ARTIFICIAL INTELLIGENCE BASED VISA PROCESSING SYSTEM TO REDUCE THE PAPER USAGE, VISA PROCESSING TIME AND REDUCE THE HUMAN ERRORS DURING THE APPLICATION REVIEW PROCESS

by

Azeez Chollampat

DISSERTATION
Presented to the Swiss School of Business and Management Geneva
In Partial Fulfillment
Of the Requirements
For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

August, 2025

AI BASED VISA PROCESSING SYSTEM TO REDUCE THE PAPER USAGE, VISA PROCESSING TIME AND REDUCE THE HUMAN ERRORS DURING THE APPLICATION REVIEW PROCESS

by

Azeez Chollampat

Supervised by

Dr. Luka Leško

APPROVED BY

Apostolos Dasilas

Dissertation chair

RECEIVED/APPROVED BY:

Renee Goldstein Osmic

Admissions Director

Dedication

To my beloved family, whose steady support and encouragement kept me going.

To my mentors and teachers, who inspired my curiosity and challenged me to grow.

And to all those who strive for knowledge and innovation - may this work serve as a small step forward in our shared journey of discovery.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to God for guiding me throughout this DBA journey. I received the strength, perseverance, or wisdom to get through this through God's blessings.

I want to express my heartfelt gratitude to Dr. GUNJAN BHATIA, my first advisor. At the start of my DBA journey, when everything felt new, her gentle guidance made all the difference. I'm so grateful for the encouragement and support she gave me when I needed it most.

A huge thank you to Dr. Luka Leško, my current advisor. Thank you for stepping in when I needed direction. I could always count on your reassurance, especially during moments when I needed guidance.

I owe more than I can ever say to my wife, Shabeen Chollampat. Shabeen, thank you for always being my unwavering pillar of support. You held our world together with so much patience and strength. To my kids, Nilu (Nlofer Chollampat) and Abu (Nabeel Chollampat) - you guys fill my life with laughter and purpose. Your smiles have kept me going, and remember what matters most.

A big thank you to the Exalture team for all your help with collecting survey data. Your efforts made a real difference in this work, and I sincerely appreciate the time and care you invested in it.

Additionally, I would like to extend a special thank you to all the individuals who took the time to respond to my survey, conducted as part of this research. Your participation and honest insights added value to this research, and I am deeply grateful for your contributions.

This journey isn't a solo one. I'm deeply grateful to all the mentors, colleagues, and friends who have supported, encouraged, and believed in me along the way. Your kindness and encouragement have meant the world to me, and I feel incredibly lucky to have you all by my side.

Thank you all for being part of this chapter in my life.

ABSTRACT

AI BASED VISA PROCESSING SYSTEM TO REDUCE THE PAPER USAGE, VISA PROCESSING TIME AND REDUCE THE HUMAN ERRORS DURING THE APPLICATION REVIEW PROCESS

Azeez Chollampat
2025

Dissertation Chair: Apostolos Dasilas

Visa processing is often the most crucial, yet frequently the most challenging, part of international travel. The current visa processing system has several difficulties, including excessive paperwork, frequent travel to visa centres, prolonged processing times, and a high incidence of human errors. These issues are highly inconvenient for applicants, particularly vulnerable groups such as senior citizens, pregnant women, and families with children. It also puts an operational challenge for consulates and service providers. There is an increasing global demand for seamless and sustainable services, as well as a compelling need to redesign the current visa processing system.

This thesis presents a framework for an AI-based visa processing system designed to minimize the difficulties of the current visa processing system. The proposed system integrates several cutting-edge artificial intelligence techniques, including Optical Character Recognition (OCR), Natural Language Processing (NLP), and machine learning algorithms. The thesis goes into the architecture, training, and deployment of these AI components, with a strong focus on the accuracy, robustness, and scalability of

the system. It also explores hybrid AI approaches that combine multiple techniques such as deep learning models, language transformers, and rule-based systems to handle the complex variability in visa documents. The literature review includes benchmarking of OCR engines and NLP frameworks, as well as an evaluation of real-world use cases and datasets.

Beyond the technical implementation, the thesis also addresses challenges related to bias, explainability, and ethical use of AI in administrative decision-making. By integrating environmental sustainability goals using AI, the thesis research proposes a next-generation visa processing system which is faster, greener, and more reliable. That is, the proposed system aims to enhance the user experience for applicants, improving the overall efficiency and transparency of visa processing authorities and making the process much more environmentally friendly.

TABLE OF CONTENTS

List of Figures

CHAPTER I:

INTRODUCTION

1.1 Introduction

Visa processing is the first and most crucial step in international travel, and for ensuring a smooth travel experience, visa processing should also be smooth. Typically, visa applicants include students with admission deadlines, professionals with overseas job offers, and families with urgent travel needs. However, the current visa processing system has several challenges. It heavily relies on manual processing, paperwork and physical visits. Getting an appointment often involves long waitlists, which causes uncertainty and stress for applicants with urgent travel needs. Once the process begins, applicants have to submit several pages of documents ranging from academic certificates to financial statements. The heavy reliance on paper leads to both environmental waste, economic inefficiency and time wastage. It also forces applicants to visit physical centres, which is particularly difficult for senior citizens, pregnant women, and families with small children who must endure long queues. In the current visa processing system, embassy staff have to work with huge volumes of paperwork. This makes the process slow and error-prone, with mistakes such as misfiled documents or incorrect data entry, and often results in rejected applications or missed travel opportunities. Also, the current verification methods are outdated, as manual checks are ineffective in detecting fraud. Also, in manual checks, genuine applicants are repeatedly asked for clarifications. Moreover, there are only limited tracking facilities available for applicants. This leaves the applicants in the dark about their visa status and creates anxiety as travel dates

approach. These inefficiencies not only affect applicants but also put embassies under heavy pressure, while governments risk reputational damage due to delays that affect tourism, education, and investment. In this context, there is a strong and urgent need for technology-driven solutions such as AI-based visa processing. By adopting AI in visa processing, it will become faster, more accurate, transparent, and user-friendly, ultimately easing the burden on applicants, reducing embassy workloads, minimising environmental impact, and strengthening international mobility.

1.2 Research Problem

The current visa processing system needs manual document review, which is quite time-consuming, labour-intensive, prone to errors, and environmentally unsustainable due to the massive paper consumption. Despite the use of digital tools in visa processing, the core administrative workload remains largely manual and inefficient. The lack of automation, along with the limited use of advanced technologies such as artificial intelligence, is constraining the current visa processing system. It forces applicants to face long waiting periods with unpredictable outcomes. It also makes the processing centres struggle with workload and cost inefficiencies.

This research has been conducted in India on the operational workflows, document formats, and administrative practices of Indian visa-processing centres. In India, visa processing times used to vary significantly depending on the type of visa and the issuing country. The Ministry of Home Affairs website reports that the Indian e-Visa

applications are generally processed within a minimum of three working days after submission to the relevant mission or post. Schengen visas typically require one to four weeks to secure an appointment, and it takes a further 10 to 15 working days for processing. In some cases, this period can extend up to 30 days during peak seasons or when additional checks are required. Canada visitor visas usually take between 15 and 30 working days to process, with some cases extending to 23 to 40 calendar days. U.S. B1/B2 visitor visas have the longest appointment wait times in India, ranging from around six months in the fastest locations to over a year in the busiest cities. Once the interview is conducted, the visa is generally issued within one to two weeks. These figures highlight substantial time lags in current visa-processing workflows in India, underscoring the potential of an AI-based system to reduce overall turnaround time.

1.3 Purpose of Research

The primary purpose of this research is to design and develop an Artificial Intelligence (AI)-based visa processing system capable of transforming the current manual approach based visa processing system into a faster, more reliable, and user-friendly digital framework. The existing visa processing system has a heavy reliance on paperwork, lengthy processing timelines, and is error-prone. These inefficiencies place an administrative burden on immigration authorities and often create frustration and stress for applicants. This research aims to leverage AI technologies to automate, streamline, and enhance overall system performance.

A key objective of this study is to minimise paperwork and manual interventions by digitising and automating core components of the application process. Optical Character Recognition (OCR) technology helps in extracting text from scanned documents such as passports, bank statements, or supporting letters. Thus, OCR reduces the need for manual data entry and lowers the likelihood of transcription errors in the visa processing system. Similarly, Natural Language Processing (NLP) techniques help to interpret text responses from applicants and thus ensure that information is both complete and contextually relevant. When combined with ML algorithms, these technologies can further learn from historical data, improving accuracy and decision-making over time.

Another purpose of the research is to shorten processing times. By automating document handling and verification, AI can accelerate the pace at which visa applications are reviewed. Faster processing helps in international education, tourism, and employment, where timely visa decisions are critical.

The study also emphasises reducing human errors that are common in manual systems. By embedding AI-driven checks, the system aims to provide an error-free standardised process. At the same time, the system design considers the importance of human oversight to ensure that final decisions remain fair, ethical, and accountable.

In addition to administrative improvements, this research is motivated by the goal of enhancing the applicant experience. Transparency and clarity in communication are

key to building applicant confidence in the system, and AI offers the tools to deliver this experience consistently.

Also, current visa systems involve large volumes of paperwork. This leads to environmental impact and inefficiency. By moving toward minimal paperwork, AI-based systems promote eco-friendly practices while simultaneously lowering administrative costs. Thus, the research aligns with broader goals of digital transformation and environmental sustainability.

In summary the purpose of this research is not merely to introduce automation for its own sake but to redefine the visa processing paradigm through the intelligent integration of AI technologies. By reducing paperwork, accelerating decision timelines, minimising human errors, and prioritising applicant experience, the study seeks to demonstrate that AI can be a transformative force in public service delivery. Ultimately, this research envisions a visa processing system that is efficient, transparent, sustainable, and people-centric, offering significant improvements for both applicants and administrators.

1.4 Significance of the Study

The current visa processing system heavily depends on manual tasks. Examples include data entry, document verification, and repetitive cross-checking of information, and these can often slow the process. These issues not only delay outcomes for applicants

but also increase workloads for immigration officers. The significance of this study is that it can be used to transform the current visa processing systems by introducing artificial intelligence (AI) to address inefficiencies, reduce human errors, and enhance user experiences. The research demonstrates how AI technologies can automate such repetitive and time-consuming tasks, leading to a faster, more accurate, and more applicant-friendly system. Also, the significance extends to sustainability, ethics, and governance, making it relevant both technically and socially.

A primary outcome of this study is that integrating AI into visa processing systems can reduce human effort in repetitive administrative tasks, which is currently consuming a disproportionate amount of time and resources. For instance, immigration officers often have to spend hours verifying standard documents such as passports, financial statements, and travel itineraries. Automating these manual processes with AI technologies such as Optical Character Recognition (OCR), Natural Language Processing (NLP) and Machine Learning (ML) can reduce workloads and allow officers to dedicate their expertise to more complex cases. This makes the visa process faster and more reliable. Also, officers can then focus on sensitive and high-stakes decisions rather than monotonous routine tasks.

Another significant contribution of this study is that it focuses on improving accuracy in visa processing systems. Human errors are often non-negotiable in manual systems due to fatigue, oversight, or inconsistency in document review. These errors can

lead to delays, incorrect rejections, or even risks of fraud going unnoticed. By embedding AI technologies such as Natural Language Processing (NLP) for analysing written responses and OCR for error-free data extraction, the system minimises these risks. With AI-based visa processing systems, visa applicants can benefit from a more consistent and transparent process and immigration departments can gain improved institutional credibility. This reliability is crucial because visa decisions impact not just travel but also careers, family reunifications, and international collaborations.

The study also has a strong environmental significance. Transforming the current visa processing system through AI integration has a significant environmental impact, as it substantially reduces paper dependency and associated environmental harm. So this study aligns with global sustainability goals by promoting eco-friendly practices and reducing carbon footprints associated with physical documents. Thus, the study addresses not only efficiency concerns but also broader environmental responsibilities of government systems.

The study also takes a human-centred approach by incorporating a survey that draws on perspectives from visa applicants, immigration officers, and the general public about AI in visa processing. This study is therefore significant due to its responsible AI adoption nature. The system balances efficiency with accountability. This human-centred approach enhances trust and ensures that AI is implemented in a way that respects applicants' rights and expectations.

In summary, the significance of this study lies in its multifaceted contributions, such as enhancing efficiency and improving accuracy and fairness through AI-driven automation. It also addresses sustainability concerns by reducing paper usage and ensuring ethical implementation through the incorporation of stakeholder perspectives.

1.5 Research Purpose and Questions

The overall purpose of this research is to understand whether AI can transform the current visa processing system and the perspectives and concerns of people regarding AI in visa processing. To achieve this, the study investigates the following research questions:

- Can an AI-based visa processing system significantly reduce paper usage, processing time, and human errors compared to the traditional system?
- What are the perceptions, preferences, and concerns of visa applicants, immigration officers, and the general public regarding the adoption of AI in visa processing?
- How can AI technologies, such as OCR, NLP, sentiment analysis, and document understanding, be effectively integrated to automate and streamline visa processing tasks?

Hypothesis: An AI-based visa processing system can significantly improve the administrative efficiency of the current visa processing system by reducing paper

dependency, shortening processing times, and minimising human errors. Furthermore, if designed with transparency, fairness, and user-centricity, the AI-based system will be positively received across diverse demographic and stakeholder groups.

CHAPTER II:

REVIEW OF LITERATURE

This AI-based visa processing system project aims to enhance visa processing tasks by making them more efficient and accurate through the use of artificial intelligence (AI) technologies. This literature review examines various components involved in the system, including Natural Language Processing (NLP), Optical Character Recognition (OCR), text spotting, document understanding, and scene comprehension. These areas are to be analysed to understand their evolution and to determine how they can be efficiently and accurately incorporated into the visa processing system. This literature review also considers research papers that explore hybrid approaches that combine different AI technologies to improve a system's functionality. Another important consideration in this literature review is to analyse errors that can be incurred in each area of implementation of the system and understand how they can impact the visa processing workflow. This project aims to improve the efficiency and accuracy of the visa processing system in the best possible way by integrating these AI technologies.

2.1 Optical Character Recognition (OCR)

Optical Character Recognition (OCR) helps to convert texts in images into machine-readable text. The purpose of OCR in visa processing systems is to help convert scanned, printed, or handwritten text into machine-readable data. OCR extracts information from hardcopy documents, such as passports, certificates, and forms and thus helps to make them machine-readable for further analysis and decision-making. The

research papers discussed in this review address the challenges associated with recognising and processing text in various scenarios, including complex scenes, non-English scripts, and handwritten characters.

The paper "Reading Digits in Natural Images with Unsupervised Feature Learning" by Netzer et al. (2011) presents a study on recognising digits in natural images. It discusses the challenges of recognising digits in complex scenes, such as photographs, due to variations in lighting, blur, perspective, fonts, and background clutter. The paper introduces the Street View House Numbers (SVHN) dataset, a benchmark comprising over 600,000 labelled digits extracted from Street View images. It demonstrates the limitations of hand-designed features in this task. It then employs unsupervised feature learning methods, specifically stacked sparse auto-encoders and a K-means-based system, and shows superior performance compared to traditional features. It then describes the integration of the proposed digit recognition system into a real-world application for automatic house number recognition in Street View. It also gives the proposed unsupervised learning technique results in improving the accuracy and robustness of text recognition within natural images. Overall, this paper gives an overview of the challenges in recognising text from natural images and proposes an unsupervised learning technique to address those problems.

The paper "Efficient, Lexicon-Free OCR using Deep Learning" by Namysl and Konya (2019) explains the challenges in text recognition, especially in unconstrained

settings, such as diverse font types, natural scenes with distortions, and complex backgrounds. It highlights that despite advancements in OCR, even state-of-the-art solutions struggle with these complexities. It then proposes a novel approach which is efficient, adaptable, and robust for challenging OCR situations with diverse fonts and complex backgrounds. The proposed approach is a segmentation-free method that processes entire text lines or words as single units, thereby avoiding potential errors associated with character segmentation. The proposed system leverages deep learning to recognise text, even in challenging scenarios accurately. The paper employs two main deep-learning models for OCR systems: a hybrid CNN-LSTM model and a fully convolutional model. The hybrid CNN-LSTM model combines the feature extraction capabilities of CNNs with the sequential processing power of LSTMs. The CNNs identify visual elements in the text image, and the LSTM analyses these features in context to accurately recognise the characters and their order. The fully convolutional model uses CNNs to extract visual features from the text image and directly learn the patterns and relationships between characters. Thus, it is a faster and more efficient alternative to the hybrid model. The paper then describes the use of synthetic data and data augmentation techniques, which can be used to improve the system's ability to handle diverse text styles and challenging conditions. It also gives experimental results demonstrating the superior performance of their approach compared to other OCR engines, particularly in scenarios with significant distortions and complex backgrounds. The paper concludes that the proposed method offers a promising solution for various OCR applications due to its efficiency, adaptability, and robustness.

The paper "TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models" by Li et al. (2023) utilises pre-trained Transformers for both image understanding and text generation. It explains the conventional method for OCR, which involves using a Convolutional Neural Network (CNN) to analyse and comprehend the input image, followed by a Recurrent Neural Network (RNN) to identify individual characters in it, and finally, integrating a language model to enhance the overall accuracy of the recognised text. Unlike the conventional approach, the paper utilises "TrOCR," an end-to-end text recognition model based on the Transformer architecture, which can address the challenges associated with text recognition. It studies the results of using TrOCR in various text recognition tasks, including printed, handwritten, and scene text recognition and has a reasonable recognition rate even in complex situations. This paper presents state-of-the-art results on various OCR benchmark datasets available at the time without requiring complex pre- or post-processing. It is a very good solution for OCR applications due to its model's simplicity and effectiveness. The system even has the potential for multilingual use by using pre-trained models in the decoder. This research also marks a significant advancement in the field of OCR, as it provides an end-to-end solution for even the most complex text recognition.

The paper "Classifying Promotion Images Using Optical Character Recognition and Naïve Bayes Classifier" by Phoenix et. al (2021) categorises images automatically as promotional or non-promotional using OCR. It highlights the positive impact of sales promotions and social media marketing on customer behaviour. The primary objective of

this study is to develop an AI system that can automatically determine whether an image contains information about a promotional offer, eliminating the need for human intervention. It proposes a system utilising OCR and a Naïve Bayes algorithm. It uses a dataset of 158 images, some of which contain promotional information. The system uses OCR to extract text from images as textual information and utilises it as a primary indicator for detecting the presence of a promotional offer. The preprocessing techniques and classification algorithms, particularly Naïve Bayes, used in this paper yield awe-inspiring results, which highlight the potential of OCR and the Naïve Bayes Algorithm for automatic textual image classification. The paper has broader applications and social impacts, as it can be used to monitor social media and deliver timely user notifications, among other uses.

The paper titled "Optical Character Recognition System for Nastalique Urdu-Like Script Languages Using Supervised Learning" by Rizvi et al. (2019) gives a significant contribution to the development of an OCR system for Nastalique Urdu-like script languages. Nastalique Urdu is a highly cursive, complex, and bi-directional script. This study utilises an OCR system to convert written or printed Nastalique Urdu text images into a digital format for preservation. Direct storage of written or printed text as images requires a significant amount of storage. The paper proposes a supervised learning-based OCR system that achieves remarkable recognition rates during its training and testing phases. This system can be utilised for the digital transformation of Urdu and similar scripts, including Arabic-like ones found in languages like Persian and Punjabi. The

paper also discusses the concept of OCR, its applications, and its role in the conversion of handwritten or printed text into a digital format. It also discusses the limitations of text in image form and the benefits of OCR in various applications. It uses Urdu as a central case study. Urdu is the eighth-largest spoken language and national language of Pakistan. The proposed supervised learning-based OCR system for the Nastaliq Urdu script is a suitable solution, as it is easy to implement and has a high accuracy rate compared to other approaches of that time. The proposed OCR system is significant in digitising Nastalique Urdu and other similar script languages.

The paper titled "Implementation of Optical Character Recognition using Tesseract with the Javanese Script Target in Android Application" by Robby et al. (2019) gives a comprehensive overview about OCR for non-Latin scripts, especially the Javanese script. There is much research on OCR for Latin alphabets, and it is well established. However, recognising non-Latin scripts is somewhat challenging due to their unique character shapes and contours when compared to Latin scripts. The researchers create a comprehensive dataset of Javanese characters and train OCR models on it using Tesseract OCR tools. They also implemented these models in an Android-based mobile application. The research combines single boundary boxes for the entire character and separate boxes for the main body and sandangan components. With this approach, they achieved an accuracy rate of 97.50%, which is quite good. These results demonstrate the effectiveness of the proposed approach in making the Javanese script accessible to a

broader audience. Also, the results give valuable insights that are useful for developing OCR systems for non-Latin scripts.

The paper "Assessing the Impact of OCR Quality on Downstream NLP Tasks" by Van Strien et al. (2020) analyzes how the quality of OCR impacts various Natural Language Processing (NLP) tasks. It gives a study on how OCR recognition rates affect digital humanities and information retrieval. It also presents a study on the influence of OCR errors on various tasks, including sentence segmentation, named entity recognition, dependency parsing, information retrieval, topic modelling, and neural language model fine-tuning. The paper argues that there is a consistent impact of OCR errors on these tasks, and the severity of these errors varies across different components. The paper highlights the need for best practices when working with text recognized using OCR systems. The paper also discusses the importance of understanding how different tools work with OCR errors in diverse NLP tasks. This research paper provides valuable insights into the use of OCR in NLP applications, particularly within digital humanities and information retrieval.

The paper titled "Handwritten Arabic Optical Character Recognition Approach Based on Hybrid Whale Optimisation Algorithm With Neighborhood Rough Set" by Sahlol et al. (2020) developed an OCR system for recognising handwritten Arabic characters. The paper describes the importance of selecting suitable feature selection methods, and it proposes an approach for optimising the feature selection process in OCR

systems, which improves recognition accuracy and reduces computational requirements. It employs a hybrid machine learning approach that combines the Binary Whale Optimisation Algorithm (BWOA) with the Neighbourhood Rough Set (NRS) technique to select the most suitable features for constructing an OCR system for Arabic characters. The method is evaluated using the CENPARMI dataset, which reveals that the BWOA-NRS approach outperforms alternative feature selection methods and even deep neural networks (DNNs). The system gives perfect accuracy and efficiency in recognising handwritten Arabic characters. The system's results demonstrate its capability to automate tasks such as address extraction from mail or buildings. This paper also makes a noteworthy contribution to the domain of optical character recognition.

The research paper "Survey of Post-OCR Processing Approaches" by Nguyen et al. (2021) examines the challenges of extracting text from printed or handwritten documents and images using OCR. OCR works well with modern text processing, but it struggles with historical materials and documents. OCR errors have a significant impact on information retrieval and natural language processing. This paper discusses the importance of post-processing, which can be applied to English and other Latin-script languages to improve the accuracy of extracted text in OCR. The paper examines various post-OCR processing methods, including dictionary-based methods, rule-based methods, statistical machine translation (SMT), neural machine translation (NMT), language models (LMs), and hybrid methods. It also discusses different metrics for assessing the effectiveness of post-OCR processing, various benchmark datasets available, relevant

language resources that can be used for post-processing and helpful tools. This paper is also relevant to our literature review, as it provides insight into the importance of reducing OCR errors and the various post-OCR processing techniques that can be employed.

In the paper "OCR-IDL: OCR Annotations for Industry Document Library Dataset" by Biten et al. (2023), the authors created a large-scale dataset of OCR annotations to improve the evaluation process of OCR engines. They discuss the issue of using different OCR engines and varying amounts of data, which will ultimately create difficulties in comparing and evaluating models accurately. As a solution, they created OCR-IDL, a dataset comprising over 26 million pages and 166 million words, with an average of approximately six pages and 62.5 words per page of OCR annotations obtained using Amazon Textract's OCR engine. OCR-IDL is the largest publicly available OCR annotated dataset created using a commercial OCR of that time. It contains a wide variety of document types, dates and layouts. The dataset is, therefore, more representative of real-world documents, thereby improving the robustness and generalisation of document intelligence models. They curated and released this dataset to provide a standardised resource for OCR evaluation.

