Immigration AI Advisor: Personalized Immigration Assistance for International Students

Matt Borkowski, Miles ZhengFerrari,

Nuvorish Paul, Raj Rahul Parekh, Santhosh Charudatta

Seidenberg School of Computer Science and Information Systems

Pace University, New York, NY, USA

[1]borderlessai@googlegroups.com

Abstract— **The Immigration AI Advisor is an advanced application using artificial intelligence (AI) to assist international students in navigating complex US immigration related questions and managing critical documentation such as I-20 form and other immigration forms in the format of PDF. The system automates the extraction, and processing of structured data, offering high quality personalized advice while ensuring data integrity through blockchain-based techniques. By incorporating very up-to-date technologies, including Sentence Transformers, vector databases, and rank similarity for semantic search, the AI Advisor bridges the gap between technical innovation and practical application. Furthermore, when using technologies such as TensorFlow and PyTorch etc to enhance statistical and analytical capabilities, while the front-end, developed using Python, Flask, FastAPI, ReactJS, HTML, and CSS, provides a dynamic and intuitive user interface. This paper contains the methodologies, technological integration, and the potential of the Immigration AI Advisor to improve and streamline administrative tasks in international student offices, with the ultimate goal of improving the student experience.**

Keywords— **Immigration Management, Artificial Intelligence, Blockchain, Semantic Search, TensorFlow, Education.**

I. INTRODUCTION

International students significantly add to the cultural and academic expertise of educational institutions in the United States, bringing diverse perspectives and experiences. However, they often encounter considerable red-tape and issues in understanding immigration policies and complying with documentation requirements of USCIS. Among the various documents, the I-20 form stands out as a fundamental document for maintaining legal student F-1 or J-1 status in the United States. Its complexity and with the necessity for precise compliance, can be overwhelming and very hard to understand for international students navigating the very unfamiliar US immigration systems.

The challenges faced by international students are further enlarged by the limited resources of international student offices, which are often understaffed and overwhelmed by the volume of inquiries with a very limited operating budget. These offices struggle to provide timely support to students, especially during peak periods such as enrollment times, leading to delays, inefficiencies, and student anxiety which can further trigger mental health aspects of issues. Traditional administrative tasks relying on manual data handling are not only labor-intensive but also prone to human error during processing and advising, further complicating the process issues.

To address these issues, the Immigration AI Advisor was developed. This system automates repetitive tasks such as document parsing and immigration questions, providing a scalable and efficient solution for managing immigration-related questions, tasks and processes. By leveraging advanced AI frameworks, blockchain technology to for secure data management, and with a robust front-end design, the Advisor ensures that students receive accurate, personalized, and timely immigration and visa status related assistance. The integration of technological products such as TensorFlow, PyTorch, and large language models (LLMs) will enhance the system's capabilities, making it a disruptive and transformative important tool in the field of immigration management in the United States.

II. LITERATURE REVIEW

The integration of artificial intelligence or AI and blockchain technologies into administrative systems has demonstrated applicable potential in addressing challenges related to data integrity, accuracy and efficiency. Shinde and Gurrala [1] explore the application of blockchain in forensic environments, emphasizing its role in securing sensitive data and maintaining integrity. That is applicable in the field of immigration and technology. Their work highlights the relevance of blockchain-based solutions for managing immigration-related information, which often involves highly confidential and sensitive user data.

From Maruti and Rao [4] 's discussion of the importance of preserving evidence integrity in cloud-based systems using blockchain. They purposely provided enhancements that align with the Advisor's approach to data security. By incorporating blockchain, the Immigration AI Advisor ensures that extracted data from I-20 forms remains tamper-proof and verifiable, providing an additional layer of trustworthiness to its operations. Those two studies addressed the security issue of our project.

In the edTech domain, Chen, Chen, and Lin [2] highlight the transformative role of AI in enhancing personalized learning tools in the field of educaion. These advancements form the foundation for adaptive systems like the AI Advisor, which tailors its responses to individual student needs. Ahmed et al. [3] extend this type of studies by exploring the potential of generative AI in education, focusing on mapping student perceptions and opportunities for AI-driven improvements. These ideas are important fundations for integrating LLMs into the AI Advisor's AI knowledge base, enabling adaptive and personalized interactions for international students.

