Ontology, Robotic Process Automation, Automatic Speech Recognition and their integration with Large Language Models

Zuber Purahoo *2421459*, Tanvee Proag *2422814* and Oudayrao Ittoo *2420501*
under the supervision of Dr. Baby Ashwin Gobin-Rahimbux

SIIS 6118 Artificial Intelligence Applications Assignment '24
Faculty of Information, Communication and Digital Technologies
University of Mauritius

Abstract

Ontology originates from philosophy, referring to a theory of existence (e.g., Aristotle's ontology with substance and quality as categories explains all that exists). In computer science, ontology designs and models knowledge about a domain, real or imagined. Robotic Process Automation (RPA) uses software robots to automate repetitive business processes, increasing efficiency and reducing human error. Automatic Speech Recognition (ASR) converts spoken language into written text, enabling voice-activated systems and hands-free interaction with devices. A Large Language Model (LLM) is an artificial intelligence model trained on vast amounts of text data, enabling it to generate human-like text, understand context and perform tasks. This paper investigates and critically evaluates ontology, RPA, ASR, and their connection to LLM, offering a brief analysis of their application in the year 2024.

Ontology

An ontology is like a blueprint for knowledge. It is a formal representation of a domain of interest. It defines the concepts, properties, and relationships that govern a specific domain. In simple terms, imagine building a house. To ensure everyone understands the plan, a blueprint is created. This blueprint outlines the layout of each room, their interconnections, and the specific components of each room.

An example of a domain of interest is the field of agriculture. Suppose data about crops, pests, diseases, weather conditions, and farming practices are stored in a database. A farmer might want to know the specific diseases affecting their crops based on their characteristics. The farmer would want an intelligent system to inform them about that specific disease and how to treat it.

Ontologies are generalized data models, meaning they only model general types of things that share certain properties but do not include information about specific individuals in the domain. For example, instead of describing a specific cat, Simba, and all his individual characteristics, an ontology focuses on the general characteristics of cats, trying to capture the characteristics that most cats have.

To model real-world information, a knowledge graph is used. Using the ontology as a framework, real data about individual cats, owners, vets, and shops can be added to create a knowledge graph [Schrader, 2022]. In other words,

ontology + data = knowledge graph

To further enhance their understanding and reasoning abilities, Large Language Models (LLMs) can be integrated with knowledge graphs which are based on ontologies which are going to be elaborated in further sections.

Foundations of ontology

As the web grew in popularity, the need for a common vocabulary to describe and share information became apparent. Ontologies emerged as a solution, providing a structured framework for representing knowledge in a machine-readable format. The Semantic Web initiative, led by Tim Berners-Lee, aimed to create a web of data that could be understood and processed by computers. Ontologies played a crucial role in this vision, enabling machines to reason about and integrate information from diverse sources [Berners-Lee].

Today, ontologies are widely used in various domains, including biology, medicine, social sciences, and engineering. They are essential for tasks such as information integration, knowledge management, and semantic search. The integration of ontologies with LLMs has opened up new possibilities for natural language understanding and generation, further expanding the scope and impact of ontological approaches.

Popular languages used to write ontologies are Web Ontology Language (OWL) and Resource Description Framework (RDF). OWL uses RDF syntax but it extends it with additional features.

Key features of OWL:

- **Classes** OWL allows you to define classes, which represent groups of individuals with common properties.
- **Properties** Relationships between individuals (e.g John hasFather Paul).
- **Individuals** Individuals are instances of classes.
- **Reasoning** OWL supports various reasoning services, such as classification (determining the subclass relationships between classes), consistency checking (ensuring that an ontology is logically consistent), and instance retrieval (finding individuals that satisfy certain criteria).

Introduction to LLMs

An LLM is a model trained to understand human sentence structure and meaning. The model can understand text inputs and generate outputs that conform to correct grammar and language [Team, 2024].

LLMs represent text as vectors, known as embeddings. These embeddings are stored in an index within a vector database. Let's break this down into digestible chunks.

Embeddings are a way to store data of all types (including images, audio files, text, documents, etc.) in number arrays called vectors.

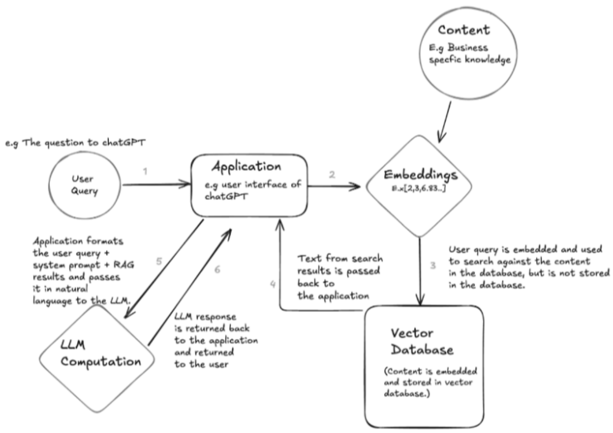

Figure 1: LLM integration with vector databases

The sentence "Artificial Intelligence course at the University of Mauritius is worth it" can be represented as an embedding [0.00233 0.86723, 0.00002 ... -1.232], with "..." indicating hundreds to thousands of other numbers. The specific meaning of these numbers is known only to the transformer model (embedding algorithm) that generated them. These numbers represent words, their context (relationship to other words), and their meaning. The embedding contains all the information the Retrieval Algorithm requires to search users' queries and find relevant information to provide answers.

Vector embeddings

Vectors serve as a method to organize embeddings in a meaningful manner. An example of an embedding was provided, which may appear as one-dimensional data, but in reality, these embeddings possess multiple dimensions. Humans face difficulties in visualizing more than three dimensions, yet the number of dimensions in a vector is determined by its length, denoted as n. For instance, if an embedding consists of 100 numbers, the corresponding vector has 100 dimensions. Embeddings are vectors stored within an index in a vector database. The term "index" refers to the order in which these vectors are stored. The origin of the vector (its starting point), its direction, and its magnitude (length) collectively determine the vector's relationship with other vectors. This relationship facilitates the retrieval algorithm to compare two vectors and assess their similarity. For example, the embedded vector representing "eat" might be stored in close proximity to the vector representing "food," and their distance within the index reflects their semantic relationship. When the algorithm queries the vector database for "eat," it may also return "food" due to their proximity [Besen, 2024].