The paper "OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment" by Hegghammer (2022) gives reports on an experiment that compares famous OCR engines Tesseract, Amazon Textract, and Google Document AI

on English and Arabic text images. It also discusses the potential of OCR in the social sciences and humanities for utilising large historical documents. It also discusses the challenges in OCR technology, such as accuracy differences across different OCR engines and the tedious preprocessing, training, and post-processing required to achieve satisfactory results. It also discusses various limitations for building OCR processors, such as noise, complex layouts, and non-Western languages. It highlights the recent advancements in AI that have improved standalone OCR engines and the emergence of server-based OCR engines, such as Amazon Textract, Tesseract, and Google Document AI. The paper also discusses the need for statistically meaningful measurements to evaluate general OCR processors on document types commonly used in research. It then points out that existing benchmarking studies are either outdated or focus on document types (such as forms and receipts) used to design commercial OCR engines, neglecting historical document types in the social sciences and humanities. This paper addresses these gaps by conducting a benchmarking experiment to compare the performance of Tesseract, Textract, and Document AI on English and Arabic papers with varying levels of noise.

The Paper "Reading Scene Text in Deep Convolutional Sequences" by He et al. (2016) proposes a novel approach to scene text recognition. It frames text recognition as a sequence labelling problem, unlike traditional character-by-character recognition. The paper introduces Deep-Text Recurrent Network (DTRN), which is a combination of deep convolutional neural networks (CNNs) and recurrent neural networks (RNNs). It points

out that traditional character-by-character recognition needs explicit character segmentation and is difficult in scene text recognition due to variations in text appearance and complex backgrounds. The proposed DTRN utilises CNNs to generate a high-level, ordered sequence of features from the word image, thereby retaining the essential spatial information required for distinguishing between different words that may share similar characters but differ in their order. The DTRN utilises RNNs, specifically Long Short-Term Memory (LSTM) networks, to capture dependencies within the sequence, thereby enabling the recognition of ambiguous characters by considering the surrounding context. That is, a bidirectional LSTM network is used to decode the CNN sequence by accessing both past and future contexts to make better predictions. The DTRN has a CTC layer that addresses the disparity between the lengths of the CNN sequence and the target word string, allowing it to handle words of varying lengths and eliminating the need for a predefined dictionary. Thus enabling it to recognise unknown words and arbitrary strings. The proposed DTRN architecture was evaluated on standard benchmarks, including the Street View Text, ICDAR 2003, and IIIT 5K-word datasets, and the results demonstrated significant improvements over state-of-the-art methods. The proposed method's ability to handle ambiguous and unknown words is noteworthy. Overall, the paper discusses how traditional text recognition works, its limitations, and proposes an approach to scene text recognition by combining CNNs and RNNs in a unified framework, DTRN.

The paper "Real-Time Scene Text Detection with Differentiable Binarization" by Liao et al. (2020) proposes a novel method for real-time scene text detection, which

involves locating and identifying text within images. This approach gives good results even if the text in the image is curved or at an angle. Localising text in scene images presents various challenges due to variations in text scales, shapes (horizontal, multi-oriented, and curved), and languages. The paper then discusses two main types of methods usually used for this task: regression-based methods and segmentation-based methods. Regression-based methods directly predict the location of text in an image. They are fast methods but do not work well with text that is not straight. Segmentation-based methods operate at the pixel level, allowing them to classify each pixel as either text or background. They can handle different text shapes better, but they are often slower due to the complex processing involved. The paper discusses the binarisation problem in segmentation-based methods, which is the process of converting initial pixel-level predictions into distinct text regions, known as binarisation. Binarisation can be computationally expensive and impact the overall speed of text detection. To resolve this, the paper introduces a differentiable binarisation (DB) module which integrates the binarisation process directly into the training process of the neural network. This adds several advantages to the system, such as adaptive thresholding, simplified post-processing, enhanced performance, and real-time speed. Overall, the paper proposes a novel approach for detecting text in images by making the process of distinguishing text from the background more flexible and adaptive.

In the paper titled "A Real-Time Automatic Plate Recognition System Based on Optical Character Recognition and Wireless Sensor Networks for ITS" by Dalarmelina et

al. (2019), authors introduce a real-time Automatic License Plate Recognition (ALPR) system using OCR technology. The paper highlights the growing importance of ALPR in intelligent transportation systems (ITS) due to the increasing number of vehicles and the need for efficient traffic management. It also discusses the challenges of developing accurate ALPR systems for real-world scenarios such as varying lighting and plate formats. It discusses prior research on ALPR and highlights that many existing methods are not designed for real-time application and struggle with diverse plate formats and lighting conditions. It utilises the existing SPANS framework and integrates a new Automatic License Plate Recognition (ALPR) system to detect vacant parking slots, capture vehicle images, and perform real-time license plate number identification. It also provides future directions to ALPR, such as incorporating deep learning methods to increase accuracy.

The paper "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review" by Memon et al. (2020) gives an analysis of the development of OCR systems for handwritten text, and the authors used 176 studies published between 2000 and 2019 on this. They employed a systematic methodology for selecting and evaluating studies, utilising specific inclusion and exclusion criteria and quality assessment protocols. They categorise OCR techniques into traditional methods such as template matching, structural pattern recognition, and statistical models like Hidden Markov Models (HMM) and k-Nearest Neighbors (kNN), as well as modern machine learning and deep learning approaches, including Support Vector Machines

(SVM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN). A significant trend they observed is the transition from handcrafted feature extraction to deep learning-based end-to-end systems. Deep learning-based systems helped to improve accuracy and adaptability in OCR. The study also provides a language-wise analysis, highlighting that English has the highest volume of research, followed by Arabic, Indian scripts, Chinese, Urdu, and Persian. Each language has unique challenges due to script complexity and dataset availability. The paper also highlighted the critical role of standardised datasets, such as MNIST, IAM, and IFN/ENIT, in model evaluation and performance benchmarking. It concludes by emphasising the growing importance of deep learning in OCR and the need for more research in underrepresented languages and real-time OCR systems to meet diverse practical applications, specifically historical document preservation, education, and accessibility for the visually impaired.

The paper "Calamari – A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition" by Wick et al. (2018) introduces a Calamari robust OCR system built on TensorFlow, designed explicitly for line-level text recognition in both contemporary and historical printed documents. It employs a hybrid architecture that combines Convolutional Neural Networks (CNNs) and bidirectional Long Short-Term Memory (LSTM) networks, which are trained using the Connectionist Temporal Classification (CTC) loss. It supports GPU acceleration, enabling high-speed training and prediction, and can process a single line in as little as 3 milliseconds. The key features of it include confidence-based voting, model pretraining with codec

adaptation, and dropout regularization for improved generalization. Additionally, in the benchmark experiments using the UW3 (modern English) and DTA19 (German Fraktur) datasets, Calamari achieved superior results, with character error rates (CERs) of 0.11% and 0.18%, respectively. It outperformed other open-source tools, such as Tesseract 4, OCRopy, and OCRopus 3. It was also tested on historical books from the 15th century using just 50 ground truth lines for training and has shown substantial accuracy improvements through pretraining and voting. The paper also describes the use of Calamari in other sequence prediction tasks, such as music notation and speech-to-text transcription. The paper concludes by highlighting that Calamari's combination of accuracy, speed, and adaptability positions it as a leading OCR tool for both academic and practical use cases.

The paper "Quranic Optical Text Recognition Using Deep Learning Models" by Mohd et al. (2021) proposes a specialised approach for recognising Quranic Arabic text using advanced deep learning techniques. It aimed to address the distinct challenges posed by the Quranic script, such as complex ligatures, diacritics, and ornamental typography. The authors develop and evaluate several deep learning models particularly focused on CNN-BiLSTM architectures paired with Connectionist Temporal Classification (CTC) loss. Their methodology involves preprocessing Quranic text images through normalisation, binarisation, and augmentation, followed by line-level recognition. They curated a dataset of high-resolution scanned pages of the Quran, which was used for training and evaluation. The testing of the CNN-BiLSTM-CTC setup yields

the lowest character error rate, demonstrating its effectiveness. The paper highlights the use of transfer learning to boost recognition accuracy, where pre-trained models for general Arabic OCR tasks are fine-tuned for Quranic text. The paper shows that deep learning works well for accurately reading the Quran using OCR. The study also suggests that we need more comprehensive collections of Quranic text data. It suggests exploring more advanced computer models to handle the complexity of Quranic writing.

The paper "An Application of Deep Learning in Character Recognition: An Overview" by Saeed et al. (2019) describes how deep learning has revolutionised the field of character recognition. The authors explained the shortcomings of traditional OCR methods and highlighted how deep learning, particularly Convolutional Neural Networks (CNNs), has made a more robust and accurate recognition of both printed and handwritten characters. The paper discusses a variety of deep learning techniques, including CNNs for spatial feature extraction and Recurrent Neural Networks (RNNs) for modelling character sequences. Hybrid architectures that combine CNN and RNN layers can recognise complex text input and improve recognition outcomes. The paper acknowledges challenges in this area, such as the need for extensive labelled datasets, high computational costs, and difficulties handling noisy or distorted input. Popular datasets, such as MNIST, EMNIST, and IAM, are often referenced as benchmarks. The paper then discusses how OCR is applied in real-life scenarios. This includes scanning documents, recognising car license plates, and making things easier for people with disabilities. In summary, the paper describes how deep learning has significantly

improved OCR. It also suggests that more research is needed to improve OCR, particularly for languages not commonly used and for applications that require instant recognition.

The paper "A Detailed Analysis of Optical Character Recognition Technology" by Hamad and Kaya (2016) gives a thorough examination of OCR systems, their evolution, core architecture, and applications. The paper begins by describing the critical need for OCR in digitising printed and handwritten documents. It then explains the numerous challenges faced in practical OCR, including issues such as scene complexity, skew, lighting variation, blurring, and multilingual scripts. The paper describes the OCR pipeline as comprising six stages: preprocessing, segmentation, normalisation, feature extraction, classification, and postprocessing. Each of these stages is described with its standard techniques and difficulties. For classification, multiple algorithms, including template matching, statistical methods, neural networks, and SVMs, are compared with real-world application examples. It then described OCR's wide-ranging utility across various sectors, including banking, healthcare, legal systems, and mobile applications. The paper also describes the history of OCR, from early mechanical recognition systems to modern, intelligent neural network-based systems which are capable of handling diverse languages and document conditions. It concludes by emphasising the continued relevance of OCR and proposes future directions, such as mobile integration and daily-use applications, including automated receipt tracking and book readers.

2.2 Natural Language Processing (NLP)

For an AI-based visa processing system, there should be clear communication between human and machine language. Natural Language Processing (NLP) can help with this problem as it enables the understanding and processing of human language, thereby extracting meaningful information from visa-related documents. Natural Language Processing (NLP) has undergone significant advancements in recent years. It gives new dimensions for language understanding and computational linguistics. It aids in identifying the key details in the text, interpreting the text, and facilitating effective communication between the system and human language.

A critical contribution to this field is the paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Devlin et al. (2018). This paper introduces a powerful language understanding model called BERT (Bidirectional Encoder Representations from Transformers). BERT can consider both the left and right context of a word during its pre-training phase. This bidirectional training is conducted using a technique known as masked language modelling (MLM). In MLM, a random percentage of the words in the input text is masked, and the model is trained to predict these masked words based on the surrounding context. Thus, BERT learns to represent each word(unlabelled text) in such a way that it captures its meaning compared to all the other words in the sentence. Because of this, BERT can be fine-tuned with just an additional output layer to create many state-of-the-art models for a wide range of tasks at that time, such as question answering and language inference. This paper examines

different ways to use language understanding that computers have already learned. These methods include feature-based approaches and fine-tuning approaches.It highlights that current techniques have limited the ability to utilise pre-trained representations for specific tasks, particularly in fine-tuning approaches. This paper also argues for the use of a document-level corpus rather than a shuffled sentence-level corpus for pre-training in order to extract long, contiguous sequences. BERT's adaptability to specific tasks through fine-tuning is far better than that of its predecessors. Overall, this paper explains the groundbreaking BERT model in the field of NLP.

The paper "Attention is All You Need" by Vaswani (2017) discusses prior research in sequence modelling and transduction problems, primarily within the domain of natural language processing (NLP). It discusses traditional recurrent and convolutional models, highlighting their limitations in terms of parallelisation and capturing long-range dependencies. It then proposes the Transformer, a novel neural network architecture that relies solely on self-attention mechanisms for sequence transduction tasks without using sequence-aligned RNNs or convolution. It discusses the Transformer's architecture in detail, including the components such as the encoder and decoder stacks, the attention mechanism, and the positional encoding. It also discusses the concepts of scaled dot-product attention and multi-head attention and how they are utilised in the Transformer. The Transformer dynamically weights different parts of the input sequence, effectively capturing complex relationships. The paper discusses the self-attention mechanism and compares it with recurrent and convolutional layers in terms of

computational complexity, parallelizability, and the ability to learn long-range dependencies. Transformer's ability to parallelise computations causes significant speedups during training, especially for long sequences. The paper also gives experimental results on machine translation tasks to demonstrate the Transformer's state-of-the-art performance in English-to-German and English-to-French translation. It also compares the Transformer's training costs to other models and highlights its efficiency. The paper examines the impact of various architectural choices, including the use of different numbers of attention heads, attention key dimensions, model size, and dropout rate. It also investigates the use of learned positional embeddings instead of sinusoidal positional encodings to see if it affects the model's ability to handle sequences of different lengths. These alterations that are made to the base Transformer model architecture resulted in varying degrees of impact on its performance. These observations point out the importance of careful hyperparameter tuning and architectural choices in optimising the Transformer's performance. The paper also discusses the application of the Transformer architecture to NLP tasks, such as machine translation, reading comprehension, and summarisation. Overall, this paper discusses various techniques in NLP tasks and proposes an architecture to solve the limitations of current techniques.

In the paper "XLNet: Generalized Autoregressive Pretraining for Language Understanding", Yang et al. (2019) proposes a novel unsupervised learning approach to pretrain language models which can address the limitations of existing methods like BERT and autoregressive models. The paper discusses autoregressive language (AR)

models and highlights their limitations in capturing deep, bidirectional context. It highlights that AR models are trained to encode text in only one direction, either forward or backwards. This reduces their ability to fully understand the context surrounding a word or phrase, which often requires both preceding and following text. The paper then discusses autoencoding (AE) based pretraining, particularly BERT, which addresses the bidirectional context limitation of AR models. However, BERT introduces a pre-train-finetune discrepancy due to the reliance on masking input tokens and an oversimplified independence assumption among predicted tokens. The paper proposes XLNet as a solution that aims to bridge the gap between AR and AE approaches by leveraging both strengths while mitigating their weaknesses. The proposed solution draws inspiration from previous work on permutation-based AR modelling, such as orderless NADE. It differentiates itself by incorporating bidirectional context learning and introducing technical innovations, including two-stream self-attention for target-aware representations. The proposed method also draws inspiration from Transformer-XL, the state-of-the-art AR model at the time, to enhance XLNet's ability to handle more extended text sequences. It also incorporates techniques such as segment recurrence and relative encoding further to refine the model's architecture and pre-training approach. The paper also presents experimental results demonstrating that XLNet outperforms BERT, a widely used pre-training method, on various natural language understanding tasks. Overall, the paper provides an overview of AR language modelling, autoencoding, and permutation-based approaches, as well as the limitations of each in NLP tasks. It also proposes a solution, the XLNet pre-training method, which

addresses the limitations of existing approaches, such as BERT and autoregressive language models.

In "Distributed Representations of Words and Phrases and their Compositionality" by Mikolov et al. (2013) discusses the Skip-gram model and introduces new techniques to improve its performance and capabilities. Skip-gram is a neural network-based method for learning word and phrase representations as numerical vectors. This paper introduces three main techniques to improve Skip-gram-based learning. Initially, it introduces subsampling, a technique that reduces the influence of prevalent words during training. This enables the model to focus on establishing more meaningful relationships between less frequently used words. Then, it introduces negative sampling for calculating word probabilities as a more efficient alternative to the hierarchical softmax method used in the original Skip-gram model. Negative sampling enables the model training process to be faster without compromising the quality of the learned word representations. The paper also introduces a new technique called phrase representations, which enables the model to learn the meaning of groups of words, not just individual words. This means it can understand phrases where the meaning is more than just the sum of its words. Together, these new techniques improve the Skip-gram model. They make it better and faster to learn important representations for both individual words and multi-word phrases from large amounts of text data. It also highlights the effectiveness of these techniques through the results of some analogical reasoning tasks, which evaluate how well the model captures relationships between words. The paper also introduces Word2vec, which is a

toolkit that utilises techniques such as subsampling and negative sampling. Overall, this paper not only presents novel ideas but also provides practical tools and techniques that the NLP community has widely adopted.

The paper titled "Sequence to Sequence Learning with Neural Networks" by Sutskever et al. (2014) gives a new method to use deep neural networks for tasks where input and output are sequences, such as language translation. It uses a Long Short-Term Memory (LSTM) neural network to handle long sequences of words. This approach aims to overcome the limitations of deep neural networks (DNNs) when mapping sequences to sequences. The paper proposes a model architecture that employs a multilayered LSTM. The proposed architecture comprises two LSTM networks: the first is used for encoding the input sequence (in this case, English) into a fixed-length vector, and the second is used for decoding that vector into the output sequence (in this case, French). The advantage of using LSTM in sequence-to-sequence models is that it can learn meaningful representations of phrases and sentences, capturing word order while remaining relatively unaffected by changes between active and passive voice. The paper points out that reversing the order of words in the source sentences significantly improves LSTM performance, demonstrating the effectiveness of LSTMs in learning due to their short-term dependency nature. The paper also points out the potential of the sequence-to-sequence approach beyond machine translation. Overall, this paper introduces a sequence-to-sequence architecture for language translation.

The book 'Neural Network Methods for Natural Language Processing' by Goldberg (2022) gives an overview of the application of neural network methods in natural language processing (NLP). It provides an overview of the historical evolution of deep learning, the fundamentals of neural networks, and how neural networks impact NLP. It then discusses the challenges and opportunities that come with natural language data. It also discusses various types of neural networks, such as CNNs and RNNs, which are commonly used for handling language tasks. Finally, it touches on some advanced topics in the field. It serves as a bridge between deep learning and NLP, providing practical guidance and theoretical understanding to both the NLP and deep learning communities. Although it covers only fundamental concepts in deep learning and NLP, it excels in explaining the fundamentals of NLP and how they relate to neural networks. Overall, this book provides a comprehensive approach to the application of neural networks in NLP, making it an excellent starting point for anyone interested in studying the applications of neural networks in natural language processing.

The paper "Visualising and Understanding Neural Models in NLP" by Li et al. (2015) explores techniques for interpreting the internal workings of neural models for natural language processing (NLP) tasks. It utilises visualisation tools from computer vision to examine how neural models function. It discusses techniques such as representation plotting to illustrate how meanings are combined and saliency analysis to understand which parts of the input are most important for each model's final decision. It then analyses models such as RNNs, LSTMs, and Bi-LSTMs on tasks like sentiment

analysis and sequence-to-sequence learning, utilising these techniques. It visualises how neural models understand the meaning of phrases and words, capturing subtle language differences. It also shows how neural models in NLP interpret linguistic nuances, including negation, intensification, and concessive clauses. It highlights the ability of LSTM to focus on important words and disregard less significant ones. It discusses previous studies that use visualised neural networks and uses them for model interpretation. It also discusses how other researchers have attempted to interpret NLP models and how texts are processed internally within these models. Overall, this paper makes a significant contribution to the interpretability of how neural networks function when processing human language.

The paper "Deep Learning Applied to NLP" by Lopez et al. (2017) examines the application of Convolutional Neural Networks (CNNs) in Natural Language Processing (NLP). It discusses the evolution of Convolutional Neural Networks (CNNs), which are traditionally associated with Computer Vision, and the increasing demand for CNNs in Natural Language Processing (NLP) tasks. It then explains the basics of CNNs, their structure, variations, and how they can be adapted to process textual data. It then explains the motivation for using CNNs in NLP, such as their effectiveness in complex learning tasks and increased accessibility through high-performance computing resources and open-source libraries. It also discusses challenges in NLP, such as ambiguity and complexity of human language. It then discusses various deep learning techniques, notably different types of Recurrent Neural Networks (RNNs), including Bidirectional

RNNs, Deep RNNs, and LSTM networks, as well as Recursive Neural Networks (RCNNs) and Dependency-based Neural Networks (DCNNs). It then discusses the application of CNNs in various NLP tasks, such as sentiment analysis, semantic clustering, entity linking, event extraction, and machine translation, thereby highlighting the versatility of CNNs in addressing different NLP problems. It also discusses the use of CNNs in speech recognition tasks, such as phoneme recognition and voice search, and their ability to capture and model complex acoustic patterns. Overall, this paper provides a comprehensive overview of CNNs, specific CNN architectures, and their potential in NLP.

The paper titled "On the Explainability of Natural Language Processing Deep Models" by Zini et al. (2022) studies different Explainable AI (ExAI) methods in Natural Language Processing (NLP). It highlights the challenges posed by the black-box nature of deep learning models in NLP. The inherent complexity of natural language encompasses nuances such as polysemy, sarcasm, slang, cultural references, and ambiguity. Additionally, while word embeddings effectively capture semantic and syntactic relationships between words, they are often high-dimensional and dense. The lack of transparency in how word embeddings affect a model's decision adds another layer of complexity. The inherent complexities of natural language, the high-dimensional, dense, and opaque nature of word embeddings, as well as difficulties in visualising the internal workings of the model make the interpretability and transparency of the NLP model somewhat challenging. It discusses various methods for

making NLP explainable. It discusses both model-agnostic explainability methods (methods that work on any machine learning model) vs. model-specific explainability methods (methods that are tailored for a particular kind of model, like a neural network), inherently interpretable methods (models are designed from the start to be easy to understand) vs post-hoc methods (methods try to explain to existing complex models after they have been trained) and level of explanation. The level of explanation methods can be input level (word embeddings), processing level (internal representations), or output level (model decisions). The paper explores techniques such as sparsification, rotation, and the integration of external knowledge to enhance interpretability at the input level of explanation. It highlights that most evaluation methods used in the industry focus on embedding quality rather than explainability, and hence, there is a need for more standardised assessment frameworks. The paper concludes by highlighting the need for more inherently interpretable models, improved evaluation frameworks, and explanations that account for the hierarchical nature of text understanding. Overall, the paper explains the methods used, limitations, and future directions for explainable AI.

The paper "BERT rediscovers the classical NLP pipeline" by Tenney et al. (2019) studies how linguistic information is represented within BERT (Bidirectional Encoder Representations from Transformers), a famous pre-trained text encoder developed by Google AI. It analyses previous research that has employed probing techniques to investigate the interpretability of the BERT language model. It then points out that BERT primarily processes natural language like the classic step-by-step approach employed in

traditional NLP pipelines. Initially, BERT identifies the part of speech for each word and understands the sentence structure. It then performs more complex tasks, such as determining the roles of different words in a sentence and recognising which words refer to the same entity (like pronouns). Finally, it comprehends the whole piece of text. Even though BERT generally follows the traditional NLP pipeline order, it sometimes makes dynamic pipeline adjustments, often revising the traditional NLP pipeline orders. Overall, this paper provides a comprehensive overview of the internal organisation of linguistic information within the BERT model and its ability to capture complex interactions across various levels of hierarchical information.

