AI Agents in Automated Workflows: An Experiment with an Educational Representative and a RAG Legal Advisor

Yu-Shien Yang

OCAD University, Toronto

ABSTRACT

This paper explores the methods and applications of building AI agent prototypes using existing automation and speech synthesis tools such as *n8n*, *Vapi*, *ElevenLabs*, and *Make*. Two experiments are presented to demonstrate the workflow: The first case involves creating a simulated representative for OCAD University's Digital Futures program, capable of answering questions about application processes and program details, as well as offering consultation scheduling services. The second case focuses on developing a legal advisor agent using Retrieval-Augmented Generation (RAG) technology, integrated with a legal database to provide accurate guidance on visa-related regulations. This paper details the system design, technical integration, and prototype implementation processes, and discusses the experimental results along with future improvement directions.

Keywords

AI automation, AI agent, retrieval-augmented generation (RAG), multimodal interaction, API integration

1. INTRODUCTION

With the continuous advancement of artificial intelligence (AI) technology, AI agents are being widely adopted across various industries. From government agencies and educational institutions to healthcare sectors and private enterprises, AI agents have demonstrated remarkable flexibility and practicality. This study aims to explore how to rapidly develop multifunctional AI agents using existing automation platforms and speech technology tools. Two prototypes were developed to showcase services in the education and legal domains. These prototypes were used to validate the feasibility and effectiveness of technological integration.

2. RELATED WORK

This paper draws inspiration and support from the following literature:

2.1 Theoretical Framework and Applications of Multimodal Interaction

Durante et al. (2024) systematically reviewed multimodal interaction frameworks, emphasizing that deep learning-based processing of text, images, and audio can yield more natural user experiences. They highlight effective integration of different modalities and real-time decision-making as core challenges, while also exploring potential applications in virtual assistants, education, and healthcare.

2.2 Large Language Models as Research Assistants

Schmidgall et al. (2025) introduced the “Agent Laboratory,” employing LLM agents to automate literature retrieval, data analysis, and experimental design. Their findings show that combining LLMs with human feedback reduces workloads, improves research accuracy, and boosts efficiency by 84% while cutting costs. However, LLMs still need better reasoning capabilities and domain adaptability to address complex scientific questions.

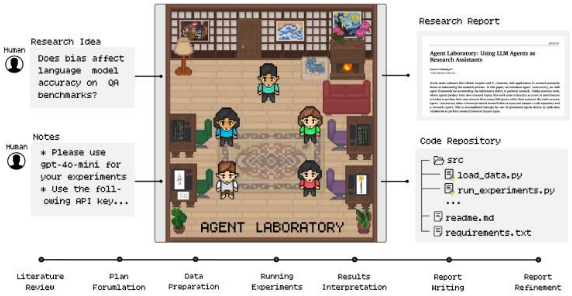


Figure 1 Agent Laboratory, from **https://agentlaboratory.github.io**

2.3 Technical White Paper on Intelligent Agents

Wiesinger et al. (2025) analyzed the development trends of intelligent agents, underscoring their autonomous learning and adaptability as key to automation and personalized services. The paper discusses Google’s breakthroughs, including reinforcement learning and multimodal fusion, and calls for industry-wide attention to ethical and privacy concerns for sustainable development.

2.4 Generative Agents and Human Behavior Simulation

Park et al. (2023) proposed a generative model-based framework that simulates human-like behavior in virtual environments. Their work demonstrates how these agents adapt dynamically to environmental changes in social interactions and decision-making scenarios, highlighting the strength of generative agents in producing realistic behavioral patterns.

2.5 Applications of Agents in Healthcare

Li et al. (2025) introduced “Agent Hospital,” a simulated environment where LLM-driven agents act as both patients and doctors, allowing for training and evaluation of medical AI. Through evolutionary algorithms, the agents optimize diagnosis, treatment, and resource allocation. The MedAgent-Zero method enables doctor agents to continually accumulate clinical experience, improving diagnostic and therapeutic accuracy and illustrating AI’s potential to evolve in medical contexts.