Vector database

Traditional databases that store information in rows and columns look for the row that matched the specific search. Unlike traditional databases, a vector database indexes and stores vector embeddings for fast retrieval and similarity search. Vector databases use an approximation function to look for the most similar vector to the query.

Improving LLMs

Under the hood, LLMs are neural networks and their performance is measured by how many parameters they contain. An LLM's parameters essentially represent the general patterns of how humans use words to form sentences. That deep understanding enables LLMs to respond to general prompts quickly. However, it does not serve users who want to dive deeper into specific topics.

In 2020, a group of researchers brought up a technique called Retrieval-Augmented Generation (RAG) which enhances the accuracy and reliability of LLMs by fetching data from external sources. It allows models to cite sources of their response and also reduces the possibility of hallucination [Merritt, 2024]. With RAG, users can now have conversations with data repositories or even business specific knowledge.

Integration with knowledge graphs

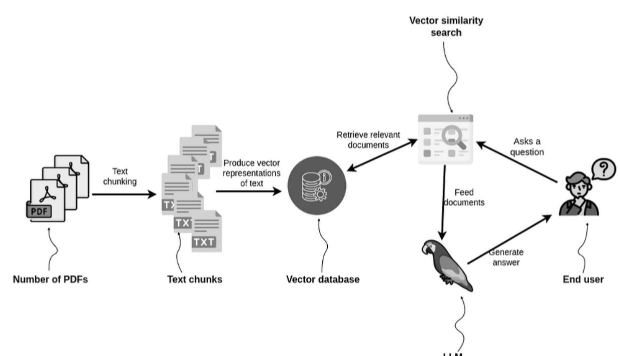

Figure 2: Integration with knowledge graphs

Now suppose a tool like Chat with Your PDF is being developed, which searches for information in a provided document. Most of these tools use vector similarity search to identify the chunks of text that contain data similar to a user's question as shown in Figure 1. Normally, the RAG might for example return three similar documents to provide context to the LLM, which enhances its ability to generate accurate answers. This approach works well when the vector search can identify relevant chunks of text. However a simple vector similarity search is not sufficient if the user has a multi-hop question as the LLM needs information from multiple documents or multiple chunks of text to generate an answer.

For instance, consider the following query: **Did any of the former OpenAI employees start their own company?**

This question is multi-part in that it contains two questions: *Who are the former employees of OpenAI?* and *Did any of them start their own company?*

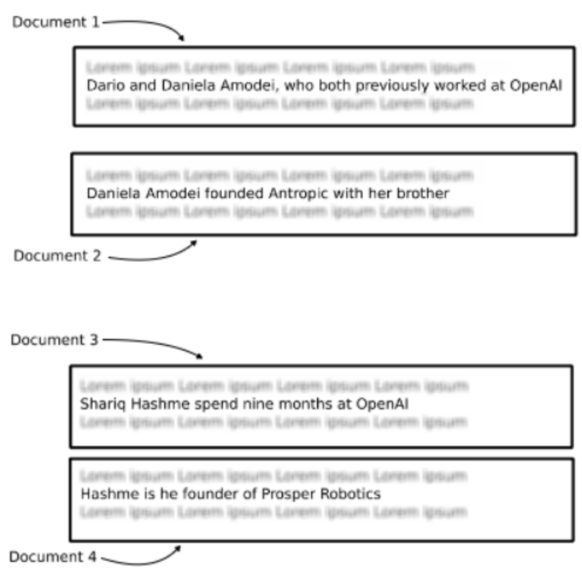

Figure 3: Information spanning multiple documents.

Answering a multi-hop question requires breaking it down into several sub questions and retrieving multiple documents to generate an accurate answer. The reason why a simple vector similarity search will not work correctly is that all the documents contain part answers and the LLM might completely ignore relevant facts from other documents as illustrated in the diagram above. Various strategies can be employed to answer these multi-part questions and one of them is using knowledge graphs as condensed storage information.

The knowledge graph (as explained in the ontology section) uses nodes and relationships to represent data as shown above. Document 1 indicates that Dario and Daniela previously worked at OpenAI, while Document 2 shows that they founded Antropic. Each record was processed separately by the LLM but the knowledge graph representation connects the data making it easier to answer questions that span multiple documents [Bratanič, 2024].

Using knowledge graphs to improve RAG applications

Contextual Understanding By processing each document separately and connecting them in a knowledge graph, we can construct structured representation of the information where entities like people, places or concepts are connected by relationships.

Semantic search Instead of relying solely on keyword matching, a retrieval system can leverage the semantic relationships in the graph. For e.g. if the user searches for "Jaguar", a simple keyword-based search might return results about the animal, the car brand or even the software company (Jaguar). With a knowledge graph, the search system might link "Jaguar" to: 1. An entity representing the animal, related to topics like wildlife. 2. An entity representing the car brand, linked to car specifications. 3. Any other entity using the term Jaguar.

Structured Data Integration By using unstructured and structured data, the RAG model can pull structured data from the knowledge graph and combine it with insights from unstructured text, leading to more informative responses.

Business-Specific Knowledge Graphs Enterprises can provide their business-specific knowledge to the RAG model which in turn enhance the quality of the generated responses.

Knowledge-graph implementation

Developing and storing a knowledge graph involves several steps, including data collection, integration, structuring, storage and management. Data Collection: Use of web scraping, APIs or other methods to gather data from databases, spreadsheets, text documents, web pages and research papers.

Data Integration Ensure that data formats are standardized and the data from various sources are aligned with a common schema or ontology.

Data Structuring Design the ontology. Define the types of entities (classes) and their relationships (e.g. "is a", "part of") within the knowledge graph. Create a schema that outlines the types of data and relationships in the graph and add metadata and annotations to enhance the context and meaning of the data.

Data Storage Use specialized databases (e.g. Amazon Neptune, Neo4j) designed to store and query graph data. Data can be stored using RDF (Resource Description Framework) format and SPARQL for querying.

Data Maintenance Data should be regularly updated and cleaned. New data sources may also be added to expand the graph.

Conclusion and future potentials

The combination of ontologies and LLMs offer significant potential to address complex challenges in various domains. However creating and maintaining large-scale ontologies can be complex and require domain expertise and specialized tools. The use of personal data in ontologies and LLMs also raises privacy concerns. Future research and development efforts should focus on developing tools and techniques to address these challenges and ensure the ethical and responsible use of these technologies.