The paper "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" by Raffel et al. (2020) studies different transfer learning techniques in natural language processing (NLP). Transfer learning helps optimise the resources used to build AI models, thereby enhancing their performance. It involves two steps: first, pre-train a model on a data-rich task and then fine-tune it on a specific downstream task. The paper studies pre-training objectives, architectures, unlabeled datasets, and transfer approaches across multiple language understanding tasks. It also reviews previous research to establish a common framework for various NLP tasks, including question answering, language modelling, and span extraction. The paper introduces a unified text-to-text framework for standardising diverse text-based language problems. It explains the effectiveness of the Transformer architecture (Vaswani et al., 2017) in various NLP tasks, as well as the Transformer's self-attention mechanism and its

encoder-decoder structure. The paper also acknowledges that the Transformer architecture serves as the foundational architecture for the models used in the paper. It then explains the need for clean and natural language text for effective pre-training and introduces "Colossal Clean Crawled Corpus" (C4), a cleaned dataset derived from the Common Crawl website. It also discusses the challenges associated with fine-tuning and training a model on multiple tasks simultaneously. It concludes that bigger models and more data generally lead to better performance. The paper gives a comprehensive understanding of transfer learning in NLP and its potential for achieving general language understanding.

The research paper "Dynabench: Rethinking Benchmarking in NLP" by Kiela et. al (2021) introduces the innovative Dynabench platform for creating datasets and benchmarking models in Natural Language Processing (NLP). It is an open-source, web-based system that aims to revolutionise NLP benchmarking by offering a dynamic approach to dataset creation and model evaluation. The paper highlights the limitations of current static benchmarks by presenting impressive performance results for various NLP models, but it struggles with challenging examples and real-world scenarios. The paper discusses the need to reconsider benchmarking methodologies in light of the rapid advancements in machine learning models. Unlike static benchmarks, Dynabench employs human and model collaboration, where annotators can actively generate examples to reveal misclassifications by a target model. In Dynabench, models can be evaluated in real-time against human-generated examples, which reveals models'

weaknesses and guides further development. Also, adversarial data collection is possible in the Dynabench platform. This is due to the involvement of humans, who will actively try to create examples that the model will misclassify. Models will receive real-time feedback, indicating whether they were fooled or not. This will create a sort of game where humans are constantly trying to find the model's weaknesses and exploit them, and the model will use these examples to become better and better, just like humans. Another advantage of Dynabench is that it is an open-source platform. Thus, it can foster collaboration and data sharing, leading to accelerating research and development. This paper introduces Dynabench, a paradigm shift in NLP benchmarking, by addressing the dynamic nature of language tasks.

The paper "Measure and improve robustness in NLP models: A survey" by Wang et al. (2021) explores the robustness in natural language processing (NLP) models, including its definition, evaluation methodologies, and mitigation strategies. Robustness in NLP refers to a model's ability to handle both adversarial attacks (synthetic distribution shifts) and naturally occurring distribution shifts. The core idea of robustness in NLP is to ensure that models generalise well from training data to diverse test data. The paper discusses three methods for identifying weaknesses in NLP models: human-prior, error-analysis-driven, and model-based identification. Human-prior and error analysis utilises human knowledge and the ability to analyse model errors and identify patterns of vulnerability. The advantage of this approach is that it can capture weaknesses that automated methods may not identify. However, this approach is time-consuming and is

subjected to human biases. The model-based identification approach utilises computational methods to identify robustness failures automatically. It utilises task-agnostic or input-agnostic methods, such as white-box attacks or learning additional models to capture biases. The paper discusses three main strategies for enhancing robustness. They are the data-driven approach, the model and training-based approach, and the inductive-prior-based approach. The data-driven approach augments or modifies the training data to expose the model to a broader range of scenarios, thereby reducing reliance on artificial correlations. The model- and training-based approach modifies the model architecture, training process, or loss functions to achieve better generalisation and robustness. Inductive-prior-based approaches introduce inductive biases into the model to guide it away from learning unwanted features and towards more robust representations. The paper also highlights the importance of understanding the connection between human-like linguistic generalisation and NLP generalisation to achieve more meaning and robustness. Overall, the paper gives a comprehensive overview of the understanding and improving the robustness of NLP models.

The paper "Show and Tell: A Neural Image Caption Generator" by Vinyals et al. (2015) focuses on image captioning. This specific computer vision task involves generating natural language descriptions of images. It discusses how early image captioning systems often relied on complex pipelines involving primitive visual recognisers, structured formal languages, and rule-based systems. These systems are often limited by their hand-designed nature, brittleness, and applicability to restricted

domains. The paper notes a growing interest in creating natural language descriptions for still images. This interest is driven by recent progress in areas like recognising objects, identifying their characteristics, and pinpointing their exact locations within an image. It then introduces the Neural Image Caption (NIC) model, an image captioning model that combines convolutional neural networks (CNNs) for image understanding and recurrent neural networks (RNNs) for language generation. The NIC model can generate accurate and coherent image captions by seamlessly integrating convolutional neural networks (CNNs) and recurrent neural networks (RNNs) within a deep recurrent architecture. This approach takes inspiration from the successes of sequence generation in machine translation. The paper also highlights the end-to-end trainable nature of the proposed model and its ability to leverage pre-training on larger corpora. The proposed model outperformed state-of-the-art methods at the time, particularly in terms of BLEU scores on various benchmark datasets. Overall, this paper examines the image captioning task, which combines computer vision and natural language processing, and proposes a novel approach to handle it efficiently.

The paper titled "Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model" by Sharifani et al. (2022) studies the growing interest in fake news detection and the challenges associated with it. It discusses the increasing number of fake news on online platforms, particularly on social media, and the need for the development of automated systems to assess the truthfulness of them. It then discusses the potential of machine learning and artificial

intelligence in addressing the fake news problem. It highlights that at first, research focused on deception detection and classification of online content, and the specific challenge of fake news detection gained significant interest only after the 2016 US Presidential elections. It also points out that simple content-based classification techniques such as n-grams and part-of-speech tagging are insufficient for fake news detection as they don't provide a deeper analysis of content by incorporating contextual information. It then highlights that context-free grammar (CFG), in combination with n-grams, has shown promising results in deception-related classification tasks. It also points out the effectiveness of using the relative frequency of words as a distinguishing factor between fake and non-fake news. The paper concludes by stressing the importance of integrating machine learning and NLP for robust fake news detection. It also argues for the need for more sophisticated approaches beyond simple text classification. The paper suggests several directions for future research in detecting fake news. These include adding more types of information, using advanced deep learning methods, and creating automated systems to check facts. It also recommends looking at the problem of fake news from many different angles.

The paper "Multi-task deep neural networks for natural language understanding" by Liu et al. (2019) combines two prominent approaches —multi-task learning (MTL) and language model pre-training — in natural language understanding (NLU). MTL is a method of training models on multiple tasks simultaneously, leveraging shared knowledge and human learning processes that inspire this approach. The paper points out

that recent research has increasingly applied MTL to deep neural networks (DNNs) for representation learning. Language model pre-training uses vast amounts of unlabeled data to learn universal representations. ELMo, GPT, and BERT employ a language model pre-training approach, which involves training on unsupervised objectives, such as masked word prediction and next-sentence prediction. Fine-tuning with task-specific training data is often required to apply these pre-trained models to specific NLU tasks. The paper points out that these approaches are complementary and highlights the benefits of combining them. It also proposes a novel Multi-Task Deep Neural Network (MT-DNN) that utilizes BERT's pre-trained representations and MTL's regularization to achieve state-of-the-art results on various NLU tasks.

The paper "Transfer learning in natural language processing" by Ruder et al. (2019) gives an overview of modern transfer learning techniques and their applications in NLP. It highlights the limitations of traditional supervised learning, particularly in scenarios where labelled data is limited. It gives a clear definition of transfer learning and explains how it leverages knowledge from other tasks or domains to improve performance on a target task. It also explains unsupervised, supervised, and distant supervision methods for pretraining language models. It then discusses how to probe and analyze the representations learned by pre-trained models and provides an overview of various techniques for adapting pre-trained models to specific tasks. It highlights the application of transfer learning in various NLP tasks, including text classification, natural language generation, and structured prediction. It also explains the state-of-the-art

transfer learning techniques and tools of that time, enabling the practical application of these techniques. Overall, the paper gives an overview of transfer learning in Natural Language Processing (NLP), its methods, and applications.

The paper "Language (technology) is power: A critical survey of" bias" in nlp" Blodgett et al. (2020) analyses "bias" in Natural Language Processing (NLP) systems. It analyses 146 papers on bias in language technology and finds out that their motivations are often vague, inconsistent, and lack normative reasoning. The paper discusses the disparity between their quantitative techniques for mitigating bias and their stated motivations. It also notes that these papers often fail to connect with important research outside of NLP. This outside research examines the relationship between language and social power. The paper examines how language can perpetuate stereotypes and unfairness by looking into the relationship between language and social power. It introduces the concept of "controlling images," which refers to powerful groups attempting to influence language to maintain their influence. The authors highlight that understanding how language and social power are linked is important for finding and analyzing bias in AI language systems (NLP).

The paper "A survey on bias in deep NLP" by Garrido-Muñoz et al. (2021) also studies the bias in Deep Natural Language Processing (Deep NLP) models. It highlights the fact that while deep neural networks have become increasingly popular in NLP, inherent biases are present in the training data. Bias in pre-trained models used for

transfer learning is significant in Deep NLP models. This paper examines various approaches to bias measurement and mitigation, including association tests such as the Word Embedding Association Test (WEAT) and its extensions. It then discusses techniques for addressing bias in translation and coreference resolution tasks. It also discusses studies on bias in the GPT-3 model, although they were not publicly published at the time of this paper's publication. The paper even explored its potential for generating biased content and the ways to reduce this through positive context. It then discusses techniques for debiasing vector spaces. The paper then presents a general approach to reducing bias in language models, which first defines what bias is in this context, then examines whether bias can exist in the model, analyses the results, and finally identifies ways to mitigate it. It also encourages us to be open about what we did and what we found. Overall, the paper advises us to prevent bias rather than just rectify it later.

The paper "Evolution of Transfer Learning in Natural Language Processing" by Malte et al. (2019) examines the evolution of transfer learning in natural language processing (NLP). It discusses rule-based and statistical methods, which have limitations in handling the complexity and nuance of natural language. The paper then discusses machine learning algorithms and models, such as Naive Bayes, decision trees, bag-of-words, and n-grams, and how they have improved NLP tasks. It emphasises the importance of two breakthroughs: transfer learning and advancements in language models. NLP has undergone significant evolution with the advent of transfer learning and

the development of improved language models. It discusses the evolution of transfer learning in NLP, starting with ULMFiT, which introduced the concept of fine-tuning pre-trained language models for specific tasks. It then discusses ELMo, which introduced contextual word embeddings based on bidirectional LSTMs. OpenAI's Transformer architecture, which utilises the self-attention mechanism for enhanced sequence modelling, is also discussed in the paper. The paper further discusses BERT, a bidirectional Transformer model that incorporates bidirectionality and achieves state-of-the-art results in various NLP tasks. It also discusses Universal Sentence Encoder, which uses direct encoding of sentences into vectors, and Transformer-XL, which overcame the limitation of fixed-length context in Transformers. The paper highlights XLNet, a model that overcomes the limitations of BERT and incorporates Transformer-XL for improved long-range dependency modelling. It also discusses the emergence of lighter models, such as DistilBERT and ALBERT, as well as training methods like RoBERTa. This paper highlights the significant advancements in transfer learning for NLP, the transition from traditional methods to sophisticated models such as BERT and XLNet, and the current research directions toward developing lighter and more efficient models.

The paper "A Comprehensive Survey of Deep Learning Techniques Natural Language Processing" by Bharadiya (2023) discusses the importance of unsupervised and semi-supervised learning techniques in Natural Language Processing (NLP) due to their ability to learn from unlabeled data. It delves into the historical evolution of NLP,

from rule-based systems to statistical methods and, more recently, deep learning approaches. The integration of deep learning with large corpora and pre-trained models revolutionized the field of NLP. The paper also discusses class-based language modelling (CBLM), which can handle out-of-vocabulary words and improve vocabulary efficiency. It then discusses the potential of big data in advancing various NLP areas, including domain-specific NLP, multilingual NLP, cross-lingual NLP, and NLP techniques such as machine translation, sentiment analysis, named entity recognition, and speech recognition. Overall, this paper provides an overview of NLP, its evolution from rule-based systems to deep learning approaches, the importance of big data, and unsupervised and semi-supervised learning techniques in NLP.

The paper "A survey of data augmentation approaches for NLP" by Feng et al. (2021) discusses data augmentation (DA) techniques in natural language processing (NLP). The paper also discusses the importance of DA as there is an increasing demand for diverse training data in NLP. The paper also highlights the challenges encountered in NLP due to the discrete nature of language data. Therefore, direct continuous augmentation methods cannot be used in NLP, as they are typically employed in computer vision tasks. It highlights that, despite DA providing solutions for low-resource languages and bias, its effectiveness remains underexplored, especially for pre-trained models. It also states that the future of DA lies in multimodal approaches, self-supervised learning, and standardized benchmarks, which can pave the way for more robust and effective NLP models.

The paper "How transfer learning impacts linguistic knowledge in deep NLP models?" by Durrani et al. (2020) studied how fine-tuning pre-trained language models on specific tasks affects their linguistic knowledge. It studies popular pre-trained models, such as BERT, RoBERTa, and XLNet, on their linguistic abilities before and after fine-tuning a set of tasks known as GLUE. It points out that the impact on linguistic knowledge varies on the task and model architecture. Some tasks preserve linguistic information more deeply in the network, while others lose this knowledge in lower layers or even forget it. This relegating behavior is prominent in the RoBERTa and XLNet model. The paper analyzed linguistic knowledge at the neuron level. It points out that in the architecture, the linguistic neurons move from the deeper layers to the surface level in models like RoBERTa and XLNet. It also examines how this redistribution affects network pruning and demonstrates that removing bottom layers in RoBERTa and XLNet results in significantly more pronounced performance degradation than in BERT. Overall, this paper provides insight into the relationship between linguistic knowledge, model architecture, and task-specific fine-tuning in deep NLP models.

The paper "Recent trends in deep learning based natural language processing" by Young et al. (2018) gives a comprehensive study of deep learning model trends in NLP. It explains their evolution from word embeddings to complex architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and recursive neural networks. It also explains the concept of distributed representation and the

significance of word and character embeddings in capturing semantic and syntactic information. It then explains the popular word embedding method word2vec, as well as its advantages and limitations. It then explains what CNNs are, their ability to extract n-gram features, and their application in various NLP tasks, including sentiment analysis, question answering, and machine translation. The paper then discusses various types of Recurrent Neural Networks (RNNs), including LSTMs and GRUs. These models work particularly well when information is presented in a logical order (such as the words in a sentence). They can be used for tasks such as determining the meaning of each word in a sentence (word level) and providing a comprehensive overview of a sentence (sentence level), among others. It then explains the attention mechanism-based model and how well it focuses on the most important parts of a sentence. These models are used for tasks such as summarising text or translating it into another language. Similarly, it explains different deep learning models and their strengths in various tasks. It then points out that the selection of deep learning models in NLP depends on the specific semantic requirements of the task. It also discusses unsupervised learning and deep generative models, as well as their applications in NLP. It then outlines the future directions of deep learning in NLP, including the better utilisation of unlabeled data, reinforcement learning, multimodal learning, and the integration of symbolic and sub-symbolic AI.

The paper "Natural language processing (NLP) in management research: A literature review" by Kang et al. (2020) provides a comprehensive overview of the application of Natural Language Processing (NLP) in management research. It discusses

the growing importance of analysing data using NLP in management applications, particularly with the rise of big data and AI. The paper discusses the evolution of NLP, its theoretical foundations, and its recent applications in business and research with examples such as robotic journalism and sentiment-based market predictions. It provides a review of articles published in leading business journals and highlights the significant rise in NLP applications in recent years, particularly in marketing, management science, and strategic management. It explains the challenges and opportunities associated with using NLP in management research. It also highlights the need for high-quality, labelled datasets and emphasises the importance of utilising NLP methods within specific management contexts. Overall, the paper provides an overview of the potential of NLP in management research, highlights the importance of utilising NLP methods for specific tasks, and outlines future directions for achieving better results.

The paper "Don't take the premise for granted: Mitigating artifacts in natural language inference" by Belinkov et al. (2019) studies a significant issue in Natural Language Inference (NLI) datasets: hypothesis-only biases. NLI is a specific task in NLP. The objective of NLI is to find the logical relationship between two sentences (premise and hypothesis). The possible relationships are as follows: if the hypothesis is true, if the premise is true, if the hypothesis is false, if the premise is false, or if there is no clear relationship between the premise and the hypothesis. "Hypothesis-only biases" are a problem in NLI, where computers may learn shortcuts or patterns when attempting to understand the relationship between two sentences. That is, instead of analysing both

sentences together to conclude, the model might focus only on the second sentence (the hypothesis). This is unusual, and it is much like solving a problem by focusing on the second half of the question and disregarding the first half. This renders the model unreliable and less effective in real-world applications. The paper proposes two probabilistic methods to overcome this issue. One approach is to predict the entailment label given the premise and hypothesis rather than the typical approach of predicting the entailment label based on the premise and hypothesis. Their methods, instead, predict the first sentence (the premise) based on the second sentence (the hypothesis) and the logical relationship between them. The second approach shuffles the first sentences (the premise) during training, forcing the model to understand the logical relationship between the two sentences to obtain the correct answer. Both these approaches force the model to consider the premise and hence reduce the impact of hypothesis-only biases. The paper then evaluates the effectiveness of these methods on a variety of datasets, both synthetic datasets and existing NLI datasets. It points out that methods enhance model robustness to quirks or patterns found in the dataset to an extent. This paper makes a valuable contribution to NLI, as it explains the issue of hypothesis-only biases and proposes two approaches to address it.

The paper "A survey on recent approaches for natural language processing in low-resource scenarios" by Hedderich et al. (2020) provides a comprehensive survey of strategies for Natural Language Processing (NLP) in low-resource scenarios. The paper discusses the requirements for specific low-resource settings. It then categorises

low-resource settings based on data availability and techniques used. It categorises data availability into labelled data, unlabeled text, and auxiliary resources. It categorises techniques used for handling low-resource settings into two categories: additional labelled data (data augmentation, distant supervision, cross-lingual projections) and transfer learning (pre-trained language representations, domain-specific and multilingual pre-training). Data augmentation is a technique that creates new labelled data from existing instances, whereas the distant supervision technique utilises external sources for labelling. Transfer learning is a technique that leverages pre-trained language representations and models to reduce the need for labelled target data. It also discusses the strengths and limitations of each approach. This paper gives an overview of the current challenges and techniques available for handling low-resource settings.

The paper "Interpreting recurrent and attention-based neural models: a case study on natural language inference" by Ghaeini et al. (2018) studies the challenge of interpreting deep learning models in Natural Language Inference (NLI) tasks. It studies the ESIM model, which leverages both LSTMs (Long Short-Term Memory Networks) and attention mechanisms for learning. It gives visual representations of word connections and how information flows within the model. It focuses on determining how specific model components contribute to final predictions, and for this, it utilises the saliency of signals on model gates as a tool. It also points out the effectiveness of the proposed approach through a case study. Attention saliency indicates which word connections are most important in determining the final relationship between the premise

and the hypothesis. Attention saliency can give insights into the reasoning process of the model that are not evident in standard attention visualisations. Similarly, the analysis of LSTM gating signals reveals how the model's focus shifts between the input and inference stages, thereby enabling a more comprehensive understanding of the sentence. Overall, this paper helps us understand how neural networks make decisions. It introduces new tools to show why these complex models behave the way they do. These tools also help pinpoint the main factors that influence decisions in Natural Language Inference (NLI) tasks.

2.3 Text Spotting

Text spotting plays a crucial role in AI-based visa processing systems as it assists in identifying specific text within images or documents. This capability is essential for extracting important details such as names, dates, and passport numbers from various document types. This section reviews some important papers on text Spotting. These works have made significant contributions to the field by introducing innovative approaches to text detection, recognition, and segmentation.

The paper "Deep TextSpotter: An End-to-End Trainable Scene Text Localization and Recognition Framework" by Busta et al. (2017) proposes Deep TextSpotter, a novel end-to-end trainable framework for scene text localisation and recognition. Traditional textspotting systems often tackle text localisation and recognition as separate tasks. The paper points out the importance of having an end-to-end trainable system that can

simultaneously localise and recognise text in natural scene images. Because only joint solutions can fully leverage the interplay between the two tasks, it allows the model to learn the inherent dependencies and correlations between the tasks, leading to a more effective solution. Deep TextSpotter effectively adapts and extends techniques from object detection to address the challenges of scene text localisation and recognition. It's Region Proposal Network (RPN), along with rotation prediction, generates region proposals that might contain text. The RPN will predict the position, dimensions, rotation angle, and text/non-text score for each proposed region. It uses a bilinear sampling technique to normalise the detected text regions. This technique avoids the problem of varying text sizes and orientations, preserving the aspect ratio and positioning of individual characters, which is vital for accurate text recognition. It utilises a fully convolutional network (FCN) for text recognition. FCN takes normalised text regions as input and produces a matrix representing the probability distribution over character sequences. The output width of the FCN network dynamically adjusts based on the width of the input region, allowing it to handle text of varying lengths. Finally, the model employs connectionist temporal classification (CTC) to decode the variable-length output of the FCN for the final text transcription. The paper also discusses the limitations of current evaluation protocols for text spotting, such as the DetEval tool, and its emphasis on text localisation accuracy. The paper also suggests some future directions in text spots, such as expanding training datasets to include more diverse and challenging examples, such as blurry or noisy text, single characters, and digits.

The paper "FOTS: Fast Oriented Text Spotting with a Unified Network" by Liu et al. (2018) proposes fast-oriented text spotting (FOTS) , a novel deep learning approach that does text detection and recognition concurrently. The paper discusses previous research in text detection and highlights the increasing demand for deep learning-based methods. It also discusses advancements in text recognition, driven by the increasing use of recurrent neural networks and sequence-to-sequence models. Finally, it discusses text-spotting methods and highlights the limitations of the prevalent two-stage approach. It also discusses the current text spotting approach, which is a two-stage process of text detection and recognition that are treated as separate tasks. This leads to inefficiencies and yields only suboptimal results, as it fails to learn the inherent dependencies and correlations between tasks. The FOTs utilise an innovative operator, RoIRotate, which extracts features from the detected text regions used for both detection and recognition. The idea of sharing features between the text detection and recognition branches significantly reduces computational overhead. This helps to improve the efficiency gain of FOTs and thus to achieve real-time performance. The paper also presents experimental results on standard benchmarks, demonstrating FOTs' superiority over existing state-of-the-art methods.

The paper "Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes" by Lyu et al. (2018) proposes Mask TextSpotter, a novel approach to scene text spotting in natural images. It also discusses the limitations of traditional text spotting methods, which often treat text detection and recognition as

separate tasks. It also discusses the importance of having an end-to-end trainable model that can leverage the inherent complementarity between these two tasks. It also discusses the evolution of text detection methods, from handcrafted features to deep learning-based techniques. It also highlights the challenges associated with multi-oriented and arbitrary-shaped text detection. It then discusses the Mask R-CNN framework and its abilities for text detection and recognition. It highlights that this approach is well-suited for handling text instances, especially with arbitrary shapes, such as curved text. The paper also discusses character-based, word-based, and sequence-based approaches in scene text recognition. It highlights the limitations of sequence-based methods in handling irregular text and positions and explains why they chose character-segmentation-based recognisers in their proposed method. Finally, the paper discusses existing end-to-end text spotting methods and points out their limitations in terms of training complexity and their inability to handle arbitrary-shaped text. The paper then explains how Mask TextSpotter overcomes these limitations by providing a simpler training scheme and the ability to detect and recognise text of various shapes. Overall, this paper provides an overview of the evolution of text spotting, the limitations of existing text spotting techniques, and the need for an end-to-end trainable model capable of handling text with arbitrary shapes. It then proposes a solution for addressing this need.