McCall et al. [5] emphasize the importance of human-machine teaming work, proposing structured frameworks to enhance collaboration and enlargement. This principle is embedded in the AI Advisor's design, which lowers the load of human effort by automating routine tasks, allowing staff to focus on complex cases that require a more human-like approach. Chaudhry and Kazim [6] address ethical considerations in AI adoption, advocating for fairness, transparency, and compliance. Those are the values that guide the AI Advisor's development and deployment.

Finally, Wangoo and Reddy [8] explore IoT-enabled smart learning environments, illustrating how interconnected systems in today's society, especially in the developed world, can revolutionize learning experiences. Their findings provide a vision for future versions and functionalities of the AI Advisor, where IoT integration could expand its capabilities to create a more holistic immigration management ecosystem for the complex US immigration landscape.

III. Relevance

International student offices are the last-mile provider in guiding students through immigration processes in the US, yet traditional workflows often fall short of meeting large demands of the student body. Privacy and security concerns, coupled with inefficiencies in manual data handling have highlighted the need for new solutions that are based on EdTech and AI. Blockchain technology, as demonstrated by Ramamoorthy and Ragu [7], offers a robust architecture for protecting sensitive data. Its integration into the AI Advisor eventually ensures that all extracted and processed data is secured, reducing the risk of breaches and unauthorized access. This is particularly critical and important in the context of immigration to the USA, where the mishandling of information can have severe legal consequences.

The AI Advisor's technical stack addresses these great challenges by combining AI and advanced programming. Sentence Transformers enable high-dimensional text embeddings, which convert user queries and document content into numerical vectors stored in Pinecone databases. This facilitates efficient retrieval and analysis, while cosine similarity ranks query-document matches based on semantic relevance. That technology works in tandem to deliver precise, context-aware responses, ensuring that students will get accurate assistance which is imperative in Immigration Advising.

Additionally, TensorFlow and PyTorch and similar technologies enhance the AI Advisor's capabilities by enabling advanced statistical processing and data transformation. These frameworks complement the system's use of LLMs which provide the foundation for a custom-trained immigration-specific closed knowledge base. By integrating programming frameworks such as Python, Flask, FastAPI, ReactJS, HTML, and CSS, the AI Advisor delivers a user-friendly interactive interface, further enhancing its accessibility, usability and quality.

IV. Methodology

The Immigration AI Advisor is a comprehensive system designed to tackle the challenges of immigration management by automating tasks such as document parsing and reading, questions, and real-time immigration advice generation. This system integrates cutting-edge technologies in document parsing, search, AI, and user interface design, ensuring efficiency, security, and safety of its users. Each component is designed to interact with others, creating a unified system capable of handling difficult and complex user requirements in the domain of immigration information.

A. Document Parsing

The document parsing component of the Immigration AI Advisor focuses on extracting critical and accurate information from I-20s, which are central to the immigration process for international students in the USA. To achieve this, the system uses advanced regular expressions and pattern matching algorithms. These tools are able to process highly structured data, such as SEVIS ID, program start and end dates, and personal information. The extraction process starts with the conversion of the I-20 form into a machine-readable format, typically in plain text. Once the data is in a readable format, regular expressions and calculations are applied to identify specific fields that match predefined patterns, such as numeric SEVIS IDs or date formats, BoD, program information etc. After the data is extracted, it undergoes a process to ensure that the formats are correct, and any errors, such as missing fields, are flagged for further review.

In addition to regular expressions, the AI Advisor in the future will integrate blockchain technology to secure the extracted data. This ensures the integrity and immutability of the processed information, providing a secure record that can be audited if necessary.

B. Semantic Search

Semantic search is a core feature of the Immigration AI Advisor, enabling users to receive very accurate responses to their immigration related questions. Traditional keyword based database search often fails to capture the personalized and meaning behind user's input. It lacks the human aspect of Immigration Advising. To address this, the Advisor employs Sentence Transformers to convert textual data into high-dimensional vectors. These vectors represent the semantic meaning of the text, capturing contextual nuances beyond surface-level keywords that are foundational to our technology used in the project.

Once the textual data is embedded into vectors, the embeddings are stored in Pinecone, a high-performance vector database optimized for fast retrieval and fast indexing. The system then uses cosine similarity to measure the angular distance between the vectors of a user's query and the vectors of document contents. The closer the vectors are to one another, the more relevant the document is to the query. We rank the result by their relevance. This allows the system to rank and generate results that are relevant to the conversation, even if the exact keywords from the query do not appear in the document or conversation.