Robotic Process Automation

Robotic Process Automation is the mimicking of human behavior by software. Robotic Process Automation (RPA) is a technology that helps automate repetitive tasks usually done by people thus making work more efficient in various industries. RPA uses software robots to perform tasks like data entry, managing files, and processing transactions much faster and more accurately than humans. By taking over these routine jobs, RPA allows employees to focus on more important and creative work that requires human judgment and problem-solving skills, increasing productivity. One of the great things about RPA is that it can be implemented quickly without needing major changes to existing computer systems, making it an attractive option for businesses looking to improve their operations. When RPA is combined with artificial intelligence and machine learning, it becomes even more powerful as these technologies allow bots to learn from data and handle more complex tasks over time. As more companies adopt RPA, they can expect to save money, reduce risks and improve compliance with regulations. The trend of hyperautomation, which combines RPA with other automation tools, is likely to further expand how automation is used in business processes. RPA is an important step toward a more automated and efficient workplace, where technology and human skills can work together to achieve better results.

Figures

Gartner projects global spending on RPA software to reach 2.4 billion by 2022, up from 680 million in 2018. Forrester estimates over 4 million robots will perform office, administrative, and sales tasks by 2021. [Gartner]

The advent of AI and robotic process automation (RPA) has challenged traditional norms, threatening many jobs, reconfiguring economic frameworks, and compelling governments to adopt new strategies. For some governments, AI and RPA have become the new standard, necessitating a balance between tradition and technology. Others remain uncertain about the AI transformation, unaware that their countries could soon face significant changes in trade and global trade networks due to AI's impact.

The economic impact of AI and RPA is undeniable. AI enhances productivity and overall GDP [PWC, 2016]. Nesbitt indicated that AI impacts trade by: (1) enabling supply chains, (2) creating efficiency in compliance software, (3) speeding up and creating better contracts, and (4) improving access to finance. These changes also pose threats to many economies. A recent report from Ball State University showed that in the United States, almost 9 out of 10 jobs were lost to robots and not to trade [Hicks and Devaraj, 2015]. New technologies threaten about 40 percent of jobs in the United States and approximately two-thirds of those in the developing world [Gershon, 2017].

Industries must navigate the challenges brought about by AI. Barriers organizations face include concerns over data protection and privacy, consumer trust and regulatory acceptance, building relevant technologies, managing the volume of unstructured data, optimizing supply chain and production systems, and overcoming potentially high investment costs [PWC, 2016].

Despite these challenges, opportunities abound. Several industries are poised to benefit from AI: Healthcare (data-based diagnostic support), Automotive (autonomous fleets for ride sharing), Financial Services (personalized financial planning), Retail and Consumer (personalized design and production), Technology, Communication, and Entertainment (media archiving and search), Manufacturing (enhanced monitoring and autocorrection), Energy (smart meters), and Transport and Logistics (autonomous trucking) [PWC, 2016]. AI enables productivity enhancement, changes work processes, and can create jobs [Rao, 2017]. The countries projected to have the highest AI gains are China (26 percent GDP boost) and North America (14.5 percent GDP boost), with a total of 70 percent of the estimated $10.7 trillion global economic impact [PWC, 2016]. Expansion in AI-related start-ups is also significant, with about 1,500 AI-related start-ups in the United States in 2016, receiving funding of around $5 billion [Rao, 2017].

With AI's strong influence on industry and governments, an organization's ability to manage and navigate change is critical. AI will set the stage for economic transformation and disruption and will be the foundation for new competitive advantages [PWC, 2016]. Some governments have started to implement strategic measures to gain an advantage in AI. The United Arab Emirates appointed a minister of AI to strategically prepare the country for technological advancements in the field [Wendel, 2017]. Singapore aims to be a pioneer as a smart nation where technology is merged with the way of life of all residents [Vaswani, 2017].

Uses of RPA

RPA can be highly beneficial in various sectors by automating repetitive tasks, enhancing efficiency, and reducing errors.

Machine Learning (ML)

RPA can be integrated with ML to enhance its capabilities. ML algorithms can be used to train RPA bots to perform more complex tasks, such as data analysis and pattern recognition. For instance, in the healthcare sector, RPA can automate the retrieval and storage of patient data, while ML algorithms can help improve healthcare diagnostics by identifying irregularities in medical imaging like X-rays or MRI scans.

Deep Learning (DL)

DL can be combined with RPA to handle tasks that require advanced data analysis and decision-making. For example, in industries with heavy machinery or infrastructure, DL can predict when equipment might fail, and RPA can automatically schedule maintenance or order replacement parts. This integration ensures that complex tasks are handled efficiently and accurately.

Invoice Processing

RPA is particularly effective in automating invoice processing, which involves extracting data from invoices, verifying information, and entering it into accounting systems. RPA bots can constantly check emails for incoming invoices, extract new invoices, and direct them to a database for recognition by document understanding algorithms. This process ensures that invoices are processed quickly and accurately, reducing the need for manual intervention and improving data recognition accuracy.

Customer Support

RPA can significantly enhance customer support by automating repetitive tasks and improving response times. For instance, RPA can create a self-service portal where customers can log in and access their account information, track orders, and find answers to frequently asked questions. RPA can also automate the processing of customer requests, ensuring that customers receive quick responses and solutions to their queries. Additionally, RPA can handle tasks such as sending personalized emails and notifications, further improving the customer experience.

Natural Language Processing (NLP)

NLP can be integrated with RPA to analyze and understand human language, enabling the automation of tasks that involve text-based data. For example, in customer service, NLP can be used to analyze customer reviews and social media comments, providing insights into customer sentiment and preferences. RPA can then integrate with CRM systems and email marketing platforms to trigger personalized email campaigns based on these insights. In document processing, NLP can help extract and structure data from unstructured documents, such as contracts and financial filings, making it easier to automate tasks like contract analysis and financial reporting.

Case Study: UiPath

UiPath, founded in 2005 in Bucharest, Romania is a global software company specializing in RPA solutions, enabling businesses to automate repetitive tasks and enhance operational efficiency through its comprehensive automation platform. RPA involves creating digital software robots to perform repetitive back-office tasks like data entry and invoice processing without altering existing IT systems. UiPath's software automates tasks using AI computer vision, APIs, and OCR engines. The UiPath Automation Platform includes Studio for process creation, Robots for execution, and Orchestrator for deployment and management. The platform also includes AI Center, Action Center, and Apps for low-code application development. This automation saves employees time and allows them to focus on critical tasks, making businesses more efficient and productive.