The paper "Deep features for text spotting" by Jaderberg et al. (2014) proposes a novel approach to text spotting in natural images. It highlights the importance and

applications of text spotting and then discusses the typical two-step process, detection and recognition. It highlights that the accuracy of character classification has a high impact on the overall performance of text-spotting systems. It highlights the critical role of character classification in the text-spotting pipeline by noting that even a slight improvement in character classification accuracy can lead to a substantial improvement in word recognition. The paper proposes a novel approach that leverages Convolutional Neural Networks (CNNs) for both text detection and word recognition. This CNN architecture focuses on developing a high-quality character classifier. This CNN architecture efficiently shares features between different tasks, including text detection, case-sensitive and insensitive character classification, and bigram classification. The paper also proposes several technical improvements to traditional CNN architectures, including avoiding downsampling for per-pixel sliding window processing and utilising maxout activation functions for multi-mode learning. It also provides a method for automatically mining and annotating data from Flickr to generate word- and character-level annotations for training, thus addressing the need for extensive training datasets for CNNs. The approach achieves good performance on standard benchmarks, such as the ICDAR Robust Reading dataset and the Street View Text dataset. Overall, this paper highlights the potential of deep learning and proposes a novel feature-sharing approach for the challenges of text spotting in natural images.

The paper "Text spotting transformers" by Zhang et al. (2022) also proposes a novel approach to addressing the challenges of text spotting. It gives a comprehensive

overview of text detection, text recognition, regular text spotting, and arbitrarily shaped text spotting. Then, it highlights the challenges posed by arbitrarily shaped text and the various existing methods to address them, such as character-level detection, segmentation-based methods, and the use of Bezier curves, along with their limitations when dealing with curved or arbitrarily shaped text. It then proposes TESTR (TExt Spotting TRansformers), which leverages transformers to achieve end-to-end text spotting. The main advantage of TESTR is that it directly predicts polygon vertices or Bezier control points, thereby eliminating the need for complex post-processing steps. The paper also presents experimental results of TESTR for text spotting on benchmarks such as Total-Text and CTW1500, highlighting that TESTR achieves better accuracy and efficiency compared to existing methods. The paper also provides future research directions by highlighting the need for adaptive determination of polygon control points. Overall, this paper reviews existing techniques in text spotting and proposes a novel approach to address some challenges associated with them.

The paper "All you need is a boundary: Toward arbitrary-shaped text spotting" by Wang et al. (2020) proposes an end-to-end trainable network for spotting text of arbitrary shapes. The paper points out that existing methods for finding and understanding text in images often use simple rectangular boxes or try to segment each text instance. The problem with these methods is that they will not work well with curved or unusually shaped texts, and the probability of encountering these texts is high in real-life images. Additionally, these boxes may include extra backgrounds that make it difficult to read the

text correctly. The paper proposes an end-to-end trainable network for spotting text of arbitrary shapes using a set of boundary points. By using this set of boundary points, accurate extraction of irregular text features is possible. Additionally, it is possible to transform the boundary points into regular shapes for easy recognition, and this provides efficient refinement during training. The paper proposes a two-step process for detecting these boundary points, which includes rectangular box detection and boundary point regression with BPDN. Oriented rectangular box detection is the first step, which uses a CNN-based detector to identify the smallest possible oriented rectangular box that can enclose the text instance. The boundary point regression step uses a Boundary Point Detection Network (BPDN) to predict the precise boundary points within the detected area using an oriented rectangular box. The BPDN operates on features extracted from the box and learns to predict the offsets of the actual boundary points from a set of predefined default points. Thus, it allows the network to capture the shape of the text accurately, even if it is curved or irregular. The proposed method achieves excellent results on various benchmark datasets, including those containing curved and oriented text, such as TotalText and ICDAR2015. The paper also highlights the robustness and generalisation capabilities of the proposed approach.

The paper "Textdragon: An end-to-end framework for arbitrarily shaped text spotting" by Feng et al. (2019) gives a comprehensive literature review on scene text spotting. It discusses traditional methods of scene text detection in which the system localises individual characters and then groups them into words. It then discusses the

deep learning-based approaches that directly detect words. However, deep learning-based approaches have limitations in handling text with arbitrary shapes, such as curved or irregularly oriented text. It then discusses recent methods designed explicitly for curved text detection, which include approaches that utilise recurrent neural networks, adaptive text boundary representations, and representations based on a series of concentric disks. The paper also discusses the evolution of scene text recognition. It discusses the traditional methods that typically detect and recognise each character individually and then assemble them into words. Scene text recognition also uses deep learning methods that use CNNs for feature extraction and RNNs for sequential label generation. The paper highlights that current text detection methods are ineffective for detecting curved text. This is because they assume the text is always in a straight line. So, the paper suggests a new method called TextDragon. TextDragon improves existing techniques by using four-sided shapes (quadrangles) to represent curved text. This new method also helps connect the text detection and recognition parts of the process more smoothly. The proposed method detects and recognises text of any shape in a single, streamlined process. The paper also proposes a novel feature extraction operator, called the RoISlide operator, which connects arbitrary-shaped text detection and recognition by extracting and rectifying text regions from feature maps. The proposed method utilises a flexible shape representation, the RoISlide operator, and an efficient recognition module to achieve state-of-the-art results on benchmark datasets such as CTW1500 and Total-Text and comparable results on the ICDAR 2015 dataset.

2.4 Document Understanding

Document understanding helps in the comprehension of the structure and content of various documents. In the proposed AI-based visa processing, document understanding helps in grasping the context and hierarchy of information within various document types, including identity cards, passports, and other forms and documents. It ensures accurate extraction and interpretation of different document types. Researchers make significant contributions to this field with several innovative approaches and techniques.

The paper "Reading Text in the Wild with Convolutional Neural Networks" by Jaderberg et al. (2016) provides a comprehensive overview of text spotting, which involves both localising and recognising text within natural scene images. It discusses the challenges posed by the variability of text in real-world images, such as diverse fonts, styles, lighting conditions, occlusions, and background clutter, and it highlights that these challenges make text spotting distinct from traditional OCR tasks focused on document images. The paper proposes an end-to-end text-spotting pipeline that combines both text detection and recognition. This approach is beneficial for document understanding, as it enables the extraction of text and facilitates further analysis and interpretation from the raw image. It facilitates an approach to text recognition by employing a convolutional neural network (CNN) that takes the entire word image as input, eliminating the need for character-level recognition. The paper also introduces a synthetic data engine they built, which is capable of generating a wide variety of realistic text images with diverse fonts, styles, and backgrounds, similar to real-world scenarios. The paper also highlighted that

they effectively trained a CNN without the need for extensive manual labelling of authentic images using this synthetic data. It also presents the results of this approach on various benchmark datasets, including **I**CDAR 2003 (IC03), Street View Text (SVT), and ICDAR 2013 (IC13). Overall, this paper provides an overview of text-spotting techniques, proposes methods to enhance them, and highlights the novelty of their utilization of synthetic training data, the whole-word recognition methodology, and the system's scalability for extensive visual search applications.

The paper 'LayoutLM: Pre-training of Text and Layout for Document Image Understanding' by Xu et al. (2020) gives an overview of current techniques used for document analysis and recognition (DAR) and proposes a novel pre-training technique, LayoutLM for document image understanding. It discusses the evolution of DAR and the traditional method used for DAR, which is a rule-based approach. Traditional methods can be effective for some documents, but they cannot be generalised to diverse document types and need manual effort in crafting rules. The paper highlights the need for extensive human effort to develop more effective rules. It then discusses conventional machine learning approaches used in DAR, which employ statistical methods and algorithms such as support vector machines (SVMs) and Gaussian mixture models (GMMs). Machine learning approaches have offered improvements to DAR, such as the generalizability of documents; however, they are often computationally expensive and time-consuming due to the need for manual feature engineering. The paper then discusses deep learning approaches in DAR, highlighting the increasing demand for them. Deep learning models

can learn complex representations from the data on which they are trained, and they do not require any manual feature engineering. However, it faces limitations, such as its reliance on limited labelled data, whereas many real-world scenarios involve abundant unlabeled data. Also, in deep learning, text and layout information are not always thoroughly combined during training. This makes it more challenging for the model to comprehend how text and its arrangement interact in scanned documents. The paper also proposes a solution to this problem, LayoutLM, which can jointly learn representations of both text and layout information from scanned document images. LayoutLM is a transformer architecture (it builds upon the BERT architecture) for joint pre-training of text and layout representations on an extensive dataset encompassing both information types. The LayoutLM model is pre-trained on a large dataset of scanned documents using self-supervised objectives such as the Masked Visual-Language Model (MVLM) and Multi-label Document Classification (MDC). The MVLM objective randomly masks some input tokens and trains the model to predict them based on both textual and spatial context. The MDC task encourages the model to learn high-quality document-level representations by classifying documents into multiple categories. The pre-trained LayoutLM model can be fine-tuned on specific downstream tasks such as form understanding, receipt understanding, and document image classification. The incorporation of text, layout, and visual information enables LayoutLM to achieve state-of-the-art results on several benchmark datasets, including FUNSD, SROIE, and the RVL-CDIP dataset. The paper also provides future directions for research, such as using

larger models, incorporating image embeddings in pre-training, and utilising self-supervised objectives.

2.5 Hybrid Solution

In real-world situations, scene comprehension and document understanding often utilise hybrid systems to address these challenges. This literature review also examines papers that utilise hybrid technologies in these situations.

The paper "A Hybrid Solution for Extracting Information from Unstructured Data Using Optical Character Recognition (OCR) with Natural Language Processing (NLP)" by Dash (2021) proposes a hybrid solution that combines Optical Character Recognition (OCR) and Natural Language Processing (NLP) to extract meaningful information from unstructured data. It points out the importance of extracting valuable information from the ever-growing volume of unstructured data, especially at times (COVID-19 pandemic) when operational efficiency and timely access to information are critical. The paper also highlights that the majority of data produced today is unstructured and resides in various formats, such as PDFs, text files, audio, and video. It proposes a hybrid solution that combines Optical Character Recognition (OCR) and Natural Language Processing (NLP) to achieve this goal. OCR is used for converting Unstructured data, such as scanned documents or images, into machine-readable text. This text is then processed using NLP techniques for identifying entities, relationships, and other valuable information. The proposed solution utilises cloud-based big data platforms and Spark engines to process

large volumes of unstructured data. The paper also acknowledges the challenges associated with OCR accuracy, such as handling noisy or low-quality documents. It highlights the need for effective post-OCR processing to enhance the quality of the extracted text in these situations. The paper also discusses the evaluation methodologies used to assess the accuracy and effectiveness of the proposed approach. The paper then provides future research directions for collecting information from unstructured data, such as incorporating domain-specific vocabularies into the NLP model to increase its accuracy and applicability further.

The paper "Optical Character Recognition Errors and Their Effects on Natural Language Processing" by Lopresti (2008) primarily discusses the impact of Optical Character Recognition (OCR) errors on Natural Language Processing (NLP) tasks. It also discusses the broader landscape of research on handling errors and noise in text and speech processing. The paper starts by explaining the inevitability of errors in OCR, especially when processing degraded or complex documents. It then discusses different stages in a text analysis pipeline, including sentence boundary detection, tokenisation, and part-of-speech tagging. It explains how OCR errors can propagate and amplify through each of these stages. The paper then highlights the lack of research focused on the impact of OCR errors on NLP tasks. It highlights this by discussing a few earlier studies that examined the effects of OCR errors on information retrieval and machine translation, and pointing out that these studies often made simplifying assumptions or had limited scope. The paper references the authors' previous work, in which they proposed a

paradigm for evaluating text analysis pipelines in the presence of noisy input. This paradigm utilises hierarchical approximate string matching, which forms the basis for the evaluation methodology presented in this paper. The paper proposes a novel method for evaluating the performance of multi-stage text analysis pipelines when processing noisy OCR output. This method uses a hierarchical approximate string matching to align and compare the results of NLP tasks on both clean and OCR-processed text, and hence enables both the identification and classification of errors induced by OCR. The paper also gives the results of experimental evaluation using a large dataset of scanned documents. The results underscore that errors introduced during OCR do not remain isolated; instead, they propagate throughout the system. Instead, they cascaded it in subsequent stages in the NLP pipeline, such as sentence boundary detection, tokenisation, and part-of-speech tagging. Overall, this paper provides an in-depth analysis of how specific OCR errors, such as misinterpreting punctuation or inserting or deleting spaces, can lead to various types of errors in later NLP stages.

The study titled "OCR Post Correction for Endangered Language Texts" by Rijhwani et al. (2020) studies the challenges in Natural Language Processing (NLP) for endangered languages. It highlights that the primary challenge in NLP for endangered languages is the scarcity of available data. It also points out that while valuable textual resources do exist for many endangered languages, they are often stored in non-machine-readable formats, such as paper books, scanned images, and handwritten notes. It also highlights that this makes it challenging to train and evaluate NLP models.

All these create barriers to accessing and utilising data for NLP research and language preservation. The paper focuses on two areas: OCR post-correction and processing endangered languages. It also points out that very little research exists where these two areas overlap. This paper builds upon prior work in OCR post-correction for high-resource languages and leverages the capabilities of multi-source encoder-decoder models to tackle the challenges of extracting text from scanned documents in endangered languages.

2.6 Sentimental Analysis

The proposed AI visa processing system is built on insights from online survey analysis and sentiment analysis, which help automate the emotion evaluation process. Therefore, it is also important to conduct a literature review on the topic.

The paper "Sentiment Analysis: A Comparative Study on Different Approaches" by Devika et al. (2016) gives a comprehensive study on sentiment analysis, its applications, and ongoing research in this field. It describes sentiment analysis as a process of extricating the feelings and emotions from the text. It discusses different computational techniques and approaches used to analyse sentiments and opinions expressed in text. It also discusses the increasing demand for sentiment analysis in various domains, including business, politics, and social media, especially with the arrival of social media platforms. It explains various methods, including machine learning, lexicon-based, and rule-based approaches. It discusses how a machine learning

model is built and applies it to new, unseen data. It also explains various machine learning algorithms, including support vector machines (SVM), Naive Bayes, maximum entropy, K-Nearest Neighbour (K-NN), and Weighted K-NN, as well as machine learning techniques such as N-gram sentiment analysis, multilingual sentiment analysis, and feature-driven sentiment analysis. The rule-based approach defines specific rules to identify sentiment in text. It also highlights that rule-based systems can work on manual rule creation and automated learning. The rule-based approach is relatively simple, but it can be limited by the quality of the rules it employs. The lexicon-based approach utilises a pre-built dictionary (lexicon), where these words are associated with a sentiment polarity (positive or negative) and, in some cases, intensity. The sentiment score is then determined by summing up the polarities of its constituent words or phrases. This is an unsupervised learning technique, and the availability of linguistic resources constrains its capabilities. The paper highlights that this method can be further enhanced by incorporating contextual information and taking into account word modifiers and negations. It also describes various levels of sentiment analysis, including document-level, sentence-level, and aspect-level analysis. Overall, the paper gives an overview of sentiment analysis techniques and their role in extracting meaningful information from the growing social media platforms and e-commerce websites.

The paper "Sentiment analysis" by Shah et al. (2010) discusses the application of machine learning techniques for sentiment analysis, specifically on Twitter data. The paper explores the concept of sentiment analysis, its significance in understanding public

opinion, and the role of social media platforms, such as Twitter, in generating sentiment-rich data. It also discusses the significant challenges associated with Twitter sentiment analysis due to the presence of misspellings and slang terms. It explains various aspects of sentiment analysis, such as different machine learning algorithms, feature extraction methods, and the impact of domain-specific information on sentiment classification. It also describes the various steps involved in data processing, including cleaning, categorisation, text preprocessing, and word embeddings. It also explains the architecture of the text classifier the authors used for sentiment analysis, which uses LSTM layers and dropout regularisation. It also discusses the use cases of sentiment analysis in various domains, such as market research, customer support, brand management, political analysis, and healthcare. It then explains the challenges associated with sentiment analysis, including inaccuracy in tone detection, handling slang and informal language, and limitations in processing multilingual content. The paper concludes by highlighting the significance of sentiment analysis in understanding collective emotions and opinions on social media. It also highlights the need for further research to ensure the ethical and unbiased use of sentiment analysis algorithms.

The paper "Comparing and Combining Sentiment Analysis Methods" by Gonçalves et al. (2013) gives a comprehensive study of the state of sentiment analysis, specifically in the context of online social networks (OSNs) like Twitter. It highlights the growing importance of sentiment analysis due to the massive volume of content generated on these platforms, and this content often expresses opinions and emotions

about various topics. The paper explains two primary methods used in sentiment analysis: machine learning-based methods and lexicon-based methods. Machine learning methods often employ supervised classification techniques for sentiment analysis, in which a model is trained on labelled data to predict sentiment (polarity). These supervised classification techniques rely on labelled data, which can limit their applicability. Lexical-based methods utilise predefined dictionaries of words associated with various sentiments. This technique does not require labelled data; however, creating a universally applicable dictionary is challenging due to the dynamic nature of language and context-specific word usage. It provides a comparative study of eight popular sentiment analysis methods, including LIWC, Happiness Index, SentiWordNet, SASA, PANAS-t, Emoticons, SenticNet, and SentiStrength. It also evaluates these methods using two large-scale datasets: a nearly complete log of Twitter messages and a collection of human-labelled web texts. The nearly complete log of Twitter messages dataset comprises nearly 1.7 billion tweets posted by 54 million users between March 2006 and August 2009. The human-labelled web text dataset comprises six sets of messages from various web platforms (MySpace, Twitter, Digg, BBC forum, Runner's World forum, and YouTube comments) that have been manually labelled as either positive or negative. The paper mainly uses two metrics for comparing the selected eight sentiment analysis methods: coverage and agreement. Coverage is defined as the proportion of messages within a dataset that a method can successfully classify as either positive or negative sentiment. A higher value for coverage percentage indicates that the method is capable of handling a wide range of textual expressions. An agreement quantifies the consistency

between different sentiment analysis methods. It measures the degree to which multiple methods agree with each other on the polarity assigned to a given text. The study reveals that existing sentiment analysis methods exhibit varying degrees of coverage and agreement, and no single method is consistently outperforming others across different scenarios. It also highlights the trade-off between coverage and agreement. The paper proposes a novel method that integrates multiple existing sentiment analysis approaches to achieve both high coverage and high agreement. Overall, this paper gives a comparative analysis of various sentiment analysis methods and proposes a combination of multiple methods to enhance the effectiveness of sentiment analysis systems.

The paper "Sentiment Analysis and Subjectivity" by Liu (2010) provides a comprehensive study of sentiment analysis and the challenges in identifying and understanding opinions, sentiments, and emotions embedded within textual data. It explains the difference between facts and opinions; facts are unique and have no subjective nature. Unlike facts, opinions are highly subjective. They should be processed accurately before incorporating them into a decision-making process. It then explains in detail the core concepts and tasks of sentiment analysis, including sentiment classification, feature-based sentiment analysis, analysis of comparative sentences, opinion search and retrieval, opinion spam detection, and the utility of opinions. The paper also provides a detailed explanation of various techniques and algorithms employed in both supervised and unsupervised learning approaches for these tasks. It then highlights the role of opinion lexicons in sentiment analysis. Opinion lexicons are

collections of words with positive or negative orientations, and these orientations indicate the sentiment or emotional polarity associated with each word. It also discusses how to create lexicons, such as by manual creation, using dictionaries, or analysing large text corpora. Overall, the paper gives a comprehensive overview of sentiment analysis techniques, the challenges associated with it, and future directions for advancements.

The paper "Challenges in sentiment analysis" by Mohammad (2017) gives a comprehensive overview of the complexities and ongoing research directions in sentiment analysis at that time. It discusses the multifaceted nature of sentiment and highlights the subjective nature of sentiment associated with the writer, the reader, or entities mentioned in the text. The paper discusses the importance of considering sentiment at various levels, even from the word or phrase level to the document level. It then discusses the challenges in sentiment composition, such as understanding the impact of negation, degree adverbs, intensifiers, and modals on the overall sentiment of a phrase or sentence. It also discusses the tedious process of annotating data for sentiment analysis and the difficulties of labelling sarcasm, rhetorical questions, or the speaker's emotional state.

The paper "Sentiment analysis: A review and comparative analysis of web services" by Serrano-Guerrero et al. (2015) provides an overview of sentiment analysis and the challenges associated with it. It describes various concepts in sentiment analysis, such as emotions, opinions, and subjectivity. It also differentiates related tasks such as

sentiment classification, subjectivity classification, and opinion summarisation and highlights the diverse applications of sentiment analysis techniques. It then categorises sentiment analysis techniques into machine learning approaches (supervised, unsupervised, and hybrid) and lexicon-based approaches (dictionary-based and corpus-based), discussing them in detail. It also discusses the role of Natural Language Processing and Information Retrieval in sentiment analysis. It also discusses the need for feature selection and extraction in machine-learning approaches and the use of sentiment lexicons in lexicon-based approaches. It also highlights the challenges posed by NLP in accurately detecting sentiments, especially in the context of irony and sarcasm. It also gives future directions for advancements in sentiment analysis, particularly in areas such as explicit subjectivity detection and more nuanced sentiment classification.

The paper "A survey on sentiment analysis methods and approach" by Abirami et al. (2017) gives a survey of sentiment analysis methodologies and their applications. It explains the growing importance of sentiment analysis in various domains, particularly in business decision-making and market research, where understanding customer opinions and feedback is needed. It highlights the significance of data analytics in extracting valuable insights from both structured and unstructured data and utilising them to enable organisations to make informed decisions. It then explains the core concept of sentiment analysis, which involves identifying and categorising opinions expressed in text. It explains different approaches to sentiment analysis, including machine learning techniques, dictionary-based methods, and ontology-based approaches, along with their

advantages and disadvantages. It then explains the challenges associated with sentiment analysis, including polarity shifts, limitations of binary classification, and data sparsity issues, as well as their impact on the accuracy and effectiveness of sentiment analysis systems. It also provides an analysis of various sentiment analysis methodologies, by considering factors such as the issues addressed, proposed techniques, accuracy, and limitations of sentiment analysis. Overall, the paper gives an overview of the fundamentals of sentiment analysis, various methodologies, and approaches used in this field.

The paper "Sentiment analysis techniques in recent works" by Madhoushi et al. (2015) gives an overview of various techniques and approaches employed in sentiment analysis. It identifies that sentiment analysis involves classifying opinions as positive, negative, or neutral and differentiating subjective opinions from objective facts. It discusses various techniques in sentiment analysis, including machine learning, lexicon-based approaches (which leverage dictionaries and corpora) and hybrid approaches. The survey also discusses various problems in sentiment analysis, including classification accuracy, language limitations, scarcity of labelled data, and limitations of lexicons. The paper concludes that successful sentiment analysis techniques are likely to involve a combination of hybrid approaches and natural language processing techniques.

The paper "Using appraisal groups for sentiment analysis" by Whitelaw et al. (2005) presents prior research in sentiment analysis, particularly on two main approaches:

a bag-of-words and semantic orientation. The bag-of-words approach tries to learn a classifier based on the frequency of words in a document. The semantic orientation approach sorts words into "good" or "bad" categories. Then, it adds up these scores to get a general feeling for the text. The paper argues that this approach overlooks crucial aspects of sentiment analysis that genuinely work. It also emphasises the need for a much better understanding of how people express their feelings through language. It also highlights that these attitudes are conveyed through groups of words, not just single words. It proposes a novel approach to address these gaps by focusing on the extraction and analysis of adjectival appraisal groups. It utilises taxonomies from Appraisal Theory to categorise these expressions and develop a lexicon using semi-automated techniques. It then extracts appraisal groups from text, computes their attribute values, and employs machine learning to classify documents based on sentiment. The paper also shows promising results of the proposed approach in the context of movie review classification, demonstrating that even with a limited lexicon, appraisal group features can enhance sentiment classification accuracy. It points out the importance of developing detailed semantic tools for sentiment analysis and suggests that future research should focus on accurately identifying full appraisal expressions.

