***Internship Final Report***

**Introduction:**

I, Rithvika Tiruveedhula, am a dedicated Computer Science major at Vellore Institute of Technology, Chennai (VIT). This internship, an integral part of my academic journey, provided me with a unique opportunity to apply theoretical knowledge in a real-world setting, fostering a bridge between academia and industry.

**Objectives:**

The overarching aim of this internship was to deepen my expertise in Artificial Intelligence (AI) and Machine Learning (ML). Beyond this, I sought to gain practical insights into the nuanced fields of Natural Language Processing (NLP) and Generative AI. My objectives included not only skill acquisition but also understanding industry practices and the application of theoretical concepts in solving real-world challenges.

**Internship Activities:**

List of all the activities done during the internship:

**1. PII Recognizer and Named Entity Recognition (NER) in NLP:**

* Project Focus:  The project centered around Named Entity Recognition (NER) within the context of Personal Information Identity (PII).
* Learning Experiences:  Engaging in the PII Recognizer project provided substantial learning experiences, bridging theoretical knowledge with practical applications.
* Initial Challenges:  Initial attempts with Presidio Analyzer highlighted challenges in accurately recognizing entities, particularly in the NER phase.
* Strategic Transition: To address entity recognition issues, a strategic decision was made to transition from Presidio Analyzer to Spacy models.
* Organization Recognizer Challenges: Specific challenges arose in the Organization Recognizer component, leading to further exploration and optimization.

**2. UI Creation and API Testing:**

* Comprehensive API Testing: Rigorous testing was conducted, complemented by the development of a User Interface (UI) using Flask for seamless integration.
* Tool Adoption:  Postman was adopted to facilitate comprehensive API testing, uncovering persistent bugs in the process.
* Code Optimization:  Extensive code optimization was undertaken to address the persistent bugs and ensure optimal functionality.
* Error Handling and Logging:  Robust error handling and logging mechanisms were implemented to enhance error understanding and facilitate more efficient debugging.
* Multifaceted Approach:  The multifaceted approach to UI creation and API testing highlighted the intricate aspects of software development and quality assurance.

**3. Generative AI and RAG-based Model:**

* Exploration of Generative AI:  The journey involved the utilization of llama index and transformers through the Hugging Face model.
* Culmination in RAG-based Model:  Efforts culminated in the development of a RAG-based AI model showcasing proficiency in handling multiple file uploads and simultaneous questioning.
* Positive Feedback:  The project presentation to Avinash Sir not only solidified understanding but also garnered positive feedback, highlighting the success and innovation achieved.
* Practical Applications:  The developed RAG-based model showcased practical applications of Generative AI in real-world scenarios.

**4. Audio Detection and eDiscovery Tasks:**

Venturing into audio detection, I explored Speech to Text capabilities using the speech\_recognition library. Additionally, I contributed to the eDiscovery team by employing Optical Character Recognition (OCR) for text recognition in images. Undertaking tasks involving numpy for data manipulation based on Excel sheet data further diversified my skill set, encompassing both audio processing and data manipulation. These experiences underscored the interdisciplinary nature of AI applications and the need for versatile skill sets.

**Challenges Faced and Solutions Implemented:**

Challenges encountered during the PII Recognizer project, particularly entity recognition issues, were successfully mitigated through the incorporation of a deny list and transitioning to Spacy models. In the domain of API testing, persistent bugs were addressed through meticulous code optimization, showcasing the importance of thorough testing and optimization in software development. These challenges not only tested technical skills but also underscored the importance of adaptability and problem-solving in real-world project scenarios.

**Skills and Knowledge Acquired:**

The internship has been instrumental in equipping me with a diverse set of technical skills. Proficiency in NLP, API testing, UI creation with Flask, Generative AI using Transformers, audio detection using speech\_recognition, OCR for text recognition, and data manipulation using numpy are among the many skills acquired. Exposure to llama index and transformers through the Hugging Face model significantly expanded my understanding of AI applications, providing a comprehensive foundation for future endeavors.

**Achievements and Contributions:**

A significant achievement during the internship was the successful completion of the PII Recognition project. Presenting the fully built UI to Avinash Sir during a comprehensive demo not only showcased the project's practical applicability but also demonstrated effective communication skills. The RAG-based AI model, designed to handle multiple files and simultaneous questions, received commendation during the presentation to Avinash Sir, highlighting the success and innovation achieved. These accomplishments not only validated the theoretical knowledge but also emphasized the practical implications of AI in real-world scenarios.