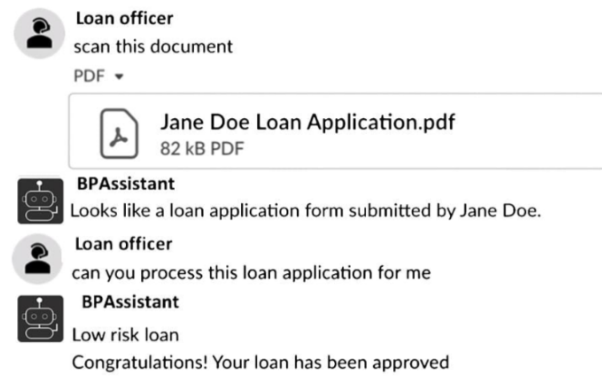

Figure 4: RPA bot in NLP

UiPath has achieved significant success, generating a record quarterly revenue of $405 million, a 31% year-over-year increase [Statista, 2024], and an Annual Recurring Revenue (ARR) of $1.464 billion, up 22% year-over-year [Stock Analysis, 2024]. This growth underscores the transformative impact of UiPath's automation solutions on businesses. Notably, UiPath has helped major companies such as *Microsoft*, *IBM*, *Uber* and *Accenture* by automating repetitive tasks, enhancing efficiency, and driving digital transformation.

RPA with NLP

Rizk et al. [2020] proposed an RPA orchestration solution using a multi-agent framework. RPA bots, considered as agents, share context and interact with users through natural language. The orchestrator, triggered by events, broadcasts to agents, evaluates their responses, and orders the best-scoring agents to execute. This approach allows multiple agents to cooperate on complex tasks. In the loan process, an RPA agent extracts information from a PDF loan application and a business rules RPA bot analyzes the information, determining loan approval or rejection, all through natural language. This approach coordinates RPA bot actions dynamically.

Government use of RPA

New Zealand's tax department is testing a $1.9 billion project using AI to simplify processes for businesses and individuals. The robo-taxman will analyze big data to detect anomalies and noncompliance, comparing third-party data with tax declarations. However, concerns have been raised about potential difficulties in communication and understanding for both users and professionals like accountants and lawyers. [Sonto, 2020]

The AI system validates documents by having users upload images and enroll their face for authentication. It uses machine learning and image functions tailored to each document type. For example, identity cards, birth certificates, driver's licenses, diplomas, and work certificates are all processed differently, with varying information extracted and presented to the user for validation.

The system learns from issues and makes adjustments to improve accuracy and reduce processing time. It can even issue certificates of redemption based on reviewed documents and individual reasons for not completing military service, a task previously done by a section chief. This digitization and automation have significantly reduced processing time from several months to just a couple of days.

Best Practices for RPA in businesses

It is important to manage change to ensure the successful adoption of RPA, as it involves significant changes to business processes and employee roles. The common challenges faced during RPA implementation are resistance to change, lack of understanding, and fear of job loss. The best practices to address these challenges would be

Communication Clear and transparent communication is crucial. Organizations should inform employees about the benefits of RPA and how it will impact their roles.

Training Provide comprehensive training to ensure employees understand how to work with RPA tools and how to adapt to new processes.

Involvement Involve employees in the RPA implementation process to build trust and ownership. This can include forming a cross-functional team to oversee the project.

Change Management Plan Develop a detailed change management plan that includes strategies for addressing resistance and ensuring smooth transition.

Pilot Projects Start with pilot projects to demonstrate the effectiveness of RPA and build confidence among stakeholders.

Continuous Monitoring Regularly monitor the implementation process and make necessary adjustments to ensure that the project stays on track.

Feedback Loop Establish a feedback loop to gather insights from employees and stakeholders, which can help in refining the RPA processes.

Successful RPA implementation requires a holistic approach that includes not only technical aspects but also organizational and cultural changes. By following these best practices, organizations can mitigate risks and ensure that RPA is adopted effectively, leading to improved efficiency and productivity.

Responsibilities of the roles involved in RPA

RPA Manager Oversees the entire RPA project, ensures smooth project delivery, monitors RAG (Red, Amber, Green) status and takes necessary actions.

RPA Business Analyst Clears process-related doubts, coordinates with clients, assesses RPA opportunities and feasibility.

RPA Solution Architect Involved in development if challenges arise, ensures timely development completion and risk callouts, checks code standard implications.

Senior RPA Developer Leads the development team, ensures high-quality code and timely completion.

RPA Developer Builds and tests RPA bots, participates in code reviews and testing.

RPA Tester Conducts unit and integration testing, ensures bots meet the required standards.

RPA Supervisor Oversees the daily operations of RPA bots, manages support and maintenance.

RPA Stakeholders Includes various stakeholders who benefit from RPA implementation, provides input and feedback on RPA processes.

- Project managers ensure smooth project delivery and monitor RAG status.
- Solution architects are involved during development if challenges arise.
- Business analysts clear process-related doubts and coordinate with clients.
- Developers build and test bots, ensuring high-quality code.
- Testers conduct unit and integration testing to ensure bots meet standards.
- Supervisors oversee daily operations and manage support and maintenance.

Challenges

Dealing with the challenges of AI and RPA involves addressing various issues such as technology absorption, job losses, the need for skills updates, security concerns, and ethical considerations. The significant impact of AI and RPA on governments is expected to persist into the future. AI is projected to contribute an estimated $15.7 trillion to the global economy by 2030 and boost local economies' GDP by approximately 26% [PWC, 2016]. Unprecedented technological advancements will lead many organizations into uncharted territories, challenging conventional wisdom. For instance, AI is predicted to drive trucks by 2027, write bestselling novels by 2049, and replace surgeons by 2053 [Gray, 2017].

However, many companies and governments are not adequately prepared for the changes brought about by AI and RPA. Only about 20% of organizations possess the essential skills to succeed with AI technology [Rao, 2017]. An effective strategy can be a game-changer for companies and governments worldwide. Strategic AI approaches that can be helpful to organizations include assessing opportunities and weighing in on technology, competitive pressures, and pain points, determining key priorities, ensuring resources related to talent, culture, and technology are in place and planning for proper governance and control with trust and transparency in mind [PWC, 2016].