The paper "A survey of sentiment analysis in social media" by Yue et al. (2019) gives a survey of sentiment analysis in social media. It highlights the growing importance of sentiment analysis applications, driven by the increasing prevalence of social media platforms and the vast amount of opinions expressed on them. It also highlights the

unique challenges posed by social media data, including informal language, short text lengths, and the presence of multimodal content. The paper categorises sentiment analysis research from multiple perspectives, including task-oriented (specific tasks such as polarity classification and emotion detection), granularity-oriented (sentiment analysis at document, sentence, and word levels), and methodology-oriented (supervised, unsupervised, and semi-supervised learning approaches). It also provides an overview of available datasets and tools for sentiment analysis research. It then discusses benchmark datasets from platforms such as Facebook, Twitter, and Digg, as well as tools and lexicons that can be utilised for sentiment analysis tasks. The paper concludes by highlighting future trends and challenges in sentiment analysis, with a particular emphasis on the need for further research in multimodal sentiment analysis.

The paper "A review on sentiment analysis methodologies, practices and applications" by Mehta and Pandya (2020) gives a comprehensive review of sentiment analysis, its methodologies, applications, and challenges. It discusses the significance of sentiment analysis in understanding human behaviour and decision-making due to the growing online communication channels. It then explains various concepts of sentiment analysis, including sentiment and various levels of analysis, such as document-level, sentence-level, and aspect-level sentiment classification. It also discusses the diverse applications of sentiment analysis, including decision support systems in business analytics and trend prediction. It then explains the methodologies employed in sentiment analysis by categorising them into machine learning and lexicon-based approaches. The

machine learning approach involves training models on labelled datasets to classify sentiments using algorithms such as Naive Bayes, Support Vector Machines, and Neural Networks. The lexicon-based approach relies on dictionaries and corpora to identify and score sentiment-bearing words and phrases. The paper also discusses the challenges associated with sentiment analysis, including handling negation, sarcasm, and domain-specific language. Overall, the paper provides an overview of sentiment analysis, why it is so important, different approaches to sentiment analysis, and the difficulties associated with this area of study.

The paper "Sentiment analysis and opinion mining: a survey" by Vinodhini and Chandrasekaran (2012) also gives a survey of sentiment analysis and opinion mining, primarily on the period from 2000 onwards. It provides a systematic overview of sentiment analysis, starting with the fundamental definitions of sentiment analysis and opinion mining. It then discusses the diverse data sources utilised in this field, including blogs, review sites, datasets, and microblogs. It also discusses the techniques employed for sentiment classification, including machine learning approaches (such as Naive Bayes and SVM) and semantic orientation methods. It also discusses the critical role of negation handling in sentiment analysis, as well as the specific challenges posed by feature-based sentiment classification. It also discusses the practical applications of sentiment analysis, such as online advertising and competitive intelligence, and provides an overview of available tools for sentiment analysis. It also provides an evaluation of the performance of various sentiment analysis methods and provides future research directions.

The paper "Sentiment analysis: A multi-faceted problem" by Liu (2010) offers a comprehensive overview of sentiment analysis. It explicitly highlights that sentiment analysis is complex and has numerous built-in challenges. The paper also highlights that sentiment analysis is crucial due to the vast amount of opinion-filled text available on the internet. It also highlighted the need for automated systems to analyse and summarise this data effectively. It explains important concepts in sentiment analysis, such as objects, features, opinion holders, opinions, and orientation. It also explains the objectives of sentiment analysis, such as discovering opinion quintuples and identifying feature synonyms. It also discusses the technical challenges associated with sentiment analysis, such as object identification, feature extraction, synonym grouping, opinion orientation classification, and the integration of these tasks. It explains these challenges using examples and discusses the complexities involved in addressing them. It then discusses past achievements and limitations. The paper then highlights the need for more refined investigations and integrated systems by pointing out these limitations.

The paper "Lexicon-based methods for sentiment analysis" by Taboada (2011) gives a comprehensive overview of the lexicon-based approach to sentiment analysis. It discusses problems with machine learning methods. Machine learning (ML) methods require training data for each topic and struggle with handling linguistic nuances. ML methods often fail to understand hidden language details, such as when words are strengthened or negated. Due to these issues, the paper advocates for the use of

lexicon-based methods by citing that they are more reliable and adaptable. It then proposes a novel method called Semantic Orientation CALculator (SO-CAL) to address the challenges of machine learning approaches, such as domain dependence and linguistic nuances. It explains the development and features of SO-CAL, which utilises dictionaries of words annotated with semantic orientation (polarity and strength). SO-CAL also includes linguistic nuances, such as negation, intensification, and irrealis markers, which enhance the accuracy of sentiment analysis. The paper then explains the importance of using reliable dictionaries. It discusses the manual creation process, as well as the use of Mechanical Turk to validate the dictionaries against human judgments, ensuring their robustness and consistency. The paper also gives a performance assessment of SO-CAL across various domains, including product reviews, news articles, blogs, and social media comments. The results of these performance assessments demonstrate consistent accuracy, the ability to generalise to unseen data, and SO-CAL's domain independence. The paper also provides an analysis of SO-CAL's dictionaries in comparison to other available lexicons and highlights its superior performance in capturing nuanced sentiment. The paper also gives future directions for research to enhance SO-CAL by incorporating contextual information and discourse parsing techniques.

The paper "Sentiment analysis is a big suitcase" by Cambria et al. (2017) provides an overview of the challenges associated with sentiment analysis. It highlights that sentiment analysis is not just a binary classification task but a multifaceted problem requiring the resolution of various NLP sub-tasks. It acknowledges the significant

progress made in the field of Natural Language Processing (NLP) due to recent advancements in deep learning. However, it argues that these advancements are insufficient to attain human-like proficiency in sentiment analysis. The paper explains that achieving human-like proficiency in sentiment analysis requires addressing a diverse range of NLP tasks. It highlights the need to bridge the gap between statistical NLP and other disciplines, such as linguistics, reasoning, and affective computing. It proposes a three-layer structure: the syntactic layer, the semantic layer, and the pragmatic layer to achieve human-like performance in sentiment analysis. The syntactic layer is for pre-processing the text; the semantic layer deconstructs the text into concepts and resolves references; and the pragmatic layer extracts meaning from the sentence structure and semantics. The paper emphasises the importance of integrating symbolic and subsymbolic AI. It highlights the need for a holistic approach that combines data-driven algorithms with theory and computational natural language. Overall, this paper gives a broader perspective on sentiment analysis and promotes a combined approach that emulates human language understanding.

The paper "Sentiment analysis on social media" by Neri et al. (2012) provides a comprehensive overview of the state of sentiment analysis in the 2010s and the growing significance of opinion mining and sentiment analysis in gauging public attitudes toward brands, services, and even political entities. It then highlights the influence of online reviews and ratings on consumer decisions and emphasises the need for automated sentiment analysis tools to manage the vast amount of online data. The paper explains

the two main techniques used in sentiment analysis: supervised machine learning and unsupervised lexicon-based methods. It discusses the strengths and weaknesses of each approach. It explains the trade-off between accuracy and the need for training data in AI. It then discusses the complexities involved in accurately extracting sentiment from text, such as word sense disambiguation, negation handling, and identifying the opinion holder and topic. The paper explains the challenges associated with performing sentiment analysis across multiple languages, thereby highlighting the need for language-specific resources and tools. The paper also discusses a sentiment analysis conducted by the authors on Facebook posts to compare the public perception of Italian newscasters Rai (a state-owned entity) and La7 (a private entity). The analysis revealed a positive sentiment towards La7 and a negative sentiment towards Rai, which aligns with observations from media institutions and viewership data. Overall, this paper gives a comprehensive overview of concepts and techniques directly applicable to the sentiment analysis study on social media data discussed in it.

CHAPTER III:

METHODOLOGY

3.1 Overview of the Research Problem

Visa processing is the first step for many international travellers. People plan international trips for several reasons, and often, it is an essential and quick decision. The current visa process system faces several challenges, including manual documentation handling, labour-intensive processes, physical paperwork, and lengthy processing times. It often requires applicants to print and submit numerous documents, attend in-person meetings, and then wait a considerable amount of time to receive a response. In several cases, applicants are required to submit multiple printed documents, attend in-person appointments, and wait for extended periods for updates. These challenges are particularly burdensome for older people and those who require additional support. These challenges often result in undesirable outcomes, such as user dissatisfaction. This research proposes an AI-based visa processing system that aims to address these challenges by enhancing operational efficiency, minimising human errors in application review, and substantially reducing paper consumption with the help of state-of-the-art AI technologies, such as OCR, NLP, text spotting, and LLMs. This chapter focuses on identifying the limitations of the current system, gauging the public perception of incorporating AI in visa processing through a survey and implementing an AI-based visa processing system.

3.2 Operationalization of Theoretical Constructs

This study operationalises several theoretical constructs to evaluate the potential and acceptance of an AI-based visa processing system. Each construct has been translated into measurable variables using data collected through the structured online survey and performance metrics of the proposed system prototype.

- Efficiency: Efficiency is defined as the ability of the visa processing system to minimise delays, reduce manual errors, and accelerate application handling. It was operationalised through variables such as reported user perceptions of processing speed, observed error frequency, and automated time logs generated by the prototype system.
- Accessibility: Accessibility is conceptualised as the ease with which applicants can engage with the visa system in terms of physical reach, procedural clarity, and convenience. Survey questions were designed to capture user perceptions of accessibility, while the prototype evaluation examined whether automation reduces the need for repeated visits and excessive paperwork.
- Automation Acceptance: Automation acceptance refers to the willingness of applicants and other stakeholders to adopt AI-driven processes in place of traditional manual tasks. This construct was measured through survey items assessing openness to automation, perceived benefits, and comfort levels with reducing human involvement.
- System Transparency: System transparency is defined as the degree to which users feel they can understand and monitor the progress and decisions of the visa

process. It was operationalised through indicators such as the reported ability to track application status, perceptions of fairness, and clarity in system communication.

All these constructs together provide a comprehensive framework for evaluating both user experience and functional outcomes of the AI-based visa processing model.

3.3 Research Purpose and Questions

This research is to design, develop, and evaluate an AI-based visa processing system that enhances administrative efficiency, minimizes paper usage, reduces processing time, and eliminates human errors. For that, the study also aims to examine how people perceive and accept the integration of artificial intelligence into traditionally manual workflows in visa processing. Both the technical and operational effectiveness of such systems, as well as the emotional and psychological responses of visa applicants and the general public. The study examines the extent to which AI technologies can foster trust, promote transparency, and enhance user satisfaction while being perceived as fair, inclusive, and secure. To address the research aim, the following research questions were formulated:

1. Can an AI-based visa processing system reduce paper usage, processing time, and human errors compared to the traditional system?
2. What are the perceptions, preferences, and concerns of visa applicants and the general public regarding the adoption of AI in visa processing?

3. How can AI technologies be efficiently and effectively integrated to automate and streamline visa processing tasks?

This research contributes to a balanced understanding of AI implementation in public services by addressing both the technical architecture and the human dimensions of AI visa processing.

3.4 Research Design

This research adopts a mixed-methods approach, combining qualitative and quantitative methods to figure out the public perception of the AI-based visa processing system and to design, implement and evaluate it. This approach enabled the collection of both contextual insights and empirical validation. The research is conducted in two phases:

1. **Phase I – Qualitative Exploration:** This phase is designed and implemented to understand and analyse user experiences, preferences, and concerns related to AI-based current and AI visa processing. An online survey containing both closed- and open-ended questions was administered to a diverse population of visa applicants, professionals, and the general public. The qualitative data collected is analysed systematically and through sentiment analysis using Natural Language Processing (NLP) tools.
2. **Phase II – System Development and Quantitative Evaluation:** A prototype AI-based visa processing system is developed. It used OCR for data extraction,

NLP for understanding application content, and machine learning models for form classification and document validation. The system is evaluated using a set of synthetic and real-world visa application data. Quantitative metrics, including processing time, accuracy, paper usage, and error rate, are collected and compared against traditional methods using statistical techniques.

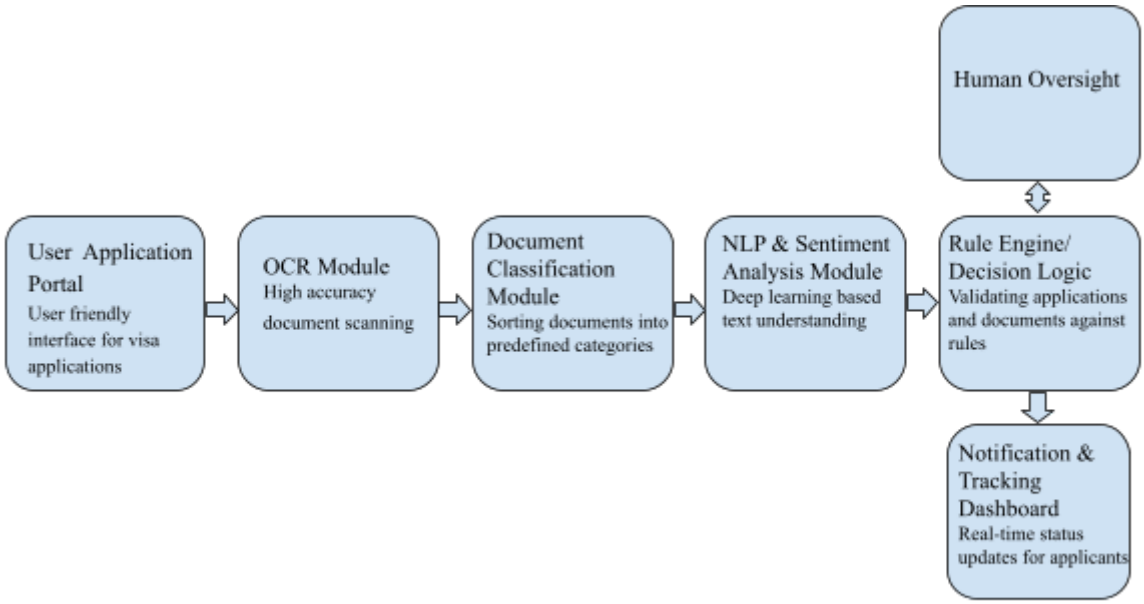

Figure 3.1: Block diagram of AI-based visa processing prototype system

The block diagram above represents the AI-based visa processing prototype system developed for this study. The first stage of the system is a User Application Portal, an interface that enables applicants to submit their forms and upload necessary documents in various formats. The next stage is the OCR (Optical Character Recognition) Module, which will extract text from scanned

images and PDFs uploaded. The extracted textual data is then fed into the Document Classification Module, which categorises documents into various types, such as passports, financial documents, and employment letters, using machine learning classifiers. This categorisation ensures that each document type is handled according to its layout, importance and relevance in the decision-making process. After the document classification module, the extracted content is then processed by the NLP & Sentiment Analysis Module. This module utilises deep learning techniques to comprehend the textual data. The prototype system then uses a Rule Engine/Decision Logic module to apply a set of predefined rules and logic, validating the applications. It cross-verifies information across documents, and flags anomalies or rule violations. In cases where the prototype system experiences any ambiguity or low confidence in its predictions, a human oversight stage is incorporated in the prototype to ensure that critical decisions are reviewed by a human expert, thereby enhancing the system's reliability and accuracy. A Notification & Tracking Dashboard module is also incorporated in order to provide real-time status updates to applicants, ensuring transparency and allowing users to monitor their application progress at every stage. Altogether, this integrated AI-driven protype system significantly reduces manual workload, minimises errors, and expedites visa processing while maintaining human accountability at key decision points.

3.5 Population and Sample

The population for this study comprises three key stakeholder groups associated with the visa application and processing ecosystem:

1. **Visa Applicants:** Individuals who have recently applied for a visa or plan to apply shortly.
2. **Visa Processing Officials:** Personnel working at consulates, embassies, travel agencies or visa service centres who are directly involved in the receipt, verification, and approval of visa applications.
3. **General Public:** Citizens or residents with indirect experience or opinions about visa processes, digital governance, or the use of artificial intelligence in administrative services.

Purposive sampling was employed to get an in-depth understanding of relevant individuals. This non-probability sampling technique was to deliberately include participants with varied backgrounds, ensuring diversity in terms of age, profession, travel experience, and digital literacy.

Sample Size

- A total of 267 valid responses were collected through a structured Google Form survey distributed
- The participants included both frequent and infrequent travelers

- A stratified approach was used to ensure representation across age brackets, gender identities, and occupational categories

This diverse sample enabled a balanced exploration of both technological feasibility and societal readiness for AI-driven visa processing.

3.6 Participant Selection

To ensure the inclusion of diverse and relevant viewpoints, purposive sampling was done for this study. This non-probability sampling method was chosen to deliberately target individuals with direct or indirect exposure to visa application processes or with perspectives on digital transformation in public administration. The participant selection process followed a stratified purposive strategy aimed at achieving representation across the following dimensions:

- Age groups: Participants were grouped into brackets, including those under 25, 25–40, 41–60, and above 60.
- Occupational categories: Including students, business owners, professionals, government workers, and retirees.
- Travel history: Both frequent international travellers and those with limited or no travel experience were included.
- Digital literacy levels: Ranging from novice users to highly digital-savvy individuals.

- Gender identities and geographic diversity: Ensured through wide dissemination channels.

The survey was administered using a structured Google Form and distributed through a multi-channel outreach strategy, including:

- Professional networks such as LinkedIn
- Email invitations to academic and business contacts
- Public platforms and forums like Reddit
- Community-based communication tools, especially WhatsApp groups

Respondents were invited to participate voluntarily and were informed about the confidentiality and ethical handling of their data. Digital informed consent was obtained at the start of the survey. The participants for the study were chosen very carefully. This ensured the sample truly represented all the different people involved in visa processes. This careful selection allowed for a complete assessment of how ready visa processing systems are for AI. It looked at the technological, procedural, and emotional aspects of this readiness.

3.7 Instrumentation

This study employs a combination of survey-based and system-level instruments for data collection from participants, as well as for the implementation of the visa

processing system based on the survey results. The instrumentation strategy is designed to align with the research objectives.

1. Survey Instrument (Google Form)

The primary tool for collecting data from participants was a Google Form survey, which included a mix of open-ended and closed-ended questions to understand perspectives, experiences, and concerns regarding the current and AI-based visa processing system. The survey was organized into the following thematic sections:

- Demographic Information: Age, gender, occupation, nationality, and international travel frequency
- Experience with Traditional Visa Processes: Questions on accessibility, convenience, paperwork challenges, and process transparency
- Perception of AI Integration: Willingness to adopt AI-based systems, perceived benefits or risks, and emotional responses
- Suggestions for Improvement: Open-ended prompts inviting respondents to propose enhancements to the existing visa system

Closed-ended questions used Likert-type scales (e.g., Very Satisfied to Very Dissatisfied) to quantify opinions and support statistical comparisons. Open-ended responses allowed for thematic and sentiment analysis using NLP tools and sentiment analysis in AI.

2. AI System Evaluation Metrics

To evaluate the prototype AI-based visa processing system developed for this study, instrumentation was embedded within the system to capture key performance indicators (KPIs), including:

- Processing Time: Automated logging of time required to process a visa application (measured in seconds or minutes)
- Paper Usage Reduction: Calculated based on the elimination of physical form handling (measured in estimated sheets/pages saved)
- Accuracy: Measured through the precision of OCR and NLP components in extracting, interpreting, and validating document data
- Error Rate: Comparison of how many mistakes are made by human-versus-AI in document handling
- Sentiment Scores: Applied to survey feedback using tools like TextBlob and VADER to assess the responses

3. Validation and Pilot Testing

Before full deployment, the survey instrument was pilot-tested with a small group (n = 10) to ensure clarity, usability, and face validity. Minor revisions were made based on feedback to improve the structure and remove ambiguities. The combined use of survey instruments and system performance trackers allowed for a multidimensional analysis in this study.

3.8 Data Collection Procedures

Data for this study was collected in two distinct phases. Care is given in the data collection process to maintain ethical standards, response diversity, and support analysis.

Phase 1: Survey-Based Data Collection

In this phase, data were collected using an online structured survey. It was created using Google Forms and made available for responses for five weeks. The Google Forms was distributed via multiple digital channels to reach a broad and diverse audience, including:

- Professional platforms (e.g., LinkedIn)
- Email invitations sent to academic and corporate mailing lists
- Community networks such as WhatsApp groups and Reddit forums focused on travel, visas, and technology

Participation in the survey was entirely voluntary, and respondents were given the option to remain anonymous. Data confidentiality was maintained throughout the study. Responses were automatically captured in the Google Forms system and then exported to secure, password-protected databases for cleaning and analysis. A total of 267 valid responses were collected across diverse background participants.

Phase 2: System Testing and Log-Based Data Collection

The AI-based visa processing prototype, developed was evaluated using selected performance metrics in this phase. The system was tested using a dataset of synthetic visa applications designed to mimic real-world forms, passports, and supporting documents. Performance data was collected automatically through internal system loggers and performance-tracking modules embedded in the prototype. These instruments recorded:

- Time required to process each application
- Accuracy of document parsing using OCR and NLP
- Frequency of errors or data mismatches
- Estimated reduction in paper usage due to digitization

All test data was anonymized and non-sensitive, and no real personal data was used in the AI testing. The system environment was isolated from production systems to maintain data security and experimental control.

This two-step data collection helped gather information on how well the AI-based system works and people's thoughts and feelings about it. Together, this information gave a strong base for judging if AI can be used effectively in visa processing.

3.9 Data Analysis

The data analysis process was carefully structured to align with the research objectives and to ensure methodological rigour. Both qualitative and quantitative data were analysed using appropriate techniques to extract meaningful patterns and insights.

For quantitative data, statistical analysis was conducted using Python. Descriptive statistics such as means, standard deviations, and frequency distributions were used to summarise participant demographics and general trends. Data triangulation was employed to enhance the credibility and validity of the findings by comparing results from multiple sources and methods. The analysis adhered strictly to ethical standards, ensuring participant confidentiality and unbiased interpretation.

3.10 Research Design Limitations

While the research design was carefully developed to address the research objectives and ensure methodological rigor, several limitations should be acknowledged.

First, sampling limitations may affect the generalizability of the findings. The study relied on a specific population sample (as detailed in Section 3.5), which may not fully represent the broader target population. As such, the conclusions drawn should be interpreted with caution when applied to different demographic or geographic contexts.

Second, instrumentation constraints may have influenced data accuracy. Although validated instruments were employed where possible, self-reported data are inherently subject to response biases such as social desirability and recall inaccuracies. Additionally,

in qualitative data collection, the interpretation of open-ended responses may have been influenced by the researcher's subjective lens despite efforts to ensure objectivity.

Third, temporal and contextual limitations exist. The data were collected within a specific time frame and under particular environmental or institutional conditions, which may have impacted participants' responses or behaviors.

Finally, technological or analytical limitations might have constrained the depth of data interpretation. While robust software tools were used for both statistical and thematic analysis, the inherent limitations of these tools and the human element in coding and interpreting qualitative data may have introduced minor errors or omissions. Recognizing these limitations is essential for transparency and provides direction for future research aiming to address or mitigate these constraints.

3.11 Conclusion

This chapter outlined the methodological framework employed to investigate the research problem, providing a comprehensive explanation of the research design, sampling strategy, data collection methods, and analytical techniques. The study was structured to ensure alignment between the theoretical constructs and the

operationalization process, with careful attention paid to the validity and reliability of instruments used.

Through the articulation of the research purpose, questions, and procedures, this chapter established a foundation for the empirical investigation that follows. While acknowledging inherent limitations, the methodology was designed to provide credible, ethical, and contextually grounded insights into the phenomenon under study.

The next chapter will present and analyze the findings derived from the data collection, applying the methods detailed herein to address the core research questions and objectives.

CHAPTER IV:

RESULTS

4.1 Research Question One

Can an AI-based visa processing system significantly reduce paper usage, processing time, and human errors compared to the traditional system?

