

A Case Study on Delegating Critical Tasks to Agentic AI

Sunyoung Kim and Hokeun Kim

IEEE 5th Cyber Awareness and Research Symposium 2025 (CARS'25)

University of North Dakota, Grand Forks, ND, USA

Monday, October 27, 2025



ASU KIM

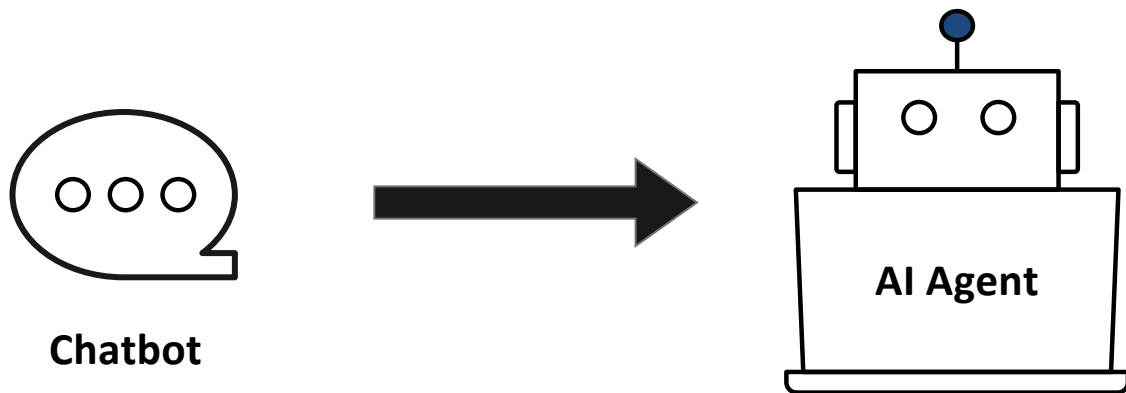
KNOWLEDGEABLE &
INTERACTIVE MACHINES

ASU Arizona State
University

Motivation

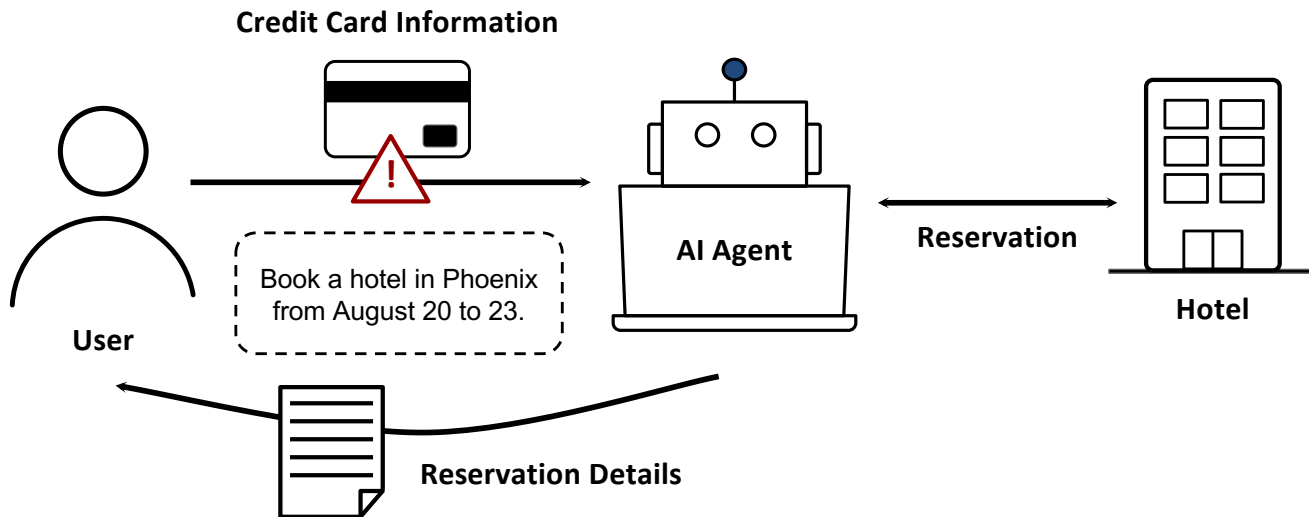
Emerging Interests in **Agentic AI**: beyond chatbots → Autonomous planning / decision making

Agentic AI systems that can handle critical tasks, but raise security & privacy concerns



Motivation

Agentic AI can interact with external resources, which often require users to provide the AI with sensitive information, such as credit card information, address, or phone numbers.