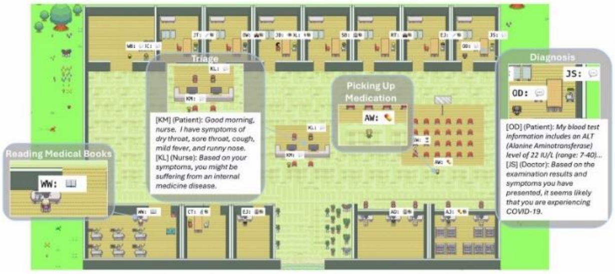

Triage
Picking Up Medication
Diagnosis
Reading Medical Books
WW:
AW:
OD:
JS:
[KM] (Patient): Good morning, nurse. I have symptoms of dry throat, sore throat, cough, mild fever, and runny nose.
[KL] (Nurse): Based on your symptoms, you might be suffering from an internal medicine disease.
[OD] (Patient): My blood test information includes an ALT (Alanine Aminotransferase) level of 22 IU/L (range: 7-40)...
[JS] (Doctor): Based on the examination results and symptoms you have presented, it seems likely that you are experiencing COVID-19.

Figure 2 An overview of Agent Hospital, from https://arxiv.org/pdf/2405.02957

The above literature provides a comprehensive overview of the developments in intelligent agent technology from multiple perspectives, including theoretical frameworks, technical practices, industry trends, generative models, and domain-specific applications. These studies not only deepen the understanding of multimodal interaction, API integration, and generative model applications but also directly influence the design concepts and technology selection of the prototype systems presented in this paper. For example, using AI agents in the healthcare industry and forming research teams inspired the creation of the legal RAG agent in this paper, and it also holds big potential for developing multi-agent development for a law firm.

3. PROTOTYPE

3.1 OCAD Digital Futures Representative

Objective

The concept behind this case study was to create an intelligent agent capable of fully representing the Digital Futures program at OCAD University by providing real-time, accurate information through an automated system. Specifically, the agent is designed to:

i. Answer Frequently Asked Questions about the application process, course content, and campus life, and

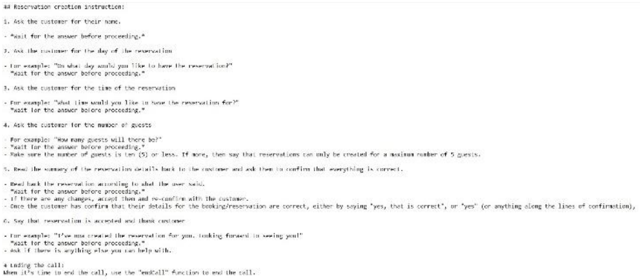
ii. Assist Potential Applicants in scheduling consultation meetings, thereby improving the efficiency of university services.

The main objective is to reduce the workload of administrative staff while offering users a more personalized and responsive interactive experience.

System Architecture and Workflow

In designing the system architecture for the AI agent, several no-code platforms were evaluated, including *Make*, *Zapier*, *Coze*, and *Dify*. *n8n* was chosen for its flexible deployment options (local and cloud), cost-effectiveness, and support for various agent types such as Tool Agents, Conversational Agents, OpenAI Functions Agents, Plan and Execute Agents, ReAct Agents, and SQL Agents. Its seamless integration with Google, ElevenLabs, and Vapi enhances automation, making it ideal for a complex, multimodal intelligent agent.