C. AI Integration

The AI integration within the Immigration AI Advisor builds upon several foundational machine learning frameworks, including TensorFlow and PyTorch, as well as large language models (LLMs). It is noted that we have yet to implement TensorFlow and PyTorch. These technologies work together to provide a robust backend which is capable of processing complex data and generating sophisticated responses. The LLMs, in particular, form the basis of the AI Advisor's knowledge base, which is a custom-trained closed resource to focus on immigration advice and informational content and terms. This enables the system to generate responses that are not only accurate but also contextually human like just as if the students are talking to an advisor from the school.

TensorFlow and PyTorch will be utilized to support advanced data transformation and statistical analysis. It allows the system to process large volumes of data very accurately. These frameworks use the LLMs by enabling them to handle quantitative data, such as calculating visa expiration timelines or determining compliance with specific immigration rules and regulations. Additionally, these frameworks provide the ability to fine-tune the models, ensuring that the responses generated by the system continue to improve over time as more data is processed, and when additional immigration policies are incorporated into the knowledge base.

D. User Interface Design

The user interface (UI) of the Immigration AI Advisor is designed with user experience in mind, utilizing a combination of Python, Flask, FastAPI, ReactJS, HTML, and CSS. The goal is to provide an intuitive, dynamic, and responsive interface that is human mincing and accessible to both students and international student office staff. The front-end features real-time updates, human-like animations, and will incorporate secure login, and personalized dashboards in the future that display the most relevant documents and information based on the user's needs. Out goal is for international students to be able to easily interact with the system, upload immigration-related documents, and receive personalized advice based on the analysis of those documents.

Security is a critical consideration for the UI, particularly in ensuring that sensitive student data is protected. Just like most US HigherEd institutions, The AI Advisor will employ multi-factor authentication (MFA) and encrypted data storage to ensure security. We hope eventually the interface also allows for error handling and user guidance etc, which are important for first-time international students users who are unfamiliar with the system.

E. Security and Scalability

Security and scalability are essential components of the Immigration AI Advisor's design thoughts. The system will incorporate blockchain technology to secure all data processed within the system. This includes not only the information extracted from immigration forms but also the user questions and responses generated by the AI. Blockchain ensures that all interactions are logged in and not in an editable format, guaranteeing data integrity and providing transparency in the tool's operations.

Scalability is another important aspect of consideration and design, given the increasing volume of international students and the growing complexity of immigration policies in the United States especially for the incoming White House administration. The Advisor's modular architecture supports the addition of new policies and data sources without disrupting existing functionality. The back-end is designed to handle a large number of student users, ensuring that the system remains responsive even during periods of high demand and high volume. Moreover, the use of Pinecone for vector storage allows the system to efficiently scale as more data is added, providing a high level of immigration advising performance does not degrade as the system expands.

V. RESULTS

The Immigration AI Advisor is currently ending in Sprint 2, with significant progress achieved in both front-end and back-end development. The front-end interface, developed using Python, Flask, FastAPI, ReactJS, HTML, and CSS, has been successfully implemented to provide a dynamic and user-friendly human-like experience. The system supports seamless document uploads. A key feature of the interface is

the real-time updates on document processed, which provides immediate feedback to users as their questions are being analyzed.

On the back-end, we have successfully implemented the core AI functionalities, particularly the document parsing and semantic search components. The integration of Sentence Transformers for converting document content and user queries into 768-dimensional vectors is functioning as expected. These vectors are stored in a scalable vector database, similarity is being used to compute the relevance of documents in response to user queries. The initial tests have shown promising results which indicate the ranking of questions with the system able to retrieve relevant documents based on the semantic meaning of user input, this improves the accuracy of search results compared to traditional keyword-based databases.

The AI processing capabilities are still in their very early stages, with future improvements planned to refine the model's understanding of user questions. We have begun using Sentence Transformers for data transformation and tokenized the text data into numerical values, which will allow for more sophisticated processing of complex data sets, such as immigration policy and regulations analysis and a real-time response generation. For the next Sprint, the integration of large language models (LLMs) is expected to further enhance the system's ability to handle natural language queries, offering users personalized advice that is both accurate and consistent with the latest immigration regulations piped down to individual institutions of the United States.