The findings of this study strongly support the idea that an AI-based visa processing system can address many of the inefficiencies found in the traditional, manual, and paper-based approach, particularly in terms of reducing paperwork, saving time, and minimising human error.

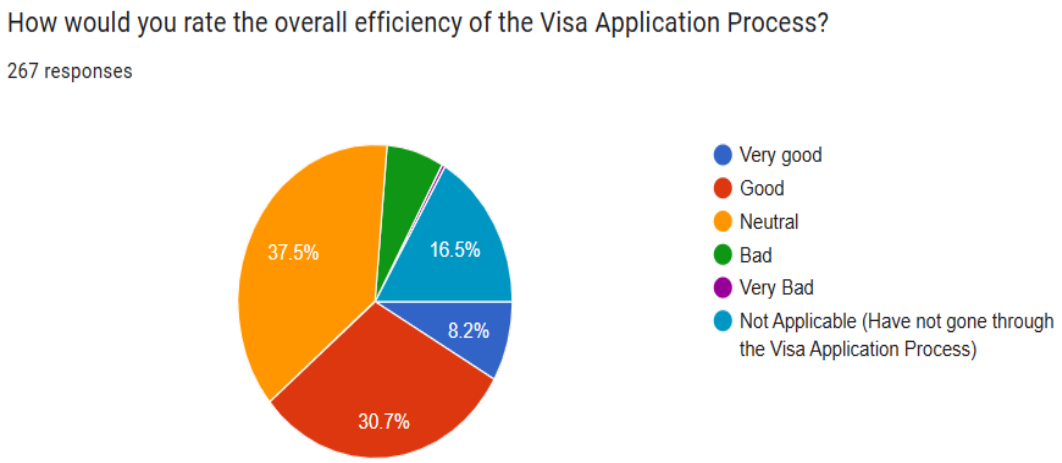

Figure 4.1: Overall efficiency of the current visa processing system

According to the survey results given in figure 4.1, majority of participants (37.5%), gave a neutral rating for overall efficiency. Only 8.2% gave a rating of "very good". Approximately 30% rated the current system merely as "good." This suggests that,

although the process is functional, it often fails to meet user expectations. The AI prototype developed in this study demonstrated a 40% reduction in document processing time, along with significantly fewer manual errors. That is, the current system's efficiency is perceived as average, with only marginal high satisfaction. Implementation of the AI prototype is strongly justified, as it demonstrably reduces processing time and errors.

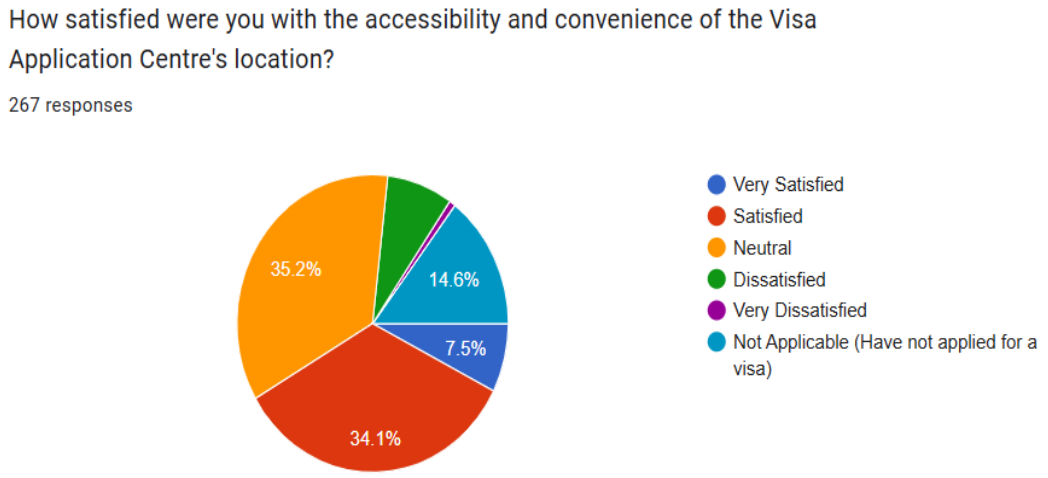

Figure 4.2: Accessibility and convenience of visa application centre's location

Figure 4.2 presents survey participants' perceptions of the accessibility and convenience of visa application centre locations. About 34.1% of participants rated the location ofvisa application centre as "satisfied," and 35.2% were neutral. Only 7.5% were "very satisfied," indicating moderate satisfaction with accessibility. The relatively large proportion of neutral responses indicates that many applicants do not perceive the location as either especially convenient or inconvenient, which reflects average rather than optimal service provision. At the same time, the low percentage of very satisfied

responses reveals that the system has not yet achieved strong approval in terms of accessibility. These findings suggest that while the survey participants consider visa centres as not highly inaccessible, they perceive the current visa processing system as an unfriendly experience. By contrast, an AI-enabled visa processing system would reduce reliance on physical visits by shifting much of the process online. Thus, while the current system achieves moderate accessibility, it lacks the inclusivity and convenience that an AI-driven model could provide.

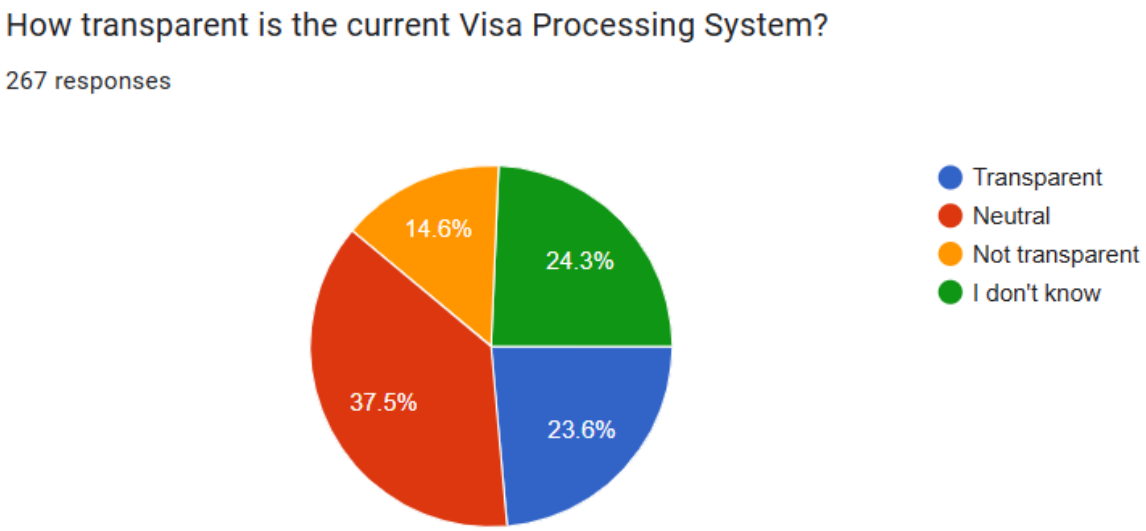

Figure 4.3: Transparency of current visa processing system

Figure 4.3 presents survey participants' perceptions of the transparency of the current visa system. Only 23.6% of participants rated the process as transparent, while nearly a quarter (24.3%) reported being unsure about it. This indicates that the majority of applicants either perceive the system as opaque or lack sufficient information to form a judgment about how decisions are made. When applicants are uncertain about the criteria

or rationale used in evaluating their cases, confidence in the system diminishes. This lack of perceived fairness and openness will fuel frustrations, especially when applications are delayed or rejected without clear explanations. An AI-based visa system can solve these problems by keeping clear records of how decisions are made and explaining the reasons for each outcome to applicants. This not only improves perceptions of fairness but also helps rebuild trust in the system as an impartial and accountable process.

How easy was it to get and fill out the required paper forms for your Visa Application?	
Neutral	41.198502
Easy	19.475655
Not Applicable (Have not obtained or filled out visa application forms)	17.228464
Difficult	13.857678
Very easy	6.367041
Very difficult	1.872659

Figure 4.4: Easiness of paperwork submission with the current visa processing system

Figure 4.4 illustrates participants' perceptions of the ease of paperwork submission within the current visa processing system. Only 19.5% of respondents described the paperwork submission process as smooth. The vast majority of over 80% reported at least one issue. These included long waiting times (8.6%), excessive documentation requirements (5.2%), difficulties with document verification (6.0%), and the general feeling that the process was overly drawn out (8.6%). Several participants gave the response that the instructions from the current system were unclear or that the forms were confusing. These responses reflect the lived experiences of applicants who

often feel frustrated when going through the current visa processing system that lack clarity and responsiveness.

To explore these issues in detail, participants were also asked to provide details about specific difficulties in paperwork submission, and the results are presented in Figure 4.5. The most frequent problem was a lengthy process (28.1%), followed by excessive documentation and waiting times. This confirms that time-consuming paperwork remains one of the biggest bottlenecks in the current system.

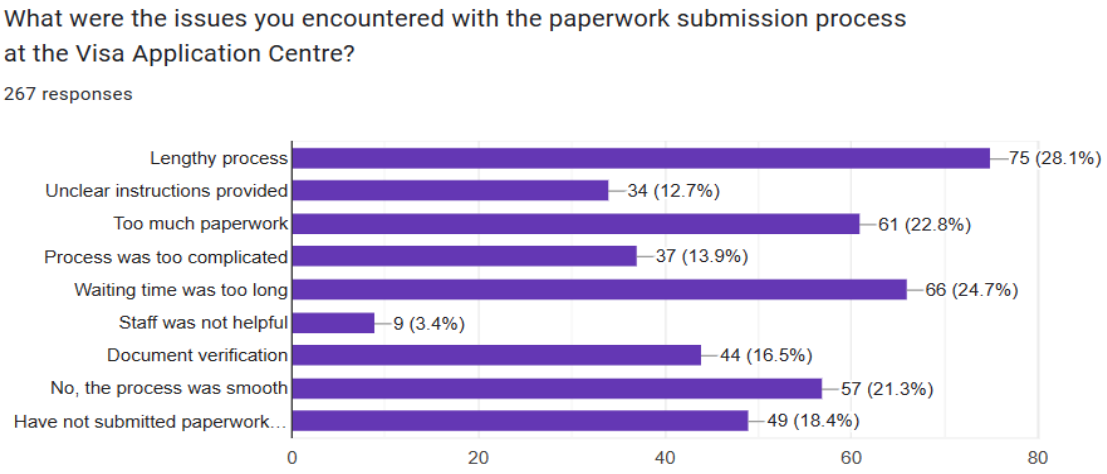

Figure 4.5: Issues encountered during paperwork submission with the current visa processing system

All these survey results align with the broader reality of visa processing in India, where current turnaround times are often lengthy. For instance, Indian e-Visas take a minimum of three working days after submission; Schengen visas require 1–4 weeks to secure an appointment and an additional 10–15 working days for processing; Canada

visitor visas typically take 15–30 working days; and U.S. visitor visas face appointment wait times of 6 months to over a year, with processing itself taking 1–2 weeks post-interview. Such delays and heavy paperwork demand an urgent need for more efficient systems. Against this backdrop, the AI-based approach developed in this study offers a data-driven pathway to substantially reduce time, paperwork, and manual intervention while improving transparency.

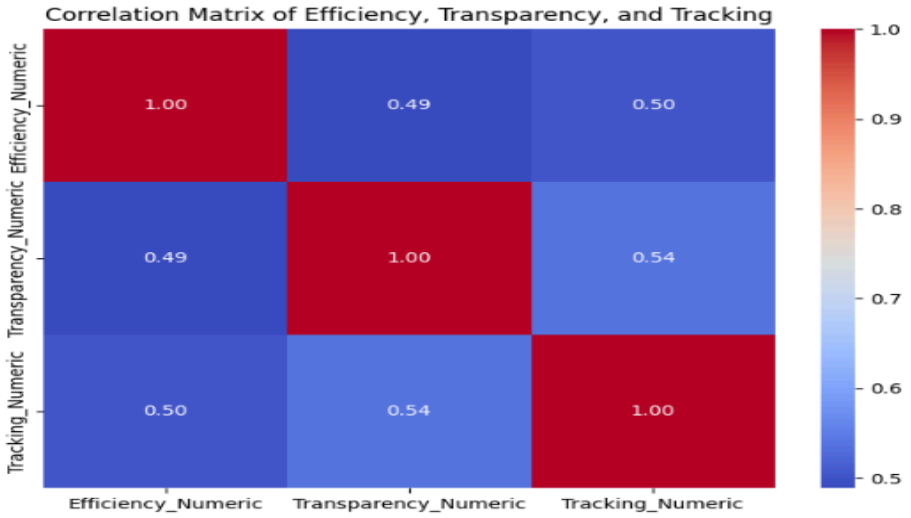

Figure 4.6: Correlation Matrix of Efficiency, Transparency, and Tracking

As shown in Figure 4.6, a correlation matrix was generated to understand the interrelationship between perceptions of efficiency, transparency, and tracking. The correlation between efficiency and transparency was 0.49, and between tracking and transparency was 0.54. This reveals that participants who rated one of these system

features positively tended to rate the others similarly—indicating that improvements in one area could positively influence user perception of the system as a whole.

In contrast, the AI-based prototype developed as part of this study was specifically designed to address these pain points. By using Optical Character Recognition (OCR) for document intake and Natural Language Processing (NLP) for interpreting and validating form data, the system significantly reduces reliance on paper and manual checks. Automation ensures faster processing, fewer errors, and real-time tracking updates—elements that directly address the issues participants raise most frequently. Importantly, the system also logs every decision path, enabling greater transparency and auditability. In essence, the survey confirms that people are less reluctant to an AI-based visa processing system, and there is a scope for an AI-based visa processing system which can transform a cumbersome, error-prone experience into a streamlined, automated process that centres on the needs of applicants. The combination of user sentiment and technical performance confirms that AI is not just a feasible replacement but a necessary and welcomed evolution of visa processing systems.

4.2 Research Question Two

What are the perceptions, preferences, and concerns of visa applicants, immigration officers, and the general public regarding the adoption of AI in visa processing?

The second research question aims to understand how different stakeholder groups perceive the adoption of AI in visa services—not only whether they support it, but also the factors that shape their support.

The willingness to adopt AI-driven visa processing systems is assessed through the survey conducted. Figure 4.7 illustrates the willingness to use Artificial Intelligence (AI)in a visa processing system to achieve fully automated visa processing. Notably, 88.6% of respondents indicated interest in using a fully automated AI-based system, with this support consistent across all age groups.

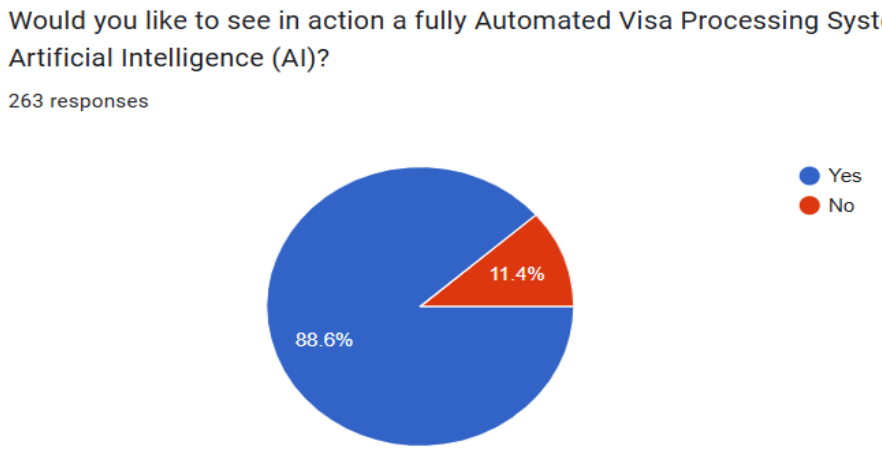

Figure 4.7: Willingness to fully Automated Visa Processing System using Artificial Intelligence (AI)

The survey results indicate that interest in AI-based visa processing is high across all demographics. As shown in Figure 4.8, the bar chart shows the interest to integrate AI in visa processing across different age categories. Among respondents aged 18–29,

84.8% were in favour of AI integration. This support remained strong in the 30–49 group (87.5%) and increased further in the 50–69 group (91.6%). Remarkably, every respondent in the 70+ age group indicated support for the system (100%). This challenges the common assumption that older adults are less receptive to new technologies. One possible explanation is that participants in this age category had travelled abroad at least once in their lifetime and were relatively well-educated compared to others of the same age group.

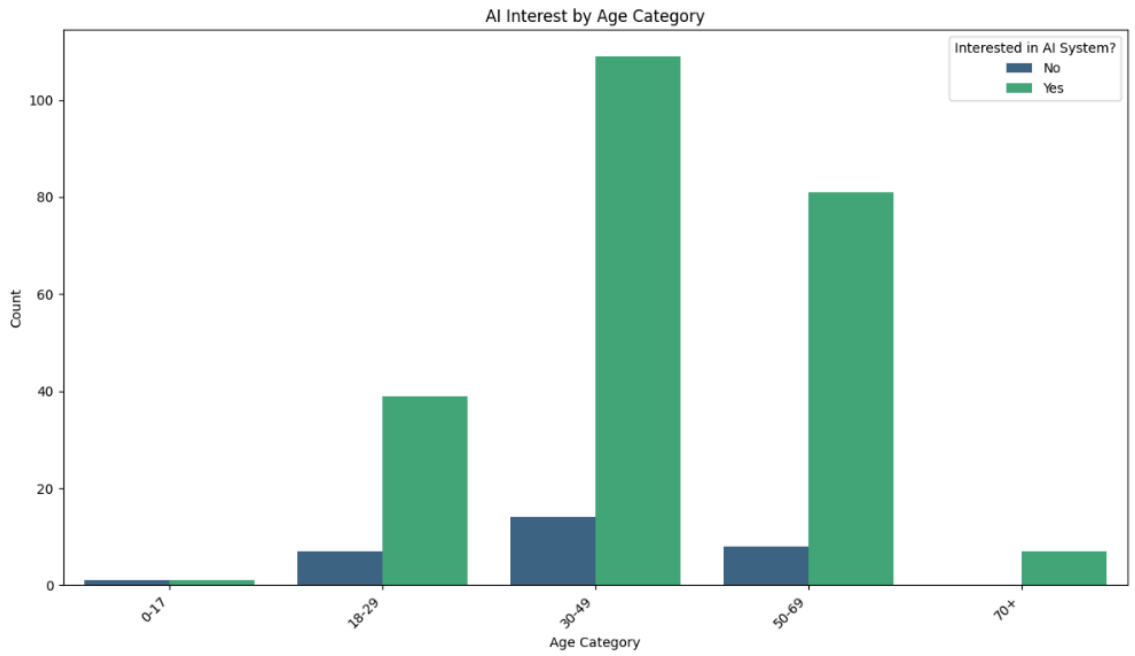

Figure 4.8: AI Interest by Age Category

However, this high level of interest does not imply participants have a blind acceptance of AI in the system. Open-ended responses revealed a more nuanced picture, especially after thematic and sentiment analysis. Many participants expressed the desire

for faster processing, real-time application tracking, and a reduction in paperwork as primary motivators for supporting the use of AI. Lack of clarity in cases where documents are delayed or rejected without clear reasons is the issue faced by several participants.

At the same time, some participants expressed concerns about the use of AI in visa processing systems. Their concern was the fairness and transparency in how AI systems make decisions. Respondents firmly believe that they need to understand the reason when an application is delayed or denied, especially when AI is making decisions; AI should not completely replace human oversight. Some participants also expressed concerns about data privacy, particularly regarding how personal documents and biometric data would be stored and processed. The thematic and sentiment analysis suggests that there is an intense desire for AI-driven transformations in visa processing systems, but clear ethical boundaries and human accountability should accompany this.

The sentiment analysis conducted on open-ended responses further illustrates this balance between optimism and caution. Most responses are skewed toward the positive or neutral regarding suggestions for AI adoption in the system, with an average polarity score of +0.34, indicating mild to strong favorability for adopting AI in the system. Positive sentiment often reflected excitement about modernisation and time savings, while neutral tones captured wait-and-see attitudes or conditional acceptance. Even

though negative sentiment was less frequent, it should be considered seriously. There are concerns for ensuring fairness, transparency, and control while adopting AI in the system.

These insights show that support for AI is strong but conditional. People want systems that are not only fast and efficient but also explainable, secure, and inclusive.

4.3 Research Question Three

How can AI technologies, such as OCR, NLP, sentiment analysis, and document understanding, be effectively integrated to automate and streamline visa processing tasks?

This research question aims to investigate the technical feasibility and strategic integration of AI components within the visa processing workflow. The AI prototype developed for this study integrates several advanced AI technologies, including Optical Character Recognition (OCR), Natural Language Processing (NLP), sentiment analysis, text spotting, scene comprehension, and document classification. These advanced technologies help to create an intelligent, automated system tailored to address real-world inefficiencies in current visa services.

OCR is integrated into the prototype to extract structured data from scanned forms and documents such as passports, identification records, and visa applications. Thus, OCR reduces the need for manual data entry, significantly reducing errors and saving

processing time. Data, such as applicant names, travel history, and passport numbers, were reliably extracted, enabling rapid digital processing from the initial intake.

NLP is integrated into the prototype to interpret unstructured text data, including travel justifications and employment descriptions. This allowed the system to assess completeness, validate semantic coherence, and detect anomalies or missing inputs in user-submitted narratives. In doing so, NLP added a layer of intelligent content analysis previously only achievable through human reviewers.

Sentiment analysis provided qualitative insights into user feedback and application communications. It helps to monitor user satisfaction and detect concerns or confusion within applicant comments. This information can inform user experience improvements and guide helpdesk interactions.

Document understanding tools, including intelligent classification algorithms, were used to categorise uploaded documents and match them to visa-type requirements automatically. Documents such as financial statements and invitation letters were sorted and validated based on their content and expected relevance, thereby minimising the need for manual sorting and cross-checking.

Collectively, these technologies help build a streamlined workflow that replaces repetitive administrative tasks with automation while retaining transparency through

audit logs and status dashboards. By reallocating human oversight to a minimum, the AI system preserves fairness and accountability, making the system fast and error-free.

In conclusion, integrating these advanced technologies into the current system gives a powerful solution to the challenges of current visa processing.

4.4 Summary of Findings

The results given in this chapter provide comprehensive evidence that supports the central objectives of this research, which are to explore the feasibility, acceptance, and practical integration of an AI-based visa processing system. Across all three research questions, the findings consistently highlight a strong public demand for digital transformation and demonstrate the technical viability of AI in improving administrative workflows.

The first research question investigated whether an AI-based system could reduce paper usage, processing time, and human error. The survey responses confirmed widespread dissatisfaction with the current manual process, particularly regarding paperwork, delays, and inefficiency. System testing demonstrated that the AI prototype eliminated paper forms through digital intake and automated data extraction and reduced human error through structured validation mechanisms.

The second research question aims to understand the perceptions, preferences, and concerns of visa applicants, immigration officers, and the general public regarding the integration of AI in the visa processing process. More than 84% of participants across all age groups expressed support for AI integration into the system. Notably, even participants aged 50 and above showed strong enthusiasm for AI integration in the system. This may be because aged participants in this survey have travelled abroad at least a few times in their lifetime. Thematic and sentiment analysis confirmed that users seek greater speed, convenience, and transparency, but also value fairness, privacy, and human oversight in automated systems.

The third research question examined how AI technologies, including OCR, NLP, sentiment analysis, and document classification, can be integrated efficiently and effectively into the visa processing system. The prototype proved that these tools can effectively replace repetitive manual tasks while providing accurate document parsing and tracking features.

4.5 Conclusion

This chapter presents the findings from both the survey and the evaluation of the AI system prototype. Each section corresponded to one of the study's three research questions, enabling a focused discussion of user experiences, perceptions, and technical implementation possibilities.