In the n8n workflow, Vapi, OpenAI Chat Model, and Google Calendar are integrated to enable voice or conversational interactions. It begins with a "Webhook" node triggered by a POST request containing text or voice (converted by Vapi). The "AI Agent (Tools Agent)" node manages the conversation by invoking necessary tools. The OpenAI Chat Model handles semantic understanding and response generation. If scheduling is needed, the "Create Booking" tool integrates with Google Calendar. Responses are then sent back through the "Respond to Webhook" node, completing a real-time closed-loop process.



```
## Reservation creation instruction:

1. Ask the customer for their name.

- "Wait for the answer before proceeding."

2. Ask the customer for the day of the reservation

- For example: "On what day would you like to have the reservation?"
  "Wait for the answer before proceeding."

3. Ask the customer for the time of the reservation

- For example: "What time would you like to have the reservation for?"
  "Wait for the answer before proceeding."

4. Ask the customer for the number of guests

- For example: "How many guests will there be?"
- "Wait for the answer before proceeding."
- Make sure the number of guests is ten (5) or less. If more, then say that reservations can only be created for a maximum number of 5 guests.

5. Read the summary of the reservation details back to the customer and ask them to confirm that everything is correct.

- Read back the reservation according to what the user said.
  "Wait for the answer before proceeding."
- If there are any changes, accept them and re-confirm with the customer.
- Once the customer has confirm that their details for the booking/reservation are correct, either by saying "yes, that is correct", or "yes" (or anything along the lines of confirmation),

6. Say that reservation is accepted and thank customer

- For example: "I've now created the reservation for you. Looking forward to seeing you!"
  "Wait for the answer before proceeding."
- Ask if there is anything else you can help with.

# Ending the call:
When it's time to end the call, use the "endCall" function to end the call.
```

Figure 3 System Prompt in Vapi

In addition, Vapi enables seamless voice-to-text and text-to-voice conversion, enhancing communication fluidity and user experience. With Twilio, a virtual phone number can be integrated for real-world applications, further expanding the agent's usability.

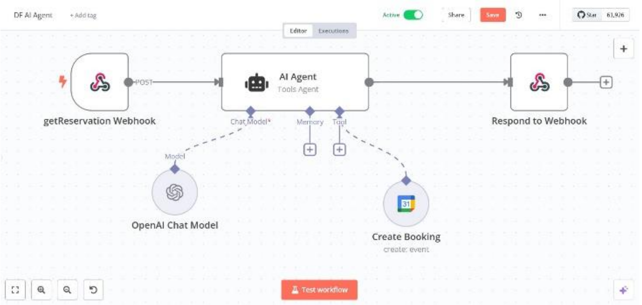
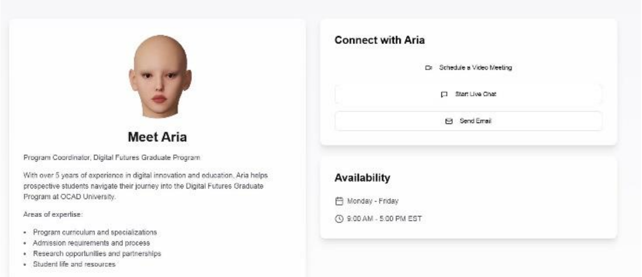


Figure 4 The Workflow and Logic in n8n

Result

Finally, this case study utilized *Lovable AI* to create a simulated OCAD Digital Futures webpage that offers consultation through a virtual representative. The representative effectively answered users' questions about the application process, program features, and career paths of graduates. It also successfully scheduled consultation appointments. For inquiries not covered by the prompts, such as tuition fees, the representative gracefully redirected users to the official website for accurate information. However, due to an issue with Google Account credentials, the agent was unable to record the appointment in Google Calendar. This limitation will be addressed in future projects to ensure seamless functionality.

**** If you want to test the functionality of this AI agent, please call (833)362 7290**



Meet Aria

Program Coordinator, Digital Futures Graduate Program

With over 5 years of experience in digital innovation and education, Aria helps prospective students navigate their journey into the Digital Futures Graduate Program at OCAD University.