**Challenges Faced and Solutions Implemented:**

**PII Recognizer Challenges:**

**1. Entity Recognition Issues:**

During the development of the PII Recognizer, a significant challenge emerged in accurately identifying entities during the Named Entity Recognition (NER) phase. This posed a potential risk to the precision and reliability of the entire recognition process.

*Solution -* Transition to Spacy Models:

To overcome the NER challenges, a strategic decision was made to transition from utilizing Presidio Analyzer to Spacy models. This transition aimed to harness the advanced capabilities of Spacy for entity recognition, ensuring a more precise and effective identification of entities within the PII Recognizer.

**2. Organization Recognizer Entities:**

Another obstacle encountered was related to the recognition of entities within the Organization Recognizer. Difficulties arose in ensuring the accurate identification of entities, which raised concerns about the overall effectiveness of the Organization Recognizer.

*Solution -* Deny List and Alternative Libraries:

In response to the challenges faced with the Organization Recognizer entities, a multi-faceted approach was adopted. Firstly, a deny list was implemented to enhance the accuracy of entity recognition. Additionally, alternative libraries, such as Spacy, were explored as viable alternatives to address and further improve entity recognition challenges within the Organization Recognizer.

**UI Creation and API Testing Challenges:**

**1. Persistent Bugs during Testing:**

The development of the User Interface (UI) and subsequent API testing encountered persistent bug issues. These bugs posed potential obstacles to the seamless functionality of the UI and the reliability of the API, introducing uncertainties in the testing phase.

*Solution -* Code Optimization:

To mitigate the persistent bugs identified during testing, a comprehensive strategy of code optimization was implemented. This involved a meticulous review and refinement of the codebase to address underlying issues, ensuring that the UI and API would function optimally and reliably.

**2. Enhanced Error Handling:**

Recognizing the importance of robust error management, an additional layer of enhanced error handling and logging mechanisms was implemented. This proactive approach aimed to provide a clearer understanding of errors, facilitating more efficient debugging processes during testing and deployment.

**Learning from Challenges:**

The challenges encountered throughout the development process proved to be invaluable learning experiences. They highlighted the dynamic nature of real-world projects and emphasized the significance of adaptability and problem-solving skills. The iterative nature of addressing challenges, transitioning between tools and strategies, showcased the importance of flexibility in navigating complex project scenarios.

**Reflection:**

Reflecting on the overall internship experience, I recognize the immense value gained in terms of practical knowledge, problem-solving skills, and adaptability. Exposure to various AI applications has not only complemented my academic coursework but has also provided valuable insights into potential career paths within the field. The hands-on experience and mentorship received have significantly influenced my academic and professional development, shaping my aspirations in AI-related domains.

**Recommendations:**

Building on the positive aspects of the internship, I recommend the implementation of a more structured task assignment and management system. Timely task assignment and clear communication from a designated task giver and manager would contribute to a more organized workflow for interns. Additionally, expanding mentorship opportunities for interns would further enhance their professional growth, fostering a collaborative and supportive learning environment.

**Conclusion:**

In conclusion, I express my sincere gratitude for the enriching opportunity to contribute to the organization's projects. The guidance from mentors, including Aditya Sir and Avinash Sir, has been instrumental in my professional development. The internship has not only equipped me with practical skills but has also shaped my aspirations, providing a strong foundation for future academic and professional endeavors.

Thank you to everyone at Iota Analytics for this enriching and transformative internship experience.

 I have also attached some sample work done for the eDiscovery team.

Text detection and Image Rotation: [[](https://iotaanalyticscom-my.sharepoint.com/:u:/g/personal/rithvika_tiruveedhula_iotaanalytics_com/EfKh7JpRyNdLsYN4qNKy7O0BcIrPmoFG4TDj3qjeQheufQ)image\_text\_detection 1.py](https://iotaanalyticscom-my.sharepoint.com/:u:/g/personal/rithvika_tiruveedhula_iotaanalytics_com/EfKh7JpRyNdLsYN4qNKy7O0BcIrPmoFG4TDj3qjeQheufQ)

Creation of Separate CSV files based on the Excel file given:  ​[[](https://iotaanalyticscom-my.sharepoint.com/:u:/g/personal/rithvika_tiruveedhula_iotaanalytics_com/EcjYf-uPWqlEt95J8DX7cuUBFvHyWFAJOS46F7iorI7QJg)docs\_creation 1.py](https://iotaanalyticscom-my.sharepoint.com/:u:/g/personal/rithvika_tiruveedhula_iotaanalytics_com/EcjYf-uPWqlEt95J8DX7cuUBFvHyWFAJOS46F7iorI7QJg)