- Sự kiện giao lưu cộng đồng
- Khảo sát

Ý kiến đóng góp của bạn có thể định hình các quyết định mà Thành phố đưa ra về loại AI nào sẽ được sử dụng và cách thức sử dụng.

Nguồn biểu tượng:
Flaticon.com

Trước phần Hỏi & Đáp: Thông tin về Nghiên cứu và Tùy chọn Từ chối

- Chúng tôi đang tiến hành nghiên cứu để tìm hiểu những thắc mắc và suy nghĩ của bạn về AI và về việc Thành phố sử dụng AI.
  - Chúng tôi sẽ ghi chú lại những gì bạn nói trong phần Hỏi & Đáp, chính xác như cách bạn nói.
  - Tất cả ghi chú sẽ được ẩn danh.
  - Mặc dù không mang lại lợi ích trực tiếp cho bạn, nhưng phản hồi của bạn có thể giúp chúng tôi cải thiện các buổi tương tác trong tương lai như buổi này.
- Nếu bạn không muốn bất kỳ câu hỏi nào của mình được đưa vào, vui lòng cho chúng tôi biết.
  - Bạn có thể cho chúng tôi biết bất cứ lúc nào (ví dụ: trước khi bạn đặt câu hỏi, sau khi bạn đặt câu hỏi hoặc bạn có thể nhắn tin riêng cho Angela hoặc Ezinne).

Hỏi & Đáp + Khảo sát

Chia sẻ suy nghĩ của bạn và đặt câu hỏi.

SỰ KHẢO SÁT

Chia sẻ phản hồi ẩn danh về phiên họp

Cảm ơn bạn đã tham gia

- Chúng tôi hy vọng bạn đã học được:
  - Thành phố đang sử dụng AI một cách có trách nhiệm như thế nào
  - AI là gì và không phải là gì
  - Cách sử dụng AI cho hoạt động cộng đồng

Giữ kết nối

Vui lòng liên hệ digitalprivacy@sanjoseca.gov nếu bạn có thắc mắc.
Tìm hiểu thêm tại www.sanjoseca.gov/govai.