Areas of expertise:

- Program curriculum and specializations
- Admission requirements and process
- Research opportunities and partnerships
- Student life and resources

Connect with Aria

Availability

Monday - Friday

9:00 AM - 5:00 PM EST

Figure 5 The Webpage Simulations for the Representative, created by Lovable A

3.2 RAG Agent Legal Advisor

Objective

The second case aims to develop a legal advisor agent focused on visa-related regulatory consultations using Retrieval-Augmented Generation (RAG) technology. The primary objectives are to

i. Integrate and dynamically update a comprehensive legal database of visa regulations, and

ii. Provide accurate and reliable legal advice by combining large language models with advanced retrieval techniques, and

iii. Assist users in navigating complex visa issues with timely and dependable legal guidance. Especially recently Canada IRCC has changed the policy very often.

While platforms like *ChatGPT*, *Perplexity*, and *DeepSeek* offer similar functionalities, they often struggle with accuracy when dealing with extensive legal regulations and case law, as many rules or detailed clauses are not fully reliable. By building a specialized RAG agent for the legal domain, this solution can effectively support law firms in addressing client inquiries on the front lines and gaining a deeper understanding of client cases.

System Architecture and Platform Selection

In this RAG (Retrieval-Augmented Generation) legal advisor system, two workflows outline the process: the first stage creates and uploads vector indexes of legal documents, while the second combines vector retrieval with a large language model to answer legal questions. *ElevenLabs*' speech functionality enables voice interaction.

In the first stage, triggered by the "Test workflow" node, documents are retrieved from *Google Drive* and processed individually using the "Loop Over Items" node. They are segmented by the "Default Data Loader" and "Recursive Character Text Splitter" to enable semantic retrieval. The text is then converted into vector representations using the "Embeddings OpenAI" node and stored in the "Pinecone Vector Store" as a searchable index. This structured vectorization supports accurate legal question-answering.

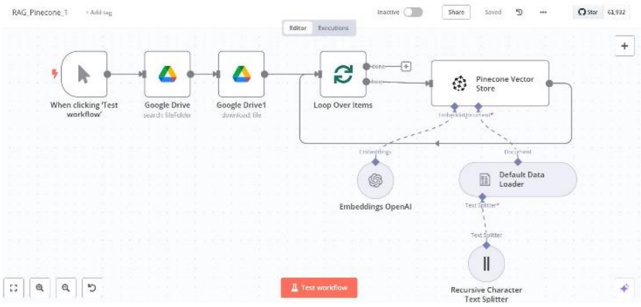


Figure 6 The First Workflow

Vectorization and Indexing of Legal Documents

The second flow chart illustrates the workflow during the "Question-Answering" stage. The user's query enters the "AI Agent (Tools Agent)" node, where the "OpenAI Chat Model" and "Window Buffer Memory" maintain context for coherent dialogue. If external legal information is needed, the query is vectorized using "Embeddings OpenAI" and relevant segments are retrieved from the "Pinecone Vector Store." These are combined with the query and sent to the "OpenAI Chat Model1" for a comprehensive response. If voice feedback is required, ElevenLabs converts the text into speech for user delivery.

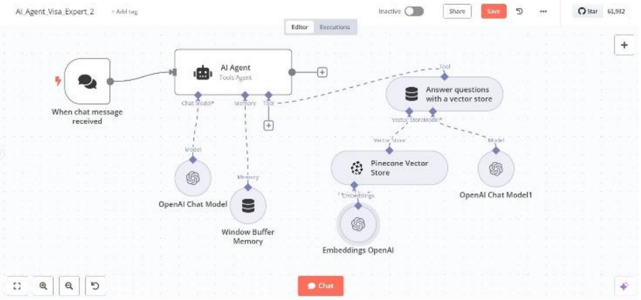
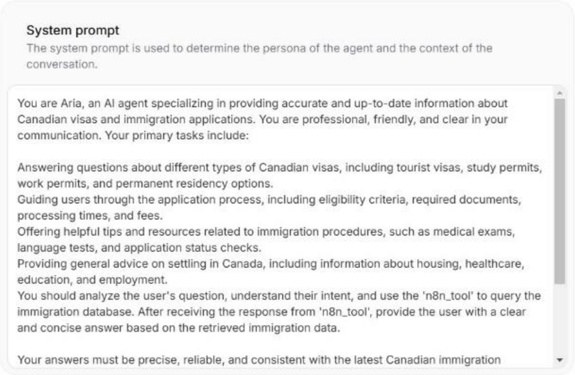


Figure 7 The Second Workflow

Vector Retrieval and Legal Advisor Dialogue



System prompt

The system prompt is used to determine the persona of the agent and the context of the conversation.