RPA with LLMs

The integration of Large Language Models (LLMs) with Robotic Process Automation (RPA) represents a significant leap forward in the automation landscape. This synergy enables RPA bots to perform predefined tasks, understand and generate human language thus revolutionizing various industries in several ways:

Enhanced Customer Support LLM-powered RPA bots can engage with customers in natural, human-like conversations. They can understand customer queries, provide personalized responses, and even handle complex issues by accessing a vast knowledge base. This level of automation can significantly improve customer support services, reducing response times and ensuring consistent quality.

Intelligent Data Analysis In industries like finance and healthcare, data analysis is a crucial task. LLM-powered RPA bots can not only analyze data but also generate insightful reports in a format that is easily understandable by humans. This can save organizations valuable time and resources while enabling data-driven decision-making.

Streamlined Document Processing Many businesses deal with a large volume of documents daily. LLM-powered RPA bots can be trained to extract information from documents, classify them, and even generate summaries or reports based on the content. This can lead to faster document processing and reduced manual errors.

Multilingual Capabilities For global businesses, language barriers can be a significant challenge. LLM-powered RPA bots can easily translate between languages, making it possible to communicate with customers, partners, and employees worldwide seamlessly.

Continuous Learning and Adaptation LLMs are known for their ability to learn and adapt to new data. This means that LLM-powered RPA bots can continuously improve their performance over time. They can stay up-to-date with industry trends, regulations, and customer preferences, ensuring that they always deliver the highest level of service.

Supply Chain Optimization In manufacturing, the integration of RPA with LLMs can revolutionize supply chain management. RPA serves as the automation engine, streamlining repetitive tasks such as order processing and inventory tracking. Meanwhile, LLMs analyze vast and complex supply chain datasets, identifying patterns, inefficiencies, and anomalies that traditional analytics tools might overlook. This integration brings forth enhanced efficiency, data-driven optimization, improved resilience, and faster decision-making.

Implementation and Integration To fully harness the potential of LLM-powered RPA, organizations must carefully plan and execute the integration. This involves identifying specific use cases where LLMs can provide value, choosing the right RPA tool and LLM, integrating the LLM's API into the RPA solution, defining workflows, developing integration logic, and ensuring robust error handling and security measures.

Challenges and Considerations

While the synergy of LLMs and RPA holds tremendous potential, there are challenges to consider. These include data privacy concerns, ethical implications of generating human-like text, and the complexity of integrating LLMs with existing RPA systems. Addressing these challenges is crucial for a successful and sustainable integration.

In conclusion, the integration of LLMs with RPA represents a transformative force reshaping the future of automation. By enhancing customer support, intelligent data analysis, document processing, multilingual capabilities, and continuous learning, this synergy can drive efficiency, accuracy, and innovation across various industries. As technology continues to advance, organizations that embrace LLM-powered RPA early on will likely gain a competitive advantage in their respective markets.

The Future of LLM-Powered RPA

As technology continues to advance, the fusion of Large Language Models (LLMs) with Robotic Process Automation (RPA) is poised to revolutionize industries. The ability to automate tasks with a high degree of language understanding and generation will unlock new levels of efficiency and productivity. Organizations that embrace LLM-powered RPA early on will likely gain a competitive advantage in their respective markets.

In conclusion, LLM-powered Robotic Process Automation represents a remarkable leap forward in the automation landscape. It has the potential to redefine the way businesses operate, interact with customers, and analyze data. While there are challenges to overcome, the benefits of this technology are too significant to ignore. As we move forward, it will be fascinating to witness the transformative power of LLM-powered RPA in action and the industries that it reshapes along the way.

Integrating LLM with RPA

This integration empowers RPA solutions to comprehend and interact with natural language, enhancing their effectiveness in handling unstructured text data, communicating with humans, and making informed decisions. By choosing the right LLM, establishing API integration, and defining specific use cases, organizations can benefit from improved natural language understanding, text data extraction, content generation, and decision support within their automated processes. However, it's essential to prioritize security, compliance, error handling, and testing, as well as consider scalability and cost management. Regular maintenance and monitoring, along with documentation and training, are critical for ensuring a successful and sustainable LLM-RPA integration that keeps pace with evolving technology.

Considerations for Integrating LLM with RPA

Define Use Case Identify the specific scenarios where integrating LLMs with RPA can bring value, such as extracting data from unstructured text, generating human-like responses, validating data, or supporting decision-making.

Select RPA Tool Choose an RPA tool that supports integrations and can interact with LLMs via APIs or other integration methods. Popular RPA tools include UiPath, Automation Anywhere, and Blue Prism.

Choose LLM Select a suitable LLM that aligns with your requirements.

API Integration Integrate the LLM's API into your RPA solution, utilizing available SDKs or libraries for various programming languages.

Design Workflows Develop RPA workflows that involve interactions with the LLM, such as extracting data from documents, answering customer queries, or providing recommendations. Define the inputs and outputs of these workflows.

Implement Integration Logic Write code or configure the RPA tool to interact with the LLM's API, sending text inputs and processing responses accordingly.

Handle Errors Implement error handling and exception handling mechanisms to manage issues with LLM responses or connectivity.

Test and Debug Thoroughly test the integration and debug any issues, fine-tuning parameters and inputs for optimal performance.

Security and Compliance Ensure security best practices and data protection regulations are followed, handling sensitive data with care and encrypting communication between the RPA tool and the LLM.

The integration of LLMs with RPA enhances automation capabilities across industries, enabling businesses to automate customer interactions, document analysis, and content generation. This fusion of technologies offers significant benefits, including increased efficiency and productivity and a competitive edge. As automation technology evolves, LLM-RPA integration is set to transform organizational operations, customer interactions and data analysis, leading to a future of intelligent automation as the norm.

Case Study: SmartFlow

SmartFlow is an AI-driven Robotic Process Automation (RPA) system that uses pre-trained LLMs and deep learning-based image understanding (Jain et al.). This system can adapt to new scenarios, including changes in user interfaces and input data, without human intervention. SmartFlow uses computer vision and natural language processing to perceive visible elements on graphical user interfaces (GUIs) and convert them into textual representations. LLMs then generate a sequence of actions executed by a scripting engine to complete tasks. SmartFlow has proven robustness across various layouts and applications, automating business processes like form filling, customer service, invoice processing, and back-office operations. By assisting organizations in automating a larger fraction of screen-based workflows, SmartFlow enhances productivity.