Background

Case in Industry:

Mastercard's Agent Pay [1]

- *Delegated payment request*
- *Tokenized transaction*
- *Authorized payment confirmation*

However, the system is limited to Mastercard payment providers only.

[1] Mastercard, "Mastercard unveils Agent Pay, pioneering agentic payments technology to power commerce in the age of AI," 2025, Accessed: 2025-08-20. [Online]. Available: <https://www.mastercard.com/us/en/news-and-trends/press/2025/april/mastercard-unveils-agent-pay-pioneering-agentic-payments-technology-to-power-commerce-in-the-age-of-ai.ht>

Image Source: <https://www.mastercard.com/us/en/business/artificial-intelligence/mastercard-agent-pay.html>



TRANSFORMING COMMERCE

Introducing Mastercard Agent Pay

Mastercard's new infrastructure for enabling secure, scalable and trusted payments in agentic commerce.



Trust

We're reimagining trust for the agentic era — anchored in Mastercard's industry-leading standards.



Security

Creating a world where agent-led payments are secure, seamless and future-ready.



Visibility

Enabling insight, personalization and confidence in every transaction.

Background

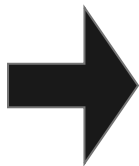
AI agents face significant limitations when performing complex tasks or end-to-end activities

- **Security Risks:**

exposing personal information

- **Technical Barriers:**

- reCAPTCHA / bot detection
- Multi-factor authentication (MFA)
- UI limitations



**High Dependence on
Human Intervention**

- **Hallucinations**

Case Study

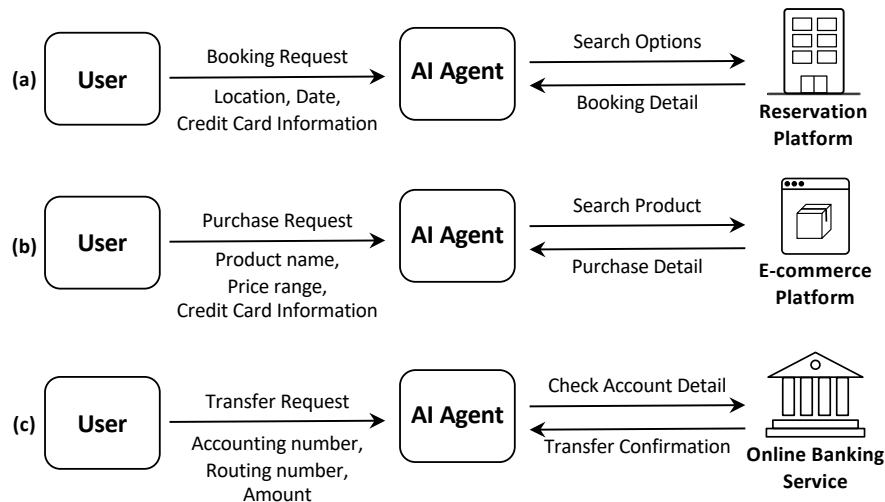
Scenarios and Setup

Overview:

- Evaluated current Agentic AI capabilities in handling sensitive data
- Conducted real-world task scenarios

Experiment Environment

- Model: Meta Llama-3.1-8B-Instruct and Llama-3.2-3B-Instruct
- Python 3.12.11
- NVIDIA A100 GPU
 - ASU Sol Supercomputer