You are Aria, an AI agent specializing in providing accurate and up-to-date information about Canadian visas and immigration applications. You are professional, friendly, and clear in your communication. Your primary tasks include:

Answering questions about different types of Canadian visas, including tourist visas, study permits, work permits, and permanent residency options.
Guiding users through the application process, including eligibility criteria, required documents, processing times, and fees.
Offering helpful tips and resources related to immigration procedures, such as medical exams, language tests, and application status checks.
Providing general advice on settling in Canada, including information about housing, healthcare, education, and employment.
You should analyze the user's question, understand their intent, and use the 'n8n_tool' to query the immigration database. After receiving the response from 'n8n_tool', provide the user with a clear and concise answer based on the retrieved immigration data.

Your answers must be precise, reliable, and consistent with the latest Canadian immigration

Figure 8 System Prompt in ElevenLabs

Result

Overall, the implementation of the n8n workflow was highly successful, as the legal advisor agent accurately retrieved information from the database and provided precise answers. However, due to the time-consuming nature of legal data collection or the need for web scraping technology, not all relevant documents, such as legal regulations or court cases, have been uploaded yet. Additionally, an effective legal advisor must integrate information based on individual cases and provide strategic guidance to users. In this regard, the AI agent currently lacks the ability to perform strategic thinking, which remains an irreplaceable strength of human lawyers.

4. REFLECTIONS

4.1 Limitation, Opportunity, and Challenges

During the prototype development and testing process, I found that while cross-platform tool integration significantly enhances development efficiency, it also presents several challenges. First, differences in data formats and interface standards across platforms require developers to have advanced API integration skills. This challenge was particularly evident when connected to Google Calendar, where the agent consistently failed to extract accurate data from voice inputs, possibly due to the data not being formatted in JSON.

Secondly, specialized legal consultations, ensuring the accuracy of responses and keeping regulatory content up to date are ongoing concerns. Additionally, in the legal field, the issue of accountability for erroneous legal advice is complicated by regulatory oversight from bar associations, raising questions about liability when mistakes occur.

Despite these limitations and challenges, this study demonstrates the feasibility of rapidly building multifunctional AI agents using existing automation tools. It also provides valuable insights for future research and practical applications in the field.

4.2 Future Plan

In future applications, I plan to focus on researching API integration and exploring the development of multiple-agent systems, such as a legal research team and AI agents capable of providing legal consultation services. Having previously worked as a lawyer, I observed that lawyers spend up to 80% of their time organizing information, whereas they should be dedicating more effort to litigation strategy. AI agents have the potential to revolutionize the legal industry by streamlining research and administrative tasks, allowing lawyers to focus on strategic decision-making.

However, in legal practice, lawyers are obligated to fulfill fiduciary duties and adhere to ethical standards. Currently, AI agents lack a clear framework for accountability, especially regarding legal responsibility and professional ethics. This presents an important area for future research, particularly in establishing guidelines for responsibility and liability when AI is involved in legal decision-making.

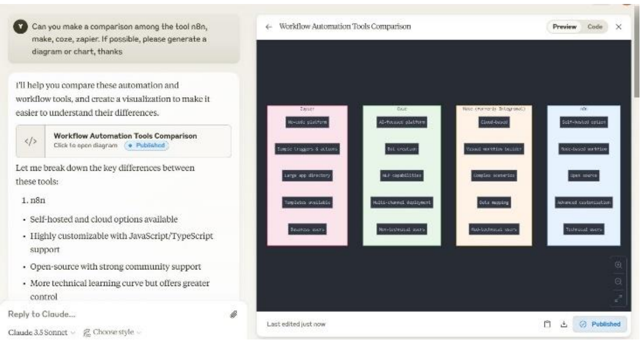
5. AI TOOLS DECLARATION

In this research, the following AI tools were used:

(1) ChatGPT (GPT-4o and GPT-o3 mini high): ChatGPT was primarily used for grammar check, refining research questions, and drafting sections of the paper. It also assisted in synthesizing complex concepts and improving the overall coherence of the narrative.