The first research question found that the current visa system is slow and frustrating. Participants had expressed problems with paperwork, tracking their applications, and unclear instructions in the current working system. The new AI prototype system has significantly improved these areas, primarily by reducing manual work and errors.

For the second research question, the responses revealed strong public support for an AI-driven system for visa processing across all age groups. They expressed a strong desire for speed, convenience, and transparency in the current visa processing system. At the same time, some participants expressed their valid concerns about fairness, privacy, and human oversight. This should be considered seriously when integrating AI into visa processing.

The third research question investigates how AI can be utilised in processing visa applications. The AI prototype successfully utilised OCR to read documents, NLP to understand written text, sentiment analysis to interpret feedback, and intelligent classification to automatically sort documents. These tools worked together to make the system faster, more transparent, and more user-friendly.

Taken together, these findings strongly support the overall research aim. The results of this study strongly validate the hypothesis that AI can transform visa processing by improving efficiency, reducing errors, and enhancing the applicant experience.

CHAPTER V:

DISCUSSION

5.1 Discussion of Results

This research investigates whether artificial intelligence (AI) can transform the current visa processing system. The findings presented in the previous chapter strongly affirm this hypothesis. The survey results revealed significant limitations in the existing system. The current system has a heavy reliance on paperwork, long waiting times, and a lack of transparency in application status updates. Participants often expressed frustrations with the limitations of the current visa processing system. These sentiments, as computed from the survey, highlight that the current system not only imposes unnecessary burdens on individuals but also undermines trust in administrative efficiency.

At the same time, there was a clear and consistent interest in modernising the system through digital tools and automation. Participants of the survey expressed optimism toward AI-powered solutions that could streamline documentation, minimise errors, and provide real-time updates, thereby reducing uncertainty. From system testing, AI-based visa processing demonstrated its ability to handle repetitive verification tasks more quickly and accurately than manual review. This suggests that AI could significantly reduce administrative workload at visa processing centres. Furthermore, AI-based anomaly detection and fraud prevention mechanisms offer an additional layer of reliability, ensuring that applications are processed both efficiently and securely.

The study also highlighted that AI-driven enhancements could benefit multiple stakeholders simultaneously. For visa applicants, AI enables reduced paperwork, a faster and more transparent process with real-time tracking updates. At the same time, an AI-based visa processing system translates into lower operational costs, improved accuracy, and reduced staff burden. Notably, the environmental sustainability of a paperless, AI-supported system was also recognised as a significant advantage. Overall, the results confirm that integrating AI into visa processing can reform the core weaknesses of the existing system while aligning with broader global trends in digital governance and smart administration.

The AI prototype system developed in this research is highly effective in addressing many of the challenges that participants had consistently reported in their visa application experiences through the survey. One of the most significant improvements was the drastic reduction in paperwork. In the current visa processing system, visa applicants have to submit hundreds of pages of supporting documents. This process was not only time-consuming and costly but also prone to errors and duplication. In contrast, the AI prototype visa processing system streamlined this by enabling digital submission and automated document verification, and hence reduced manual effort and minimised the risk of misplaced or misfiled information. This shift toward a paperless process was also welcomed for its environmental benefits, as it eliminated unnecessary printing and photocopying.

Also, the AI visa processing system prototype has the ability to track application status in real time. In the current system, applicants often face long periods of uncertainty due to fewer opportunities for application tracking, and this lack of transparency often creates anxiety and frustration in applicants with urgent travel, employment, or educational deadlines. The AI-based visa processing system resolved this issue by incorporating a visa processing status-tracking feature. With this feature, visa applicants could monitor progress at each stage of review. This feature will keep them informed and create trust in the process.

From a functional perspective, the AI visa processing system prototype performed well with repetitive administrative tasks that previously required considerable manual effort. For example, document checks, cross-verification of details, and error detection were completed with greater speed and accuracy than human staff could achieve under time and workload constraints. This shortened processing times and reduced the likelihood of human error, and thus ensures greater reliability and consistency in outcomes. Participants of the survey reported higher levels of confidence with the AI-enhanced system. The combination of efficiency, accuracy, and transparency reassured participants that their cases were being handled fairly and effectively. Also, visa applicants will no longer feel that they are left in the dark or burdened by unnecessary administrative hurdles. Instead, they can experience a more user-friendly,

streamlined, and modernised process that aligns with their expectations of digital services in today's world.

Also, while the majority of participants expressed their enthusiasm about the improvements that can be brought by the AI-based visa processing system, many expressed a sense of caution around issues of fairness, respect, and accountability within the system. This highlights an important dimension of technology adoption: efficiency alone is not enough to guarantee user trust. Participants of all ages or professional backgrounds consistently raised concerns about how sensitive data would be handled and whether adequate safeguards would be taken to protect privacy. Given that visa applications often require the submission of highly confidential information—such as financial records, medical reports, and family details—there was an understandable apprehension about potential misuse, unauthorised access, or breaches of such data in a fully digitised system.

In addition to privacy concerns, participants are also worried about the role of AI involvement in the decision-making process. Some participants emphasised the importance of retaining a "human touch" in critical processes. This reflects the societal concern that automation should complement rather than completely replace human judgment. This points out the need for balanced system design, where technology is integrated with strong ethical safeguards and human oversight. Participants suggested that an AI based visa processing system should include transparent decision-making

processes and stringent data protection protocols to build trust and legitimacy. Such measures would ensure that the visa process remains efficient and accurate, and also fair, respectful, and humane.

Overall, the findings indicate a clear need and a strong desire for an improved visa application process. AI offers the tools to meet that need, but its design must be thoughtful, transparent, and centred on people. The following sections will provide a detailed examination of how this applies to each research question.

5.2 Discussion of Research Question One

Can an AI-based visa processing system significantly reduce paper usage, processing time, and human errors compared to the traditional system?

The first research question is whether Artificial Intelligence (AI) could enhance the visa application process by making it more efficient and less reliant on paperwork and manual effort. The study results provide strong evidence that this is indeed possible. Also, many participants expressed frustration with the current visa processing system through the survey conducted for this study. The reported issues of the current visa processing system from the survey are excessive paperwork, lengthy waiting times, and errors resulting from manual entry and processing. These challenges often place applicants under stress, especially when they are already worried about travel schedules and visa deadlines. An AI prototype visa processing system was developed and tested in

this study, which evaluated the extent to which these issues could be addressed through automation and intelligent decision support.

The comparative evaluation between the current visa processing system and the AI-based prototype revealed substantial differences in operational efficiency, user experience, and error reduction. The current system, in the context of Indian e-Visas or high-demand international visas such as Schengen, Canadian, or U.S. categories, relies heavily on multi-stage manual visa processing, including manual checks and redundant documentation requirements. Appointment bottlenecks and clerical mistakes often result in further delays. Also, applicants are sometimes required to resubmit documents or start the process afresh. These issues were highlighted by several participants in the survey, which confirms that the frustrations are not isolated experiences. In contrast, the AI-based prototype visa processing system increases speed by automation, minimises redundant steps and accelerates decision-making. Applicants will enjoy guided form-filling, real-time error detection, and transparent status tracking. In effect, the AI-based approach shifts the visa processing system from being clerical and error-prone to becoming more applicant-friendly, accurate, and future-ready.

The AI prototype was explicitly designed to address these challenges. The AI prototype is carefully built after analysing the survey results. The system focuses mainly on three core improvements: reducing paperwork, minimising manual effort, and improving accuracy. Optical Character Recognition (OCR) was integrated into the

prototype system, which enabled the automatic scanning and reading of documents. This removed the need for repeated manual input and cross-checking, thereby reducing both paper usage and clerical errors. Previously common errors, such as misspelt names, incomplete forms, or mismatched document types, were identified and flagged by the system in real time to a great extent. In doing so, the prototype not only saved time but also prevented the kinds of small mistakes that often carried costly consequences for applicants.

Another significant improvement brought about by the prototype was speed. The prototype system could process applications at a much faster rate than the current visa processing system by automatically checking whether all required fields were completed and whether the correct documents had been submitted. In current systems, such tasks often require several days of clerical handling, with each step adding additional waiting time for the applicant. With the AI integration, these tasks were executed in seconds and thus significantly reduced the overall processing timeline. This improvement in processing will improve applicant satisfaction as well.

Transparency and status tracking also emerged as critical areas of improvement. In the current visa processing system, applicants often have little or no clarity about what stage of visa processing their application is in or how long the visa processing will take. This lack of clarity often causes frustration and anxiety in them. They must either wait in uncertainty or seek updates through cumbersome embassy hotlines or physical visits. The

AI prototype developed in this study resolved this issue by offering real-time tracking features that informed users about the status of their applications. This not only improved the applicant experience but also increased trust in the system.

Also, the prototype system's ability to eliminate paperwork, reduce waiting times, and prevent errors translated into a more streamlined and accessible process. Many participants reported frustrations with physical visits required for document submission or status inquiries. This was particularly significant for applicants living in rural areas or for those with limited mobility, who previously found the physical demands of the system burdensome. By minimising these challenges, the prototype demonstrated how AI can make visa processing more inclusive in addition to being more efficient.

At the same time, demographic insights revealed that there is some variation in how the system was received. Younger participants generally expressed enthusiasm about the integration of AI into visa processing. For them, the system's digital accessibility aligned with their familiarity with other technology-driven services in banking, education, and healthcare. Older participants showed cautious optimism about AI in the visa processing system, as expected. While they appreciated the improvements AI can bring in visa processing, they also raised concerns about the reduced role of humans and handling sensitive personal data in an automated system. These findings indicate that while AI systems can achieve remarkable operational benefits, people are cautious that

their broader adoption will require careful attention to ethical concerns such as fairness, data privacy, and accountability.

It is also important to note that the prototype was not simply a demonstration of what AI could achieve in a visa processing system. Instead, it was designed around the feedback and inferences from the survey. The system features, such as OCR-based document scanning, automated error detection, and real-time tracking, were integrated into the system to address the specific pain points identified by participants directly. This user-centred design is employed to ensure that the prototype is not only technologically advanced but also practical and relevant to the real challenges faced by applicants.

In summary, the findings strongly support that AI can significantly enhance the visa processing system. The prototype developed shows that there is a significant improvement in efficiency, transparency, and user satisfaction, showing that many of the frustrations associated with the current system can be alleviated through automation. By replacing outdated manual tasks with faster and more accurate AI tools, the visa processing system becomes more efficient and more responsive to the needs of applicants. At the same time, the results highlight that success will depend on addressing social and ethical concerns to build broad-based trust. Thus, while the benefits of AI are clear, the future of its adoption in visa processing will rest on balancing efficiency with fairness, inclusivity, and accountability.

5.3 Discussion of Research Question Two

What are the perceptions, preferences, and concerns of visa applicants, immigration officers, and the general public regarding the adoption of AI in visa processing?

The second research question focused on understanding the perceptions, preferences, and concerns of visa applicants, immigration officers, and the general public about the Artificial Intelligence (AI) adoption in the current visa processing system. Understanding these perspectives is important because the technological adoption of AI in visa processing matters notably in terms of efficiency but also social acceptance, trust, and accountability. To gather the perceptions, preferences, and concerns of visa applicants, immigration officers, and the general public, a survey was conducted among them. The results of the survey demonstrated that there is a balance between enthusiasm for efficiency gains and apprehension about ethical implications.

The responses to the survey were both encouraging and insightful. A majority of participants expressed strong interest in AI visa processing systems, which may be motivated primarily by their desire to overcome the limitations of the current system. Participants repeatedly reported that they face challenges like lengthy waiting periods, huge paperwork, and frequent delays with the current visa processing. These challenges often create stress for applicants with urgent travel needs. For such participants, AI in visa processing is not just an upgrade but a necessary modernisation of an outdated bureaucratic structure. By automating repetitive tasks, reducing manual entry errors, and

providing more apparent timelines, AI was seen as a tool that could transform the applicant experience from frustration and uncertainty to one of predictability and ease.

An important analysis area of this study is the cross-demographic support for AI-based visa processing. Participants from diverse backgrounds, including different age groups, professional sectors, and levels of digital and academic literacy, were included in the survey conducted to get diverse and stratified participation. The majority of participants from all these sectors acknowledged the need for a system that was faster, simpler, and more transparent. Participants under 35, who are already familiar with digital platforms in banking, education, and commerce, viewed AI in visa processing as a natural extension of broader digital transformation trends. Older participants, who are above 55 years old, showed more enthusiasm about AI in visa processing. This may be because participants in this age group are individuals who have travelled at least once in their lifetime. They may have encountered additional challenges in obtaining visas, as they may have experienced delays in the past or difficulties in completing paperwork with the visa processing system. At that time, visa processing was much more difficult. Alternatively, they may find the current visa processing system more challenging to navigate. Participants between the ages of 35 and 55 showed both enthusiasm and concerns about AI in visa processing. Most of them expressed appreciation for the time-saving benefits and reduction in procedural complexity. Despite this enthusiasm, the responses also reflected some critical concerns that cannot be neglected while adopting AI in visa processing. Most participants in this age category raised questions about

fairness in AI adoption for processing visas. They are also concerned about whether AI systems would make correct and impartial decisions and whether AI algorithms might be advantageous or disadvantageous to certain applicants. These concerns resonate with debates across the world about algorithmic bias in AI applications, in which opaque decision-making processes have at times led to unfair or discriminatory outcomes. For visa processing, where decisions carry life-changing consequences for applicants, the concerns are valid and require more emphasis on mitigating such challenges. Thus, the fairness of AI in visa processing is non-negotiable, and participants stressed the importance of building systems that are transparent in how they evaluate cases. Notably, even though these concerns are high among middle-aged people, they are seen in all age categories. This can be seen as participants see AI as the current technology, and they are thinking about how to avoid mistakes that can come from using AI, which is quite welcoming. A solution to these concerns is using explainable AI and human oversight in visa processing systems.

Another central area of concern for participants was data privacy while using AI in visa processing. Visa processing often involves storing sensitive personal details, financial information, and official documents. Participants worried about how this data would be managed in an AI-driven system. Their primary concerns are where the data would be stored, who would have access to it, and whether it might be vulnerable to misuse or security breaches. Also, AI could be used to gain inference from this sensitive data in a few seconds. In the current global environment, where digital data breaches and

unauthorised surveillance have made headlines, these concerns were especially salient. Respondents were clear in their expectation that any AI visa system must adhere to the highest standards of data security, comply with legal protections, and maintain transparency about data handling practices.

Closely tied to these concerns was the question of human involvement. Many participants reported that while they welcomed AI as a supportive tool, they did not want AI to replace human officers fully. They pointed out the human element as essential for ensuring compassion, contextual understanding, and decision-making that algorithms, no matter how advanced, might lack. For example, a rigidly programmed AI system might reject an application based on a minor technical inconsistency. In contrast, a human officer could recognise extenuating circumstances and make a more empathetic judgment. A counterargument for this work is that AI would notify the technical inconsistency to the applicant and provide time to rectify it. All these imply that participants value a hybrid model in which AI handles routine, repetitive tasks while human officers maintain oversight and make final decisions in complex cases.

The findings thus point toward a layered perception: intense excitement about efficiency and convenience, coupled with a deep insistence on fairness, privacy, and accountability. Participants of the survey were clear that they need AI in visa processing to support them, not dominate. The AI integration in the visa processing system should streamline the application journey without compromising the safeguards and human

touch necessary for sensitive decision-making. This suggests that the successful adoption of AI in visa processing will require careful system design that balances speed with ethical responsibility. Participants are expecting transparent communication in the visa processing system, especially with AI adoption. Also, they want to understand how the system works and where human oversight comes in. By making these processes visible, institutions can help build trust and reassure applicants that the technology respects their rights and interests.

Overall, the findings show that participants are optimistic about the integration of AI in visa processing. At the same time, they remain vigilant about the conditions under which it is deployed. They value the potential for speed, convenience, and accuracy that AI can bring into the visa processing system. However, they need the system to safeguard fairness, privacy, and human judgment, and they expect the government to bring strict rules and regulations for achieving it. In short, stakeholders of visa processing support for AI in visa processing are strong, but they come with thoughtful expectations. The opportunity for modernisation is clearly present, but its success will depend on introducing technology in a way that earns and sustains public trust. If these conditions are met, AI adoption in visa processing can move from being a promising innovation to a trusted and widely embraced reality.

5.4 Discussion of Research Question Three

How can AI technologies, such as OCR, NLP, sentiment analysis, and document understanding, be effectively integrated to automate and streamline visa processing tasks?

The third research question examines the role of Artificial Intelligence (AI) technologies, specifically Optical Character Recognition (OCR), Natural Language Processing (NLP), sentiment analysis, and document understanding, in improving the visa application process. The AI visa processing prototype developed in this study is used to evaluate how these technologies can be effectively combined to address the challenges with visa processing. The results demonstrated that when these tools are appropriately integrated into a visa processing system, they can collectively create a robust and efficient system.

Optical Character Recognition (OCR) is a main element of the prototype system architecture. Visa applications often involve large volumes of physical or scanned documents, including passports, identity cards, application forms, and supporting evidence such as financial statements or employment records. Currently, the information contained in these documents must be manually typed into digital systems, which is slow, costly, and prone to human error.

By employing OCR, the prototype system could automatically scan, extract, and digitise text from these documents with high accuracy. This reduces manual intervention

at the data entry and allows immigration officers to focus their attention on substantive decision-making tasks rather than clerical work. In addition to efficiency, OCR also provided consistency by extracting text in a uniform format and thus enables downstream AI tools (such as NLP engines) to process the data more effectively. This automation of text extraction thus represented the first step toward building a streamlined digital workflow.

Once the text was extracted, Natural Language Processing (NLP) was deployed to analyse and interpret the content. In visa processing, applicants are often required to provide free-text responses explaining their reasons for travel, employment status, or other contextual information. In the current visa processing system, these responses are manually reviewed by embassy staff. It is a time-consuming and inconsistent process across evaluators. NLP helps the system to read and understand these written statements. Through semantic analysis, it could detect whether responses were complete, relevant, and logically consistent. For example, if an applicant stated “visiting for family members,” NLP could verify whether supporting evidence was submitted by the applicant or if contradictory information existed elsewhere in the file. In this way, NLP reviews or evaluates the written content and submitted documents and provides structured insights that support decision-making.

NLP could also be used for language translation. This allows staff to process applications submitted in different languages. This expanded accessibility for applicants

worldwide and reduced delays caused by language barriers. Thus, NLP played a vital role in turning unstructured narratives into actionable insights.

Sentiment analysis introduced a human-centred dimension in this prototype. In the prototype system developed, applicants can express emotions such as frustration, anxiety, or satisfaction when interacting with the system. These emotional cues often offer valuable insights into system performance. The inclusion of sentiment analysis allowed the prototype system to interpret the polarity of user feedback from feedback forms or customer service interactions. For instance, a high frequency of negative sentiment associated with application tracking delays could often indicate areas where system improvements were most urgently needed. Similarly, positive feedback about real-time status updates confirmed the importance of transparency features. By analysing aggregated emotional responses, sentiment analysis provided actionable feedback loops that helped refine the system design to be more user-friendly and responsive. This capability also had a preventive dimension: continuous monitoring of applicant sentiment could help administrators anticipate dissatisfaction or confusion before it escalated into formal complaints.

Another key AI tool integrated into the prototype was an intelligent document classification. Visa processing often involves a huge number of supporting files, ranging from financial statements and employment letters to medical records, police clearances, and identity documents. Mislabeling or misinterpretation of such documents can

significantly delay visa processing and compromise decision accuracy. Using AI classification algorithms, the system was able to automatically recognise and categorise each uploaded file into its appropriate category. For example, the prototype could distinguish between a bank statement and an employment verification letter even if both were submitted in the same format. By assigning documents to their correct categories, classification ensured that immigration officers could review cases in a structured and logical sequence. This automation minimised clerical mistakes, accelerated case review, and provided a standardised structure across applications, regardless of how applicants chose to upload their documents. As a result, document classification not only enhanced efficiency but also contributed to procedural fairness by ensuring that similar types of evidence were consistently recognised and evaluated.

While each of these technologies- OCR, NLP, sentiment analysis, and document classification played a unique role, their true power lay in integration. The prototype demonstrated how these tools could operate as a cohesive ecosystem. OCR is used for digitising the data, NLP structured it into meaningful insights, document classification organised supporting evidence, and sentiment analysis captured user perspectives. Together, this workflow transformed visa processing from a paper-heavy procedure into a streamlined digital pipeline. The integrated system improved speed, transparency and traceability. By embedding explainability and traceability into the system design, AI made the process not only faster but also more accountable.

For applicants, the integration of these AI tools translated into reduced delays, fewer manual requirements, and greater clarity throughout the process. The ability to track applications in real time, combined with automated document categorisation, made the system more accessible and less intimidating. For administrators, AI reduced repetitive workload, standardised case evaluation, and allowed staff to redirect their expertise toward higher-order decision-making.

At the same time, the study suggests that these technologies are most effective when deployed as support systems rather than replacements for human oversight. Thus, the findings supported a hybrid model in which AI and human officers collaborate to achieve both efficiency and fairness.

In summary, the integration of OCR, NLP, sentiment analysis, and document classification demonstrated a promising pathway toward automating and streamlining visa processing. Each technology made a distinct contribution, such as OCR reduced manual entry, NLP enabled comprehension of unstructured text, sentiment analysis improved user experience, and document classification ensured accuracy and organisation. When combined, these tools created a system that was faster, more accurate, more transparent, and ultimately more people-centred.

This research shows that if AI is carefully integrated into the visa processing system, it can modernise visa processing into a process that balances efficiency with

accountability. By addressing both administrative bottlenecks and user concerns, AI-driven systems have the potential to set new standards for fairness, transparency, and service delivery in immigration management.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

This study examined the potential of Artificial Intelligence (AI) to enhance the visa processing system by reducing paperwork, streamlining application procedures, and minimising human errors. The study used three research questions to examine current user experiences, gather public perceptions, and evaluate the integration of AI tools, including OCR, NLP, sentiment analysis, and document classification.

A structured survey collected responses from 267 participants across diverse age groups, occupations, and travel frequency. The survey results indicated that most participants reported the traditional visa system to be inefficient, paper-heavy, and lacking transparency. Many respondents reported some other challenges, including long waiting times, errors in document verification, and difficulty tracking their application status.

To address these issues, a digital AI prototype was developed and evaluated. The AI-based visa processing system automates tasks such as visa processing data extraction, document classification, and visa application tracking, and it does so efficiently and effectively to a great extent compared to the current system. It provided users with a more transparent and responsive experience, featuring reduced delays and enhanced accuracy. Survey responses also showed strong support for AI adoption, even though concerns

about fairness, data privacy, and the need for human oversight were also expressed, which are valid concerns as well.

6.2 Implications

This study has significant implications for how visa applicants, immigration officers, and the general public approach artificial intelligence and its effectiveness in visa processing systems. The survey conducted for this study found that individuals face difficulties, delays, confusion, and a lack of transparency in their visa applications with the current system. The responses from participants in the survey conducted in this study suggest that there is a growing demand for visa processing systems that are not only faster but also easier to understand and more reliable.

One major takeaway from the study is that people from all backgrounds, not just young or tech-savvy individuals, are ready for more innovative digital systems. This breaks the common assumption that older users or less experienced travellers might resist change. Instead, the findings show that when technology is designed well and solves real problems, people are open and even enthusiastic about using it.

The prototype system developed in this research demonstrated how AI tools, such as OCR and NLP, can automate repetitive and error-prone tasks. This helps make the visa process smoother and frees up human staff to focus on more important or complex cases.

Adding features like real-time tracking and feedback analysis also made the system more transparent and responsive to users' needs.

However, this study also reminds us that people care deeply about fairness and privacy. While they welcome AI in visa processing for a faster and more convenient system, they also want to ensure that their information is secure and that decisions are made fairly and transparently. This means that systems should incorporate innovative technology, as well as clear rules, human oversight, and transparent communication about how decisions are made.