Automatic Speech Recognition

Automatic Speech Recognition (ASR), also known as voice recognition, enables machines to convert spoken language into written text. This technology has become increasingly significant in recent years, impacting various fields such as healthcare, entertainment, and human-computer interaction. ASR is a multidisciplinary field that involves both language processing and audio analysis. It not only identifies and transcribes spoken words into text but also interprets their meaning using predictive models, often powered by algorithms like neural networks. Popular voice assistants like Siri and Alexa rely on ASR technology to allow users to interact with devices through voice commands [Reddy, 2024].

ASR systems can be categorized into two types: speaker-dependent and speaker-independent. A speaker-dependent ASR is customized for a specific individual's voice, requiring training with that person's voice data. In contrast, a speaker-independent ASR is designed to recognize speech from any speaker. These systems are typically trained on large datasets that include a wide variety of voices and accents, allowing them to adapt to different speakers [Bartleby Monster, 2013].

Figure 5 illustrates the architecture of a typical ASR system, which consists of three key components [Dhouib et al., 2022]:

Preprocessing Layer This layer focuses on reducing background noise and filtering out any unwanted sounds from the recorded voice, ensuring a cleaner input for the subsequent stages.

Feature Extraction Layer This layer is responsible for identifying and selecting the most relevant features from the audio signal. Example techniques include the Mel-frequency cepstral coefficients (MFCCs) an the Discrete Wavelet Transform (DWT) [Papastratis, 2021].

Classification Layer with Language Model In the final stage, a language model aids the classification layer in recognizing and interpreting the spoken words based on the extracted features.

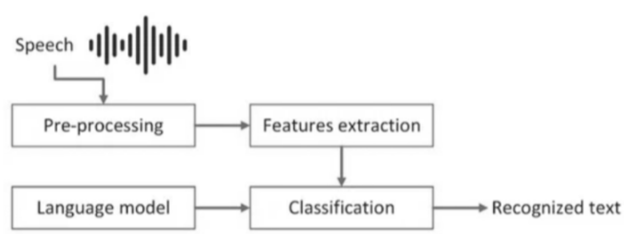

Figure 5: Speech Recognition Process

Speech recognition is one of the most challenging areas of computer science because people speak in a variety of accents, dialects, and speaking styles that make it difficult to identify voice accurately. Advanced algorithms and machine learning approaches are used to process and analyse spoken language in order to address this problem. A few traditional and state-of-the arts algorithms employed for ASR are covered in the section that follows [Cho, 2021].

Hidden Markov Models

A Hidden Markov Model (HMM) is made up of a series of states connected by transitions, where each state depends solely on the previous one. A Bayesian network can be used to represent this dependency structure. In this network, there is a start state, and each time the next state is selected randomly by a probability distribution.In ASR, phonemes are represented as states, while transitions between them reflect the probabilistic shifts, indicating the likelihood of one phoneme following another [Lamel and Gauvain, 2003, p. 305]. Since the 1960s, HMMs have been central to voice recognition. Rabiner [1989] highlighted their effectiveness by comparing them with traditional Markov models, which link each state directly to an observable event, making them predictable but rigid. In contrast to traditional Markov models, HMMs add a hidden layer that makes use of probabilities to determine observations. This feature allows HMMs to handle the variability in speech more effectively, making them particularly well-suited for voice recognition tasks. Alotaibi et al. [2008] designed a system based on HMM to recognize Arabic spoken alphabet. They made use of the telephony Arabic corpus of the Saudi Accented Arabic Voice Bank (SAAVB) as dataset to train their model. They used a Hidden Markov Model Toolkit (HTK) for building their recognizer and were able to achieve an accuracy of 93.72%.

Dynamic Time Warping

Dynamic Time Warping (DTW) is a technique used to measure the differences between two time series. In speech recognition, DTW is very useful in voice recognition since it can align words that are pronounced at varying speeds or durations, making comparison and recognition accurate. DTW helps align a spoken word with a reference template by stretching or compressing parts of the sequences to find the best match. This makes it possible for the system to identify words whether they are spoken slowly or quickly [Geek, 2024].

Researchers have explored the use of DTW to enhance ASR system. For instance, Jiang [2020] applied DTW to address the challenges of aligning spoken words with reference patterns despite variations in speech speed and pronunciation. By aligning input speech with stored templates, DTW improves recognition accuracy, effectively managing speech variability and enhancing overall performance.

State-of-the-Art Methods

In the past, machine learning for ASR relied heavily on domain-specific knowledge and traditional digital signal processing to extract features like phonemes. With the advancement of deep learning, the traditional audio processing technique is no longer required as it can directly extract features from raw data, removing the need for manual feature extraction. In deep learning, audio data is often processed using CNN architecture, and features are then extracted to create feature maps. These feature maps are fed via a classifier made up of a few fully connected layers for audio classification. In the case of voice recognition, this can be used to identify someone's mood or even detect human emotion [Doshi, 2021].

Newatia [2018] highlighted the effectiveness of CNNs in ASR, particularly focusing on the pooling layer and the weight-sharing mechanism. Weight sharing lowers the network's complexity, making it simpler to train and more effective, while the pooling layer assists in removing unnecessary data and shrinking the feature maps. Nagajyothi [2018] also proposed CNN for ASR instead of traditional neural networks.

The transformer model is another powerful deep learning method for ASR. The model includes an encoder and a decoder. The encoder processes the input speech to extract meaningful features, while the decoder translates these features into text. A popular example of a transformer-based ASR system is Whisper, developed by OpenAI [Radford et al., 2023]. Despite the fact that CNNs and transformer model have improved speech recognition, ongoing research continues to explore new algorithms. Recent research has demonstrated that LLMs are a viable improvement for ASR systems. The following section discusses the integration of LLMs within ASR technology.

ASR with LLMs

LLMs are a type of AI that can enhance ASR. Trained on large datasets, these models understand the relationships between words and sentences, which helps them predict the next word in a sequence. LLMs can be applied to speech recognition in two key ways [Pawar, 2023]:

First-pass Recognition The LLM model is used to generate a list of potential transcriptions for an input audio. This list is then handed over to a traditional recognition system to select the most likely one.

Second-pass Rescoring This approach is used after a traditional recognition system creates a list of possible transcriptions from the audio. The LLM is then used to evaluate and pick the best transcription. The initial task of converting voice to text is handled by the ASR model and not LLMs. However, LLMs play a crucial role after the ASR model generates its output. Once it provides a list of potential word sequences, LLMs evaluate these sequences by scoring them based on their likelihood of being correct and appropriate in the given context. This approach is particularly useful for resolving ambiguities when the ASR model produces multiple possible word interpretations. LLMs can be used in this situation to resolve ambuiguity [Singh, 2024].