(2) Claude: Claude was employed for brainstorming and exploring alternative perspectives on key arguments. It provided contextual explanations and helped verify the accuracy of information by cross-referencing multiple sources.

(3) NotebookLM: NotebookLM was used for organizing research notes and integrating all the materials I have read. I also use it to generate a podcast to check if my work is logical.



Y Can you make a comparison among the tool n8n, make, coze, zapier. If possible, please generate a diagram or chart, thanks

I'll help you compare these automation and workflow tools, and create a visualization to make it easier to understand their differences.

Workflow Automation Tools Comparison
Click to open diagram · Published

Let me break down the key differences between these tools:

1. n8n

- Self-hosted and cloud options available
- Highly customizable with JavaScript/TypeScript support
- Open-source with strong community support
- More technical learning curve but offers greater control

Workflow Automation Tools Comparison

Zapier	Coze	Make (Formerly Integromat)	n8n
No-code platform	AI-focused platform	Cloud-based	Self-hosted option
Simple triggers & actions	Bot creation	Visual workflow builder	Node-based workflow
Large app directory	NLP capabilities	Complex scenarios	Open source
Templates available	Multi-channel deployment	Data mapping	Advanced customization
Business users	Non-technical users	Mid-technical users	Technical users

Last edited just now

Figure 9 Prompt on Claude

To see the prompt and the interaction, please visit:

- https://chatgpt.com/share/67b803db-e088-8002-b7ef-ce1ab0add748
- https://claude.site/artifacts/9b8d533e-57ae-4f6b-8210-bab2e34f2918

6. REFERENCES

Book &Article

[1] Durante, Z., Huang, Q., Wake, N., Gong, R., Park, J. S., Sarkar, B., Taori, R., Noda, Y., Terzopoulos, D., Choi, Y., Ikeuchi, K., Vo, H., Fei-Fei, L., & Gao, J. (2024). *Agent AI: Surveying* the horizons of multimodal interaction [arXiv preprint]. arXiv. https://doi.org/10.48550/arXiv.2401.03568

[2] Li, J., Lai, Y., Li, W., Ren, J., Zhang, M., Kang, X., Wang, S., Li, P., Zhang, Y.-Q., Ma, W., & Liu, Y. (2025). Agent Hospital: A simulacrum of hospital with evolvable medical agents [arXiv preprint]. arXiv. https://doi.org/10.48550/arXiv.2405.02957

[3] Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior [arXiv preprint]. arXiv. https://doi.org/10.48550/arXiv.2304.03442

[4] Schmidgall, S., Su, Y., Wang, Z., Sun, X., Wu, J., Yu, X., Liu, J., Liu, Z., & Barsoum, E. (2025). Agent Laboratory: Using LLM agents as research assistants [arXiv preprint]. arXiv. https://doi.org/10.48550/arXiv.2501.04227

[5] Wiesinger, J., Marlow, P., & Vuskovic, V. (2025). Agents [White paper]. Google. https://ppc.land/content/files/2025/01/Newwhitepaper_Agents2.pdf

Video Tutorial

[6] Herk, N. [@nateherk]. Nate Herk [YouTube channel]. YouTube. https://www.youtube.com/@nateherk

[7] AI Foundations [@ai-foundations]. AI Foundations [YouTube channel]. YouTube. https://www.youtube.com/@ai-foundations

[8] **Ng, A.** (2024, February 20). *Andrew Ng on AI Agentic* Workflows and Their Potential for Driving AI Progress [Video]. YouTube. https://www.youtube.com/watch?v=q1XFm21I-VQ