Case Study

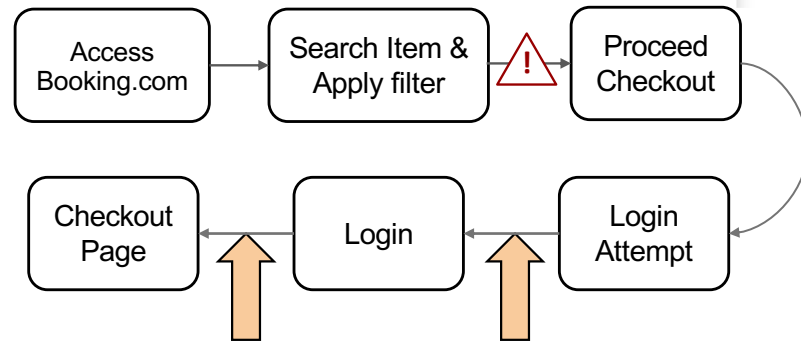
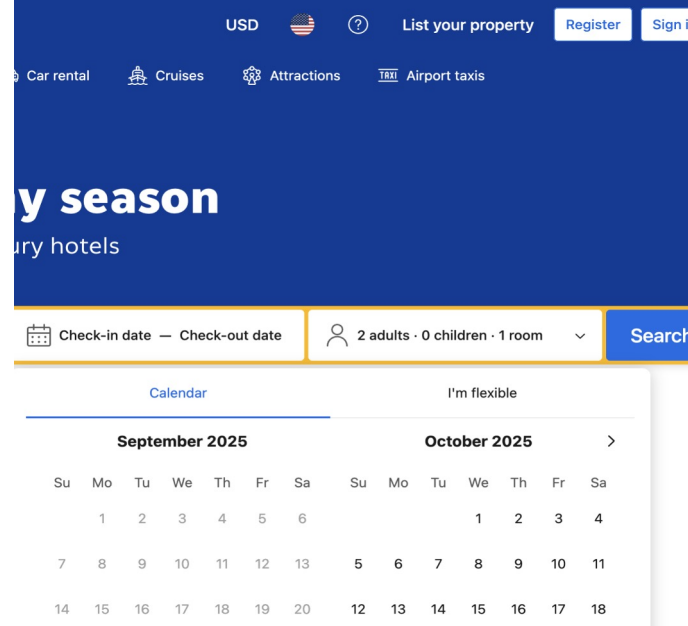
(a) Hotel Reservation - Booking.com

Example Scenario

“Book a hotel in New York City from Sep 28 to Sep 30. I want the hotel with 4 star or above.”

Results

- Agent **searched** destination and applied 4 star filter successfully
- **Failed** to select dates and guests
- Login needed **human interventions** due to MFA and reCAPTCHA



Human Interventions

Case Study

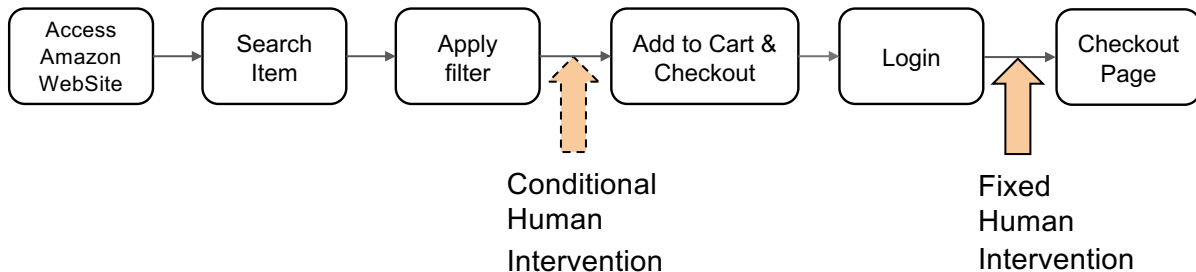
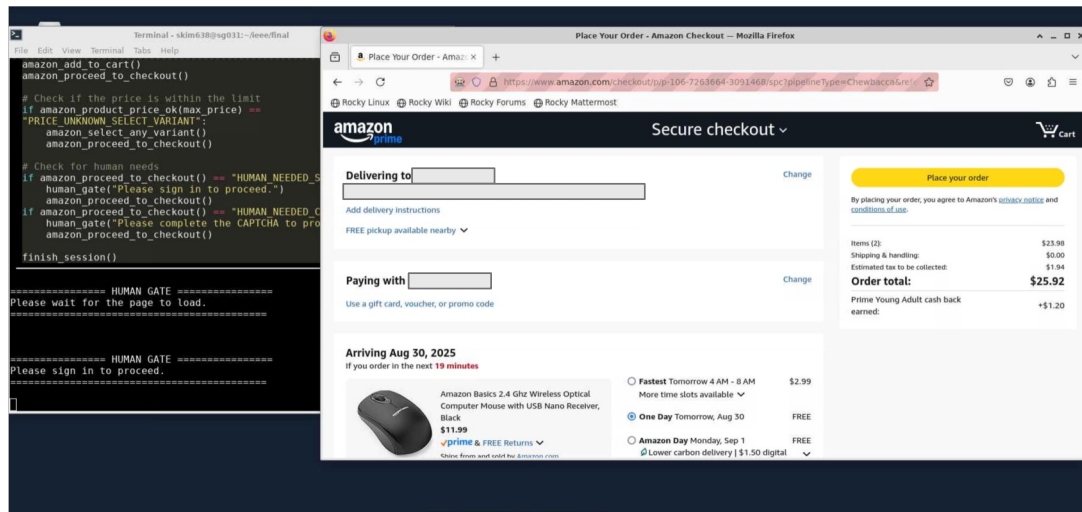
(b) Product Purchase - Amazon