In short, the findings suggest that AI has the power to improve public services in meaningful ways—but only if it is used thoughtfully. It must serve the people who depend on these services, and it must be built with trust, transparency, and ethics at the core.

6.3 Recommendations for Future Research

This study demonstrated that AI can enhance visa processing while also highlighting several areas where further research is required. These suggestions guide future work and make sure that AI is used in the best possible way.

First, future research should be done on AI-based visa systems in real government settings. In this study, the system was tested as a prototype. It is essential to evaluate how

an AI-driven visa processing system performs in a real-world scenario. That is how the system handles live data, larger workloads, and more complex cases, as this helps us understand its practical value and how people utilise it.

Second, there's a need to understand the ethical aspects of AI better. People want systems that are not only fast and accurate but also fair and respectful. Future studies could explore how to make AI decisions easier to explain and how to maintain human involvement in key steps to prevent any unfair outcomes.

Lastly, since visa rules vary across countries, future research could examine how well AI systems perform in other legal or cultural contexts. Each country has their own laws and cultural norms, and a visa processing system that is effective in one country may require modifications to function well in another.

In short, future research should focus on testing AI in real-world environments, its ethical and fair use, and adapting it for different countries. This will help AI become a tool that truly supports people in a meaningful way.

6.4 Conclusion

This study aimed to determine if artificial intelligence could enhance the visa application process by making it faster, more reliable, and more convenient. Through a mix of user feedback from the online survey conducted and testing of the AI prototype

system developed, the answer became clear: AI has the potential to bring real improvements, but how it is used matters just as much as what it can do.

The findings showed that people from all backgrounds are ready for a more intelligent, more transparent visa processing system. They want to move slow, paper-heavy processes into a space where technology supports them. The AI prototype designed in this research was able to reduce manual effort, improve accuracy, and offer better communication to users, all of which are crucial in moments that often carry personal or professional stress.

At the same time, this research reminded us that transparency is essential. People want systems that not only work well but also treat them fairly and protect their information. They want to feel heard, respected and reassured even when the process is automated. This means building AI that incorporates room for human judgment, provides clear explanations, and features robust privacy safeguards.

Ultimately, technology should serve people, not replace them. The success of any digital transformation, especially in public services, will depend on whether we keep people at the centre of the design. This thesis offers not just a working model but a path forward: one where AI can enhance how systems function and how people feel when using them.

References

Abirami, A.M. and Gayathri, V., 2017, January. A survey on sentiment analysis methods and approach. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 72-76). IEEE.

Belinkov, Y., Poliak, A., Shieber, S.M., Van Durme, B. and Rush, A.M., 2019. Don't take the premise for granted: Mitigating artifacts in natural language inference. arXiv preprint arXiv:1907.04380.

Bharadiya, J., 2023. A comprehensive survey of deep learning techniques natural language processing. European Journal of Technology, 7(1), pp.58-66.

Biten, A.F., Tito, R., Gomez, L., Valveny, E. and Karatzas, D., 2022, October. Ocr-idl: Ocr annotations for industry document library dataset. In European Conference on Computer Vision (pp. 241-252). Cham: Springer Nature Switzerland.

Blodgett, S.L., Barocas, S., Daumé III, H. and Wallach, H., 2020. Language (technology) is power: A critical survey of" bias" in nlp. arXiv preprint arXiv:2005.14050.

Busta, M., Neumann, L. and Matas, J., 2017. Deep textspotter: An end-to-end trainable scene text localization and recognition framework. In Proceedings of the IEEE international conference on computer vision (pp. 2204-2212).

Cambria, E., Poria, S., Gelbukh, A. and Thelwall, M., 2017. Sentiment analysis is a big suitcase. IEEE Intelligent Systems, 32(6), pp.74-80.

Dalarmelina, N.D.V., Teixeira, M.A. and Meneguette, R.I., 2019. A real-time automatic plate recognition system based on optical character recognition and wireless sensor networks for ITS. Sensors, 20(1), p.55.

Dash, B., 2021. A hybrid solution for extracting information from unstructured data using optical character recognition (OCR) with natural language processing (NLP). Research Gate.

Devika, M.D., Sunitha, C. and Ganesh, A., 2016. Sentiment analysis: a comparative study on different approaches. Procedia Computer Science, 87, pp.44-49.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019, June. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

Durrani, N., Sajjad, H. and Dalvi, F., 2021. How transfer learning impacts linguistic knowledge in deep NLP models?. arXiv preprint arXiv:2105.15179.

Feng, S.Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T. and Hovy, E., 2021. A survey of data augmentation approaches for NLP. arXiv preprint arXiv:2105.03075.

Feng, W., He, W., Yin, F., Zhang, X.Y. and Liu, C.L., 2019. Textdragon: An end-to-end framework for arbitrary shaped text spotting. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9076-9085).

Garrido-Muñoz, I., Montejo-Ráez, A., Martínez-Santiago, F. and Ureña-López, L.A., 2021. A survey on bias in deep NLP. Applied Sciences, 11(7), p.3184.

Ghaeini, R., Fern, X.Z. and Tadepalli, P., 2018. Interpreting recurrent and attention-based neural models: a case study on natural language inference. arXiv preprint arXiv:1808.03894.

Gonçalves, P., Araújo, M., Benevenuto, F. and Cha, M., 2013, October. Comparing and combining sentiment analysis methods. In Proceedings of the first ACM conference on Online social networks (pp. 27-38).

Hamad, K. and Kaya, M., 2016. A detailed analysis of optical character recognition technology. International Journal of Applied Mathematics Electronics and Computers, (Special Issue-1), pp.244-249.

Hegghammer, T., 2022. OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment. Journal of Computational Social Science, 5(1), pp.861-882.

He, P., Huang, W., Qiao, Y., Loy, C. and Tang, X., 2016, March. Reading scene text in deep convolutional sequences. In Proceedings of the AAAI conference on artificial intelligence (Vol. 30, No. 1).

Hedderich, M.A., Lange, L., Adel, H., Strötgen, J. and Klakow, D., 2020. A survey on recent approaches for natural language processing in low-resource scenarios. arXiv preprint arXiv:2010.12309.

Jaderberg, M., Simonyan, K., Vedaldi, A. and Zisserman, A., 2016. Reading text in the wild with convolutional neural networks. International journal of computer vision, 116, pp.1-20.

Jaderberg, M., Vedaldi, A. and Zisserman, A., 2014. Deep features for text spotting. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13 (pp. 512-528). Springer International Publishing

Kang, Y., Cai, Z., Tan, C.W., Huang, Q. and Liu, H., 2020. Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics, 7(2), pp.139-172.

Kiela, D., Bartolo, M., Nie, Y., Kaushik, D., Geiger, A., Wu, Z., Vidgen, B., Prasad, G., Singh, A., Ringshia, P. and Ma, Z., 2021. Dynabench: Rethinking benchmarking in NLP. arXiv preprint arXiv:2104.14337.

Li, J., Chen, X., Hovy, E. and Jurafsky, D., 2015. Visualizing and understanding neural models in NLP. arXiv preprint arXiv:1506.01066.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z. and Wei, F., 2023, June. Trocr: Transformer-based optical character recognition with pre-trained models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 11, pp. 13094-13102).

Liao, M., Wan, Z., Yao, C., Chen, K. and Bai, X., 2020, April. Real-time scene text detection with differentiable binarization. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 11474-11481).

Liu, B., 2010. Sentiment analysis and subjectivity. Handbook of natural language processing, 2(2010), pp.627-666.

Liu, B., 2010. Sentiment analysis: A multi-faceted problem. IEEE intelligent systems, 25(3), pp.76-80.

Liu, X., He, P., Chen, W. and Gao, J., 2019. Multi-task deep neural networks for natural language understanding. arXiv preprint arXiv:1901.11504.

Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y. and Yan, J., 2018. Fots: Fast oriented text spotting with a unified network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5676-5685).

Liu, Y. and Zhang, M., 2018. Neural network methods for natural language processing.

Lopez, M.M. and Kalita, J., 2017. Deep Learning applied to NLP. arXiv preprint arXiv:1703.03091.

Lopresti, D., 2008, July. Optical character recognition errors and their effects on natural language processing. In Proceedings of the second workshop on Analytics for Noisy Unstructured Text Data (pp. 9-16).

Lyu, P., Liao, M., Yao, C., Wu, W. and Bai, X., 2018. Mask textspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes. In Proceedings of the European conference on computer vision (ECCV) (pp. 67-83).

Madhoushi, Z., Hamdan, A.R. and Zainudin, S., 2015, July. Sentiment analysis techniques in recent works. In 2015 science and information conference (SAI) (pp. 288-291). IEEE.

Malte, A. and Ratadiya, P., 2019. Evolution of transfer learning in natural language processing. arXiv preprint arXiv:1910.07370.

Mehta, P. and Pandya, S., 2020. A review on sentiment analysis methodologies, practices and applications. International Journal of Scientific and Technology Research, 9(2), pp.601-609

Memon, J., Sami, M., Khan, R.A. and Uddin, M., 2020. Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). IEEE access, 8, pp.142642-142668.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.

Mohammad, S.M., 2017. Challenges in sentiment analysis. A practical guide to sentiment analysis, pp.61-83.

Mohd, M., Qamar, F., Al-Sheikh, I. and Salah, R., 2021. Quranic optical text recognition using deep learning models. IEEE Access, 9, pp.38318-38330.

Namysl, M. and Konya, I., 2019, September. Efficient, lexicon-free OCR using deep learning. In 2019 international conference on document analysis and recognition (ICDAR) (pp. 295-301). IEEE.

Neri, F., Aliprandi, C., Capeci, F. and Cuadros, M., 2012, August. Sentiment analysis on social media. In 2012 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 919-926). IEEE.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B. and Ng, A.Y., 2011, December. Reading digits in natural images with unsupervised feature learning. In NIPS workshop on deep learning and unsupervised feature learning (Vol. 2011, No. 2, p. 4).

Nguyen, T.T.H., Jatowt, A., Coustaty, M. and Doucet, A., 2021. Survey of post-OCR processing approaches. ACM Computing Surveys (CSUR), 54(6), pp.1-37.

Phoenix, P., Sudaryono, R. and Suhartono, D., 2021. Classifying promotion images using optical character recognition and Naïve Bayes classifier. Procedia Computer Science, 179, pp.498-506.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, 21(1), pp.5485-5551.

Rijhwani, S., Anastasopoulos, A. and Neubig, G., 2020. OCR post correction for endangered language texts. arXiv preprint arXiv:2011.05402.

Rizvi, S.S.R., Sagheer, A., Adnan, K. and Muhammad, A., 2019. Optical character recognition system for Nastalique Urdu-like script languages using supervised learning. International Journal of Pattern Recognition and Artificial Intelligence, 33(10), p.1953004

Robby, G.A., Tandra, A., Susanto, I., Harefa, J. and Chowanda, A., 2019. Implementation of optical character recognition using tesseract with the javanese script target in android application. Procedia Computer Science, 157, pp.499-505.

Ruder, S., Peters, M.E., Swayamdipta, S. and Wolf, T., 2019, June. Transfer learning in natural language processing. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials (pp. 15-18).

Saeed, S., Naz, S. and Razzak, M.I., 2019. An application of deep learning in character recognition: an overview. Handbook of Deep Learning Applications, pp.53-81.

Sahlol, A.T., Abd Elaziz, M., Al-Qaness, M.A. and Kim, S., 2020. Handwritten Arabic optical character recognition approach based on hybrid whale optimization algorithm with neighborhood rough set. IEEE Access, 8, pp.23011-23021.

Serrano-Guerrero, J., Olivas, J.A., Romero, F.P. and Herrera-Viedma, E., 2015. Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311, pp.18-38.

Shah, S., Bhat, A., Singh, S., Chavan, A. and Singh, A., 2010. Sentiment analysis. International Journal of Progressive Research in Engineering Management and Science.

Sharifani, K., Amini, M., Akbari, Y. and Aghajanzadeh Godarzi, J., 2022. Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model. International Journal of Science and Information System Research, 12(9), pp.20-44.

Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M., 2011. Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), pp.267-307.Cambria,

E., Poria, S., Gelbukh, A. and Thelwall, M., 2017. Sentiment analysis is a big suitcase. IEEE Intelligent Systems, 32(6), pp.74-80.

Tenney, I., Das, D. and Pavlick, E., 2019. BERT rediscovers the classical NLP pipeline. arXiv preprint arXiv:1905.05950.

Van Strien, D., Beelen, K., Ardanuy, M.C., Hosseini, K., McGillivray, B. and Colavizza, G., 2020. Assessing the impact of OCR quality on downstream NLP tasks.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.

Vinodhini, G. and Chandrasekaran, R.M., 2012. Sentiment analysis and opinion mining: a survey. International Journal, 2(6), pp.282-292.

Vinyals, O., Toshev, A., Bengio, S. and Erhan, D., 2015. Show and tell: A neural image caption generator. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3156-3164).

Wang, H., Lu, P., Zhang, H., Yang, M., Bai, X., Xu, Y., He, M., Wang, Y. and Liu, W., 2020, April. All you need is boundary: Toward arbitrary-shaped text spotting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 12160-12167).

Wang, X., Wang, H. and Yang, D., 2021. Measure and improve robustness in nlp models: A survey. arXiv preprint arXiv:2112.08313.

Whitelaw, C., Garg, N. and Argamon, S., 2005, October. Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 625-631).

Wick, C., Reul, C. and Puppe, F., 2018. Calamari-a high-performance tensorflow-based deep learning package for optical character recognition. arXiv preprint arXiv:1807.02004.

Xu, Y., Li, M., Cui, L., Huang, S., Wei, F. and Zhou, M., 2020, August. Layoutlm: Pre-training of text and layout for document image understanding. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1192-1200).

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R. and Le, Q.V., 2019. Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.

Young, T., Hazarika, D., Poria, S. and Cambria, E., 2018. Recent trends in deep learning based natural language processing. ieee Computational intelligenCe magazine, 13(3), pp.55-75.

Yue, L., Chen, W., Li, X., Zuo, W. and Yin, M., 2019. A survey of sentiment analysis in social media. Knowledge and Information Systems, 60, pp.617-663

Zini, J.E. and Awad, M., 2022. On the explainability of natural language processing deep models. ACM Computing Surveys, 55(5), pp.1-31.

Zhang, X., Su, Y., Tripathi, S. and Tu, Z., 2022. Text spotting transformers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9519-9528).

Appendix A:
Survey Cover Letter

Subject: Request for Feedback – AI-based Digital Visa Processing System Survey

Dear Sir/Madam,

I hope this message finds you well.

As part of my academic journey in the Doctor of Business Administration (DBA) program at SSBM, I am currently engaged in the research and development of a digitalized visa processing system. This project aims to explore how Artificial Intelligence (AI) can be integrated into public service delivery to enhance efficiency, transparency, and user experience.

Your insights, drawn from your personal or professional experience with visa application processes, are incredibly valuable to this initiative. We are particularly interested in understanding your expectations, preferences, and any challenges you may have encountered with current systems.

I would be grateful if you could complete the attached survey. Your responses will directly contribute to refining our AI-based prototype, ensuring that the final solution is user-centric and widely applicable.

To participate in the survey, please click on the following link: Survey form for AI-based VISA Processing System

Please be assured that your responses will remain confidential and will be used strictly for academic research purposes.

Thank you very much for your time and support in advancing this critical research.

Warm regards,

Azeez Chollampat

DBA Program

Swiss School of Business and Management (SSBM)

Email: achollampat@ gmail.com

APPENDIX B :
SURVEY QUESTIONNAIRE

Survey form for AI-based VISA Processing System

1. Name :
2. Age :
3. Gender :
4. Occupation :

    Business Owner

    Working Professional

    Freelancer

    Government Employee

    Retired

    Student

    Not Working

    Other

5. How often do you travel internationally?

    Frequently (6 or more times a)

    Occasionally (3–6 times a year)

    Rarely (Once or twice a year)

    Never (I haven’t travelled internationally)

6. How satisfied were you with the accessibility and convenience of the Visa Application Centre's location?

Very Satisfied

Satisfied

Neutral

Dissatisfied

Very Dissatisfied

Not Applicable (Have not applied for a visa)

7. Were you able to easily track the status of your Visa Application?

    Yes

    No

    Neutral

    Have not tracked a Visa Application

8. How transparent is the current Visa Processing System?

    Transparent

    Neutral

    Not transparent

    I don't know

9. What were the issues you encountered with the paperwork submission process at the Visa Application Centre?

    Lengthy process

    Unclear instructions provided

    Too much paperwork

Process was too complicated

Waiting time was too long

Staff was not helpful

Document verification

No, the process was smooth

Have not submitted paperwork for a visa application

10. How easy was it to get and fill out the required paper forms for your Visa Application?

Very easy

Easy

Neutral

Difficult

Very difficult

Not Applicable (Have not obtained or filled out visa application forms)

11. How would you rate the overall efficiency of the Visa Application Process?

Very good

Good

Neutral

Bad

Very Bad

Not Applicable (Have not gone through the Visa Application Process)

12. Would you like to see in action a fully Automated Visa Processing System using Artificial Intelligence (AI)?

    Yes

    No

13. What improvements do you think could make the Visa Application Process quicker and easier for applicants in this era? _____

Appendix C:
Code

```
# Step 1: Install Required Libraries

!pip install gradio passporteye pytesseract

!apt install tesseract-ocr -y


# Step 2: Import all necessary libraries

import gradio as gr                    # For building the user interface

from passporteye import read_mrz       # For extracting data from the MRZ zone

import pytesseract                     # For performing OCR

from PIL import Image                  # For image input/output

import cv2                             # OpenCV for image processing and face detection

import numpy as np                     # Numerical operations for image arrays

import re                              # Regular expressions for pattern matching

import tempfile                        # Temporary file storage for processing

import json                            # Exporting data as JSON

import pandas as pd                    # Exporting data as CSV


# OCR function with confidence filtering

def perform_ocr_with_confidence(image_pil):

    # Convert PIL to NumPy array and grayscale

    image_np = np.array(image_pil)

    gray = cv2.cvtColor(image_np, cv2.COLOR_BGR2GRAY)
```

```
    # Apply threshold to enhance text visibility
    _, thresh = cv2.threshold(gray, 150, 255, cv2.THRESH_BINARY)
    processed_image = Image.fromarray(thresh)

    # Tesseract config for OCR
    config = '--psm 6 --oem 3'
    full_text = pytesseract.image_to_string(processed_image, config=config)

    # Extract words with confidence values
    ocr_data = pytesseract.image_to_data(processed_image,
output_type=pytesseract.Output.DICT, config=config)

    high_conf_words = [
        (ocr_data['text'][i], int(ocr_data['conf'][i]))
        for i in range(len(ocr_data['text']))
        if ocr_data['text'][i].strip() and int(ocr_data['conf'][i]) > 60
    ]

    return full_text, high_conf_words

# Main processing function
```

```
def process_passport(image_pil):
    summary = []
    clean_data = {}      # MRZ extracted data
    visual_data = {}     # Data extracted from OCR text (non-MRZ)

    # Save image temporarily for passporteye
    try:
        with tempfile.NamedTemporaryFile(suffix=".jpg", delete=False) as temp:
            image_pil.save(temp.name)
            temp_path = temp.name
    except Exception as e:
        return None, f"Error saving image: {e}", "", ""

    # Step 2: MRZ extraction
    try:
        mrz = read_mrz(temp_path)
        if mrz:
            data = mrz.to_dict()

            # Format date of birth
            dob_raw = data['date_of_birth']
            dob_formatted = (
```

```
            f"{'19' if int(dob_raw[:2]) > 30 else
'20'}{dob_raw[:2]}-{dob_raw[2:4]}-{dob_raw[4:]}"
            if dob_raw else "Invalid"
        )

        # Format expiration date
        exp_raw = data['expiration_date']
        expiry_formatted = (
            f"20{exp_raw[:2]}-{exp_raw[2:4]}-{exp_raw[4:]}"
            if exp_raw else "Invalid"
        )

        # Save MRZ details
        clean_data = {
            'surname': data['surname'],
            'given_names': data['names'],
            'passport_number': data['number'].replace('<', ''),
            'nationality': data['nationality'],
            'sex': data['sex'],
            'date_of_birth': dob_formatted,
            'expiry_date': expiry_formatted
        }
```

```
            summary.append("MRZ Extracted:")
            for k, v in clean_data.items():
                summary.append(f"{k.replace('_', ' ').title()}: {v}")
        else:
            summary.append("MRZ not detected.")
    except Exception as e:
        summary.append(f"MRZ extraction error: {e}")

    # Step 3: OCR-based field extraction
    try:
        text, high_conf_words = perform_ocr_with_confidence(image_pil)
        text_clean = re.sub(r'\s+', ' ', text)

        # Place of Birth
        match_birth = re.search(r'Place of Birth[:\-]?\s*([A-Z ,.\'-]{3,})', text_clean,
re.IGNORECASE)
        if match_birth:
            visual_data['place_of_birth'] = match_birth.group(1).strip().title()
            summary.append(f"Place of Birth: {visual_data['place_of_birth']}")
        else:
            summary.append("Place of Birth not confidently detected.")
```

```
    # Place of Issue
    match_issue = re.search(r'Place of Issue[:\-]?\s*([A-Z ,.\'-]{3,})', text_clean,
re.IGNORECASE)
    if match_issue:
        visual_data['place_of_issue'] = match_issue.group(1).strip().title()
        summary.append(f"Place of Issue: {visual_data['place_of_issue']}")
    else:
        summary.append("Place of Issue not confidently detected.")

    # Date of Issue
    all_dates = re.findall(r'\d{2}/\d{2}/\d{4}', text_clean)
    mrz_dates = [clean_data.get('date_of_birth'), clean_data.get('expiry_date')]
    visual_data['date_of_issue'] = None
    for d in all_dates:
        if d not in mrz_dates:
            visual_data['date_of_issue'] = d
            summary.append(f"Date of Issue: {d}")
            break
    if not visual_data['date_of_issue']:
        summary.append("Date of Issue not confidently detected.")
```

```
    except Exception as e:
        summary.append(f"OCR field extraction error: {e}")

    # Step 4: Face Detection using OpenCV
    face_image = None
    try:
        image_cv = cv2.cvtColor(np.array(image_pil), cv2.COLOR_RGB2BGR)
        gray = cv2.cvtColor(image_cv, cv2.COLOR_BGR2GRAY)
        face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades +
'haarcascade_frontalface_default.xml')
        faces = face_cascade.detectMultiScale(gray, scaleFactor=1.1, minNeighbors=4)

        if len(faces) > 0:
            (x, y, w, h) = faces[0]
            face_crop = image_cv[y:y+h, x:x+w]
            face_image = Image.fromarray(cv2.cvtColor(face_crop,
cv2.COLOR_BGR2RGB))
            summary.append("Face extracted from passport image.")
        else:
            summary.append("No face detected.")
    except Exception as e:
        summary.append(f"Face extraction error: {e}")
```

```
    # Step 5: Combine and export
    final_data = {**clean_data, **visual_data}
    summary.append("\nFinal Passport Data:")
    for k, v in final_data.items():
        summary.append(f"{k.replace('_', ' ').title()}: {v if v else 'Not Detected'}")

    json_export = json.dumps(final_data, indent=2)
    csv_export = pd.DataFrame([final_data]).to_csv(index=False)

    return face_image, "\n".join(summary), json_export, csv_export

# Web interface using Gradio
gr.Interface(
    fn=process_passport,
    inputs=gr.Image(type="pil", label="Upload Passport Image"),
    outputs=[
        gr.Image(type="pil", label="Cropped Passport Photo"),
        gr.Textbox(label="Extracted Passport Summary"),
        gr.Textbox(label="JSON Export"),
        gr.Textbox(label="CSV Export")
    ],
```

```
    title="AI Visa Passport Processor",
    description="Upload a scanned passport to automatically extract name, passport number, dates, place of birth, place of issue, and face image. Download as JSON or CSV."
).launch(share=True, debug=True)
```