The below image illustrates how the second-pass rescoring works. The decoder generates a list of possible sequences, and these are sent for rescoring using LLM, which picks the best one.

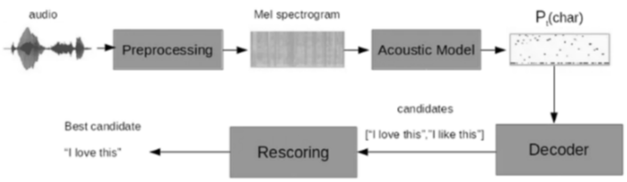

Figure 6: Second-pass Rescoring

Numerous studies, including Pu et al. [2023], have employed LLMs to enhance traditional ASR systems. By using their approach, the ASR system first analyzes the speech and generates a list of n-best hypotheses. These hypotheses are then rescored on a second pass in the first stage. After then, only transcripts that pass a predefined threshold are sent to the LLM for correction. Their approach demonstrated a 10%–20% improvement.

Conclusion

LLMs can be combined with knowledge graphs in various real-world applications to enhance their performance, accuracy, and contextual understanding. Some examples of applications using a combination of LLMs and knowledge graphs are IBM Watson for oncologists, Google's Knowledge Graph in Medical Search and even with ChatGPT's model, we can customise it to provide to provide domain specific knowledge. Given the critical importance of ontologies and large language models (LLMs) in today's technological landscape, integrating these fields will undoubtedly yield significant benefits. LLMs, with their powerful language understanding and generation capabilities, can be enhanced by the structured, explicit knowledge representation provided by ontologies. This synergy will lead to more intelligent, context-aware, and capable AI systems, enabling better data management, decision-making, and knowledge retrieval across various industries. The integration of LLMs with RPA is poised to revolutionize industries by enabling the automation of tasks with advanced language understanding and generation, thereby unlocking new levels of efficiency and productivity. As organizations adopt LLM-powered RPA, they will likely gain a competitive advantage, but it is crucial to address challenges such as security, compliance, and scalability to ensure a successful and sustainable integration. AI has played a key role in advancing ASR. In the past, features from audio signals were manually extracted using methods such as MFCCs from audio signals, and these features were fed into models such as HMMs to map them to specific phonemes. This approach required significant domain knowledge for feature engineering and model tuning. With the invention of CNN, models can now learn these features on their own, eliminating the need for manual feature extraction. Adding to that, the use of transfer learning techniques helps these deep learning models learn more relevant features and perform better. On the other hand, LLM can further enhance ASR systems by using the power of natural language processing to construct more meaningful phrases. From conventional machine learning methods to deep learning and LLMs, there has been a notable advancement in the field of speech recognition, and this trend is expected to continue with even more significant advancements in the future, driven by innovations such as multilingual models, enhanced natural language understanding, and integration with other modalities like gestures and facial expressions.

References

1 Team, E. (2024) The role of ontologies with llms, Enterprise Knowledge. Available at: https://enterprise-knowledge.com/the-role-of-ontologies-with-llms/ [Accessed: 13 August 2024].

2 Schrader, B. (2022) What's the difference between an ontology and a knowledge graph?, Enterprise Knowledge. Available at: https://enterprise-knowledge.com/whats-the-difference-between-an-ontology-and-a-knowledge-graph/ [Accessed: 13 August 2024].

3 Berners Lee, T. (PDF) the semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. Available at: https://www.researchgate.net/publication/225070375_The_Semantic_Web_A_New_Form_of_Web_Content_That_is_Meaningful_to_Computers_Will_Unleash_a_Revolution_of_New_Possibilities [Accessed: 13 August 2024].

4 Besen, S. (2024) LLM embeddings - explained simply, Medium. Available at: https://pub.aimind.so/llm-embeddings-explained-simply-f7536d3d0e4b[Accessed: 14 August 2024].

5 Merritt, R. (2024) What is retrieval-augmented generation aka rag?, NVIDIA Blog. Available at: https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/#:~:text=Retrieval_2Daugmented_20generation_20(RAG),facts_20fetched_20from_20external_20sources [Accessed: 18 August 2024].

6 Bratanič, T. (2024) Knowledge graphs & LLMS: Multi-hop question answering, Graph Database & Analytics. Available at: https://neo4j.com/developer-blog/knowledge-graphs-llms-multi-hop-question-answering/ [Accessed: 18 August 2024].

7 Aa, H. van der, Aalst, W.M.P. van der, Agostinelli, S., Auth, G., Bensberg, F., Chakraborti, T., Chircu, A., Croon, P. and Czarnecki, C. (2021). Robotic Process Automation : Management, Technology, Applications. München: De Gruyter Oldenbourg.

8 Government use of Artificial Intelligence in New Zealand. (n.d.). Available at: https://data.govt.nz/assets/data-ethics/algorithm/NZLF-report.pdf [Accessed 20 Aug. 2024].

9 IBM (2023). What Is Robotic Process Automation (RPA)? — IBM. [online] www.ibm.com. Available at: https://www.ibm.com/topics/rpa.

10 Jain, A., Paliwal, S., Sharma, M., Vig, L. and Shroff, G. (2024). SmartFlow Robotic Process Automation using LLMs. [online] Webflow.io. Available at: https://smartflow-4c5a0a.webflow.io/ [Accessed 20 Aug. 2024].

11 Jain, A., Paliwal, S., Sharma, M., Vig, L. and Shroff, G. (n.d.). SmartFlow: Robotic Process Automation using LLMs. [online] Available at: https://arxiv.org/pdf/2405.12842 [Accessed 20 Aug. 2024].

12 Naqvi, A. and J. Mark Munoz (2020). Handbook of Artificial Intelligence and Robotic Process Automation. Anthem Press. Statista. (2024). UiPath global revenue 2024 — Statista. [online] Available at: https://www.statista.com/statistics/1231627/worldwide-uipath-revenue/ [Accessed 20 Aug. 2024].

13 Stock Analysis. (2024). UiPath Inc. (PATH) Stock Price, Quote & News - Stock Analysis. [online] Available at: https://stockanalysis.com/stocks/path/ [Accessed 20 Aug. 2024].

14 Tangentia —. (2023). Tangentia Inc. [online] Available at: https://www.tangentia.com/llm-powered-robotic-process-automation-unleashing-the-industry-transformation-revolution/ [Accessed 20 Aug. 2024].