Example Scenario

"Purchase a cup. The price range should be \$1 - \$10."

Results

- Intermittent success after resolving MFA with human intervention
- Agent found and added items within price range
- Duplicate cart times caused by unintended actions



Case Study

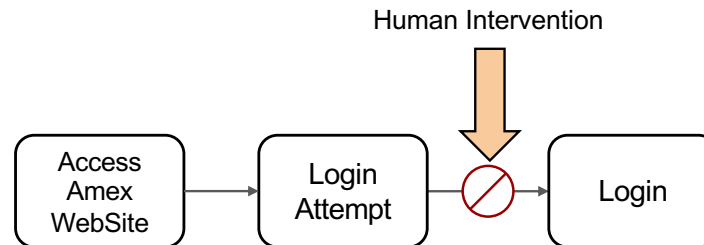
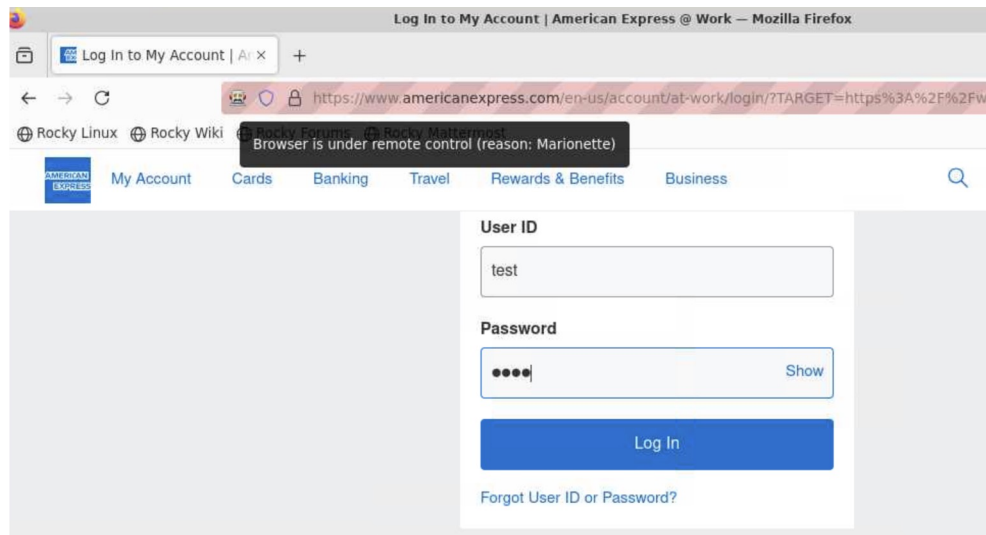
(c) Banking - American Express

Example Scenario

"I want to transfer from my checking account to my debit account."