Despite these advancements so far, the AI model is still being built and improved. As we move forward with Sprint 3, additional focus will be placed on improving the semantic search algorithm to account for more intricate questions and incorporating more data sources for better responses. We are hoping integrating advanced models for policy analysis, potentially using TensorFlow for predictive capabilities, and refining the back-end AI models will be key tasks for the next phase. We also plan to increase the number of the dataset, adding more document types and expanding the range of US immigration rules, policies and reguations the system can address.

Another major milestone in Sprint 2 has been the system's ability to scale. As the volume of documents increases, the architecture of the database and AI models is designed to support an increasing number of student user clients without compromising response time or accuracy. The vector database and when we use blockchain-backed systems have been designed to handle large-scale data processing efficiently, ensuring the system will grow as more users employ the system.

In addition to the core features, a detailed evaluation of the UI design was incorporated into our sprints. The ReactJS-based front-end provides a responsive and interactive interface, it is human-like and we think students will find it very easy and familiar to use. International students appreciate the real-time feedback and the customized experience provided by the system, especially for document upload and answer of questions. However, further refinement is very necessary to improve responses and provide additional features to accurately answer and help students with immigration related advice.

VI. Conclusion

The Immigration AI Advisor is a significant progress in automating the immigration management process for international students. It is also disruptive to the industry. By using AI technologies such as Sentence Transformers and use vector storage, and possibly blockchain for data integrity, the system provides efficient, context-aware, advisor-like responses to international student user clients. We hope the next step, the integration of TensorFlow and PyTorch further enhances the system's ability to process complex data and deliver personalized immigration advice and responses.

Looking ahead, the AI Advisor will focus on expanding its knowledge base, incorporating additional immigration documents, rules and regulations etc, and refining its AI models to improve accuracy. The system's potential extends beyond academic institutions, offering a scalable solution for immigration management in all countries. With further advancements, the Immigration AI Advisor could significantly reduce the administrative burden on student offices while empowering international students with the tools to help them manage their immigration journey in the United States.

References

[1] S. M. Shinde and V. R. Gurrala, *"Securing trustworthy evidence for robust forensic cloud-blockchain environment for immigration management with improved ECC encryption,"* Expert Systems with Applications, vol. 229, p. 120478, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423009806

[2] L. Chen, S. Chen, and X. Lin, *"Artificial Intelligence in Education: A Review,"* IEEE Access, vol. 8, pp. 75264–75278, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9069875.

[3] Z. Ahmed et al., *"The Generative AI Landscape in Education: Mapping the Terrain of Opportunities, Challenges, and Student Perception,"* IEEE Access, vol. 12, pp. 147023–147050, Sep. 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10681094.

[4] S. S. Maruti and G. V. Rao, *"Preserving Integrity of Evidence with Blockchain Technology in Cloud Forensics for Immigration Management,"* Journal of Telecommunications and Information Technology, no. 1, pp. 78–87, 2023. [Online]. Available: https://jtit.pl/jtit/article/view/211.

[5] F. McCall et al., *"Towards a Systematic Educational Framework for Human-Machine Teaming,"* in Proc. 2021 IEEE Int. Conf. Artificial Intell. Educ. (ICAIE), 2021, pp. 123–128. [Online]. Available: https://ieeexplore.ieee.org/document/9678853.

[6] M. A. Chaudhry and E. Kazim, *"Artificial Intelligence in Education (AIEd): A High-Level Academic and Industry Note 2021,"* AI and Ethics, vol. 2, pp. 157–165, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s43681-021-00074-z.

[7] S. Ramamoorthy and G. Ragu, *"A blockchain-based cloud forensics architecture for privacy leakage prediction with cloud,"* Healthcare Analytics, vol. 4, Dec. 2023, Art. no. 100220. [Online]. Available: https://doi.org/10.1016/j.health.2023.100220.

[8]   D. P. Wangoo and S. R. N. Reddy, "*Smart Learning Environments Framework for Educational Applications in IoT,*" in Proc. 2020 IEEE Int. Conf. Comput., Power Commun. Technol. (GUCON), 2020, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9342150.

[9]   *TensorFlow Documentation*. Available: https://www.tensorflow.org.