15 Gershon, L. (2017). The automation resistant skills we should nurture. BBC. Available at: http://www.bbc.com/capital/story/20170726-the-automation-resistant-skills-we-should-nurture [Accessed 20 Aug. 2024].

16 PWC (2016). Sizing the prize. PWC's Global Artificial Intelligence Study: Exploiting the AI Revolution. Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html. [Accessed 20 Aug. 2024].

17 Rao, A. (2017). A strategist's guide to artificial intelligence. Strategy + Business. Available at: https://www.strategy-business.com/article/A-Strategists-Guide-to-Artificial-Intelligence?gko=0abb5&utm_source=itw&utm_medium=20170523&utm_campaign=respB. [Accessed 20 Aug. 2024].

18 Vaswani, K. (2017). Tomorrow's cities: Singapore's plans for a smart nation. BBC. Available at: http://www.bbc.com/news/technology-39641262. [Accessed 20 Aug. 2024].

19 Wendel, S. (2017). U.A.E. appoints minister of artificial intelligence in cabinet reshuffle. Available at: https://www.forbesmiddleeast.com/en/u-a-e-appoints-minister-of-artificial-intelligence-in-cabinet-reshuffle/. [Accessed 20 Aug. 2024].

20 Sonto, L. and Biggemann, S. (2020). Applications of Artificial Intelligence and RPA to Improve Government Performance. Cambridge Core, [online] pp.141–150. Available at: https://www.cambridge.org/core/books/abs/handbook-of-artificial-intelligence-and-robotic-process-automation/applications-of-artificial-intelligence-and-rpa-to-improve-government-performance/BD44D7A5E11446819D5A1134F3C807D0 [Accessed 20 Aug. 2024].

21 Gray, R. (2017). How long will it take for your job to be automated? BBC. Available at: http://www.bbc.com/capital/story/20170619-how-long-will-it-take-for-your-job-to-be-automated. [Accessed 20 Aug. 2024].

22 Hicks, M.J. and Devaraj, S., 2015. The myth and the reality of manufacturing in America. Center for Business and Economic Research, Ball State University, 6..

23 Nesbitt, J. (2017). 4 ways artificial intelligence is transforming trade. Available at: http://www.traderead y.ca/2017/topics/import-export-trade-management/4-ways-artificial-intelligence-transforming-trade/. [Accessed 20 Aug. 2024].

24 Reddy, R. (2024). What is speech recognition in artificial intelligence?, Medium. Available at: https://medium.com/@rrathnakar707/what-is-speech-recognition-in-artificial-intelligence-8579eaa98a3c#:~:text=Speech%20recognition%20allows%20computers%2C%20applications,into%20text%20for%20business%20solutions. [Accessed: 14 August 2024].

25 Cho, H. (2021). Having Problems with Voice Recognition? You're NOT Wrong, Medium. Available at: https://medium.com/@haejin2909/having-problems-with-voice-recognition-youre-not-wrong-c26f9eec8d4 [Accessed 14 Aug. 2024].

26 Dhouib, A. et al. (2022) 'Arabic automatic speech recognition: A systematic literature review', Applied Sciences, 12(17), p. 8898. doi:10.3390/app12178898.

27 Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286. doi: 10.1109/5.18626.

28 Lamel, L. and Gauvain, J.L., 2003. Speech recognition. The Oxford Handbook of computational linguistics, pp.305-322.

29 Alotaibi, Y., Al-Ghamdi, M., & Alotaiby, F. (2008). Speech Recognition System of Arabic Digits based on A Telephony Arabic Corpus. IPCV, 35–38. `https://dblp.uni-trier.de/db/conf/ipcv/ipcv2008.html#AlotaibiAA08`.

30 Geek, T.P. (2024) Unlocking the potential of time series with dynamic time warping and ai, Medium. Available at: `https://medium.com/@theprogramminggeek/unlocking-the-potential-of-time-series-with-dynamic-time-warping-and-ai-d9bc0a072712` [Accessed: 16 August 2024].

31 Jiang, S., & Chen, Z. (2023). Application of dynamic time warping optimization algorithm in speech recognition of machine translation. Heliyon, 9(11).

32 Doshi, K. (2021) Audio Deep Learning Made Simple (Part 1): State-of-the-Art Techniques, Medium. Available at: `https://towardsdatascience.com/audio-deep-learning-made-simple-part-1-state-of-the-art-techniques-da1d3dff2504`.

33 Newatia, S. and Aggarwal, R.K., 2018, March. Convolutional neural network for ASR. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 638-642). IEEE.

34 Nagajyothi, D. and Siddaiah, P. (2018) "Speech Recognition Using Convolutional Neural Networks", International Journal of Engineering & Technology, 7(4.6), pp. 133–137. doi:10.14419/ijet.v7i4.6.20449.

35 Pawar, S. (2023) How to Use Large Language Models for Speech Recognition?, Medium. Medium. Available at: `https://medium.com/@pawarsa5098/how-to-use-large-language-models-for-speech-recognition-86eb9d7ff2a` [Accessed: 18 August 2024].

36 Pu, J., Nguyen, T.S. and Stüker, S., 2023. Multi-stage Large Language Model Correction for Speech Recognition. arXiv preprint arXiv:2310.11532.

37 Papastratis, I. (2021). Speech Recognition: a review of the different deep learning approaches. [online] AI Summer. Available at: `https://theaisummer.com/speech-recognition/`.

38 Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C. and Sutskever, I., 2023, July. Robust speech recognition via large-scale weak supervision. In International conference on machine learning (pp. 28492-28518). PMLR.

30 Singh, S. (2024). Improving ASR Models Using LLM-Powered Workflow. [online] Labellerr. Available at: `https://www.labellerr.com/blog/asr-models-enhacement-with-llm/#integration-of-llms-with-asr` [Accessed 25 Aug. 2024].

31 Kunze, J., Kirsch, L., Kurenkov, I., Krug, A., Johannsmeier, J. and Stober, S., 2017. Transfer learning for speech recognition on a budget. arXiv preprint arXiv:1706.00290.

32 Bartleby Monster (2013). What Is Speech Recognition? [online] Plum Voice. Available at: `https://www.plumvoice.com/resources/blog/speech-recognition/#:~:text=Some%20speech%20recognition%20systems%20are` [Accessed 14 Sep. 2024]