Results

- **Completely blocked**
by **bot detection** with valid user credentials
- Security system prevented automated login



Case Study

Summary

Hallucination:

- Incorrectly judged task as failed or completed
- Requested human help unnecessarily

Unintended action:

- Continued task after completion
- Added multiple items to cart (purchase scenario)

Scenario	Trials	Task Completion Rate	Search Success Rate	Login Success Rate	Average Number of Human Intervention	Hallucination Rate	Unintended Action Rate
(a) Hotel Reservation	5	0%	0%	20%	0.4	80%	20%
(b) Product Purchase	5	60%	100%	60%	1.4	40%	60%
(c) Banking System	5	0%	N/A	0% (bot detection)	1	20%	0%

Codes and Logs are available on: <https://github.com/asu-kim/agent-ai-access-control>

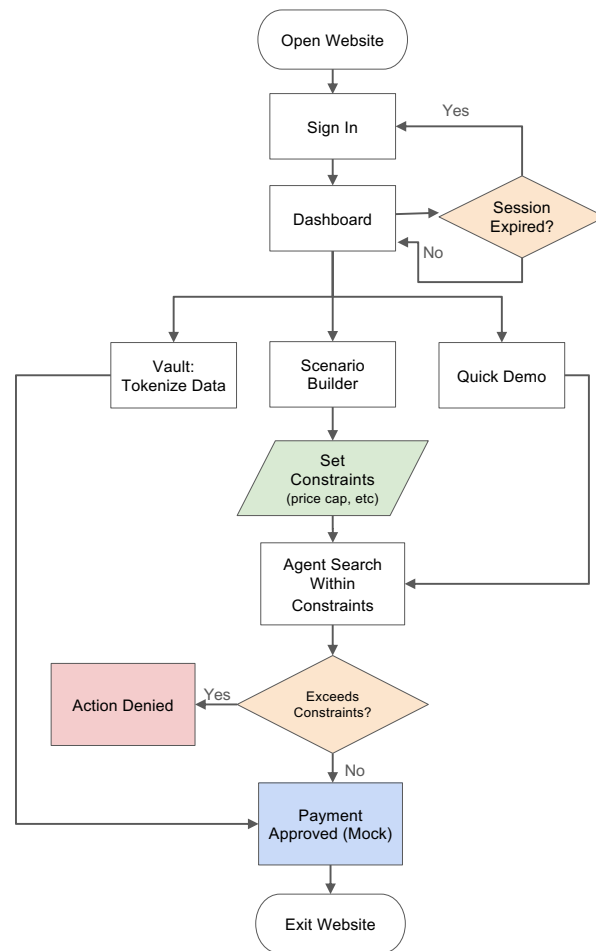
Proposed Design

Prototype Website

Prototype Website Design:

- Session auto-logout after task
- Tokenization of sensitive data
- Fine-grained access control
 - Constraints on price, rating, etc

https://github.com/asu-kim/agent-ai-access-control/tree/main/proposed_website



Proposed Design

Prototype Website Demo

The screenshot shows a web browser window with the address bar displaying '127.0.0.1:3000/vault'. The browser's address bar also shows 'Verify it's safe'. The website has a dark theme and a navigation bar with links: Home, Dashboard, Vault, Workflows, and Sign out. The main content area is titled 'Mock Credit Card / Address Vault' and contains a form with the following fields:

- Cardholder Name:
- Card Number:
- CVV:
- Address Line 1:
- City:
- State:
- ZIP:

Below the ZIP field is a 'Save to Vault' button. At the bottom of the page, there is a small copyright notice: 'Copyright © 2025, ARIZONA STATE UNIVERSITY'.

Summary

Conclusion

- Delegating critical tasks to Agentic AI requires strong security and reliability
- Case study
 - Authentication challenges
 - Restricted automated access
- Proposed prototype website
 - Session limits, tokenized data, and user-defined boundaries

Future Work

- Integrate prototype with open-source key distribution service
 - Kerberos
 - Secure Swarm Toolkit (SST)
- Combine tokenization + authorization

Contacts

Sunyoung Kim skim638@asu.edu

Hokeun Kim hokeun@asu.edu

<https://github.com/asu-kim>

Prototype Website



ASU KIM
KNOWLEDGEABLE &
INTERACTIVE MACHINES