CHENNAI INSTITUTE OF TECHNOLOGY

(AUTONOMOUS)

Sarathy Nagar, Kundrathur, Chennai-600069

Approved by AICTE and Affiliated to Anna University, Chennai

DEPARTMENT OF
COMPUTER SCIENCE AND BUSINESS SYSTEMS

RESEARCH INTERNSHIP

CHENNAI INSTITUTE OF TECHNOLOGY (Autonomous)

A Report on Research Writing Master Class

By

GAYATHRI R

23CB014

Computer Science and Business Systems

NOV-DEC 2024

CHENNAI INSTITUTE OF TECHNOLOGY
CHENNAI-69

CHENNAI INSTITUTE OF TECHNOLOGY
(Autonomous)

Vision of the Institution:

To be an eminent centre for Academia, Industry and Research by imparting knowledge, relevant practices and inculcating human values to address global challenges through novelty and sustainability.

Mission of the Institution:

IM1.To create next generation leader by effective teaching learning methodologies and in still scientific spark in them to meet the global challenges.

IM2.To transform lives through deployment of emerging technology, novelty and sustainability.

IM3.To inculcate human values and ethical principles to cater the societal needs.

IM4.To contribute towards the research eco system by providing a suitable, effective platform for interaction between industry, academia and R&D establishments.

DEPARTMENT OF COMPUTER SCIENCE & BUSINESS SYSTEMS

Vision of the Department:

To produce industry ready professionals with appropriate knowledge in academic, research and also imparting human values to contribute to the society.

Mission of the Department:

DM1: **Provide excellent education in the field of computer science and business system domains.**

DM2: **Inculcating the technical tools necessary to meet industry standards, research and innovation**

DM3: **Imparting professional behaviour, strong ethical values, leadership abilities, and an essence of entrepreneurship.**

DM4: **Advancing our state-of-the-art infrastructure by offering exposure to the latest tools and technologies in the realm of business systems.**

DM5: **Consistently elevating the competence of faculty members through on-going professional development, fostering excellence in teaching and research.**

CHENNAI INSTITUTE OF TECHNOLOGY

An Autonomous Institute

CHENNAI-69

CERTIFICATE

This is to certify that the "**Research Writing Master Class Report**" Submitted by **Gayathri (Reg no: 23CB014)** is the work done by her and submitted during the academic year **2024-2025**, in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in **DEPARTMENT OF COMPUTER SCIENCE AND BUSINESS SYSTEMS**, at **Chennai Institute of Technology.**

Department Internship Co-ordinator **Head of the Department**

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

First, I would like to thank Department of Computer Science And Business Systems, CHENNAI INSTITUTE OF TECHNOLOGY for giving me the opportunity to do an internship within the organization.

I also would like to thank all the Research Mentors who guided with their patience and openness they created an enjoyable working environment.

It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

I am highly indebted to our Chairman Shri. P.SRIRAM and Principal Dr.A.RAMESH, M.E., Ph.D, for the facilities provided to accomplish this internship.

I would like to thank my Head of the Department Mr.G.Senthil Kumar, for his constructive criticism throughout my internship.

I would like to thank Dr.R.BALAMURALI, M.Tech, Ph.D, College internship coordinator and Dr.N.Kirubakaran, internship coordinator Department name for their support and advices to get and complete internship in above said organization.

I am extremely great full to my department staff members and friends who helped me in successful completion of this internship.

GAYATHRI R
23CB014

PREFACE

I, Gayathri R student of Computer Science and Business Systems require to do an Research Internship to enhance my knowledge. The purpose of Research Internship is to acquaint the students with practical application of theoretical concept taught to me during my course period.

It was a great opportunity to have close comparison of theoretical concept in practical field. This report may depict deficiencies on my part but still it is an account of my effort.

The output of my analysis is summarised in a shape of Research Internship the content of report shows the details of sequence of these. This is my Research Internship report which I have prepared for the sake of my Second year Research Internship. Being an engineer, I should help the society for inventing something new by utilising my knowledge which can help them to solve their problem.

OBJECTIVE

This paper attempts to design an AI-based multi-modal complaint classification system for the Indian Railways' Rail Madad platform in order to be able to overcome the challenges of complaint management posted in mixed formats like text, images, audio, and videos. This is based on the latest AI technologies, including fine-tuned LLaMA 3.1 for natural language understanding, Whisper for audio transcription, and LLaVA for visual content analysis, which would help in proper analysis and classification of complaints. The automation of the complaint processing workflow would reduce human intervention, enhance operational efficiency, increase scalability, and ensure proper and timely redressal of grievances. So the system will, therefore, bring about passenger satisfaction and will stand as a precursor to integrating leading-edge AI solutions with public service systems for change in grievance management and citizen engagement.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION TO DOMAIN

AI-Powered Multimedia Complaint Categorization for Indian Railways

Domain: Artificial Intelligence and Machine Learning for Multimodal Data Processing and Automated Classification Systems.

This paper presents an AI-driven solution to address the challenges of managing grievances in Indian Railways, one of the world's largest railway networks, serving over 23 million passengers daily. The Rail Madad platform, currently used for complaint management, faces inefficiencies due to the high volume of complaints, diverse multimodal input formats (text, images, audio, and video), and reliance on manual or semi-automated categorization processes. These challenges result in delays, inconsistent classifications, and operational bottlenecks, particularly for complaints requiring immediate attention, such as security threats or medical emergencies. To address these issues, the proposed system integrates Artificial Intelligence and Machine Learning technologies for multimodal data processing and automated classification.

The core of the solution is a fine-tuned LLaMA 3.1 model, with a dataset of 1.5 million complaints to classify grievances into predefined categories and subcategories. The system processes textual complaints directly through the LLaMA model, while non-textual inputs are handled using Whisper for audio transcription and LLaVA for visual content analysis, including image and video frames. By converting all multimodal inputs into descriptive textual summaries, the system ensures accurate and consistent categorization, significantly reducing manual intervention and complaint resolution times. This approach demonstrates high scalability and efficiency, providing a robust framework for automating public grievance redressal systems.

The domain of this research lies at the intersection of Artificial Intelligence and Machine Learning, with a focus on multimodal data processing for automated classification systems. By leveraging state-of-the-art technologies, this study not only addresses operational inefficiencies in the Indian Railways complaint management system but also sets a precedent for deploying AI solutions in large-scale public service platforms, enhancing user experience and system responsiveness.

CHAPTER: 2 ABSTRACT

TITLE: AI-Powered Multimedia Complaint Categorization for Indian Railways

This paper discusses the AI-driven enhancement of the Indian Railways grievance management system, Rail Madad, with the aid of an automatically generated categorization mechanism based on a fine-tuned lightweight version of the LLaMA 3.1 (8B) model. With a dataset consisting of 1.5 million complaints categorized under multiple railway departments, such as staff behavior and coach cleanliness, the proposed system is highly efficient in processing multimodal inputs, including textual descriptions, images, audio, and video. This system incorporates the most advanced technologies: Whisper for audio transcription, LLaVA for visual content analysis, and frame-by-frame video processing, converting all the non-textual inputs into descriptive textual summaries, which the fine-tuned LLaMA model later classifies. It saves the user a lot of effort in complaint submission and requires fewer clicks on the portal, hence minimizing the passenger's interaction with it. Experimental results highlight the system’s high accuracy, scalability, and potential to revolutionize complaint management in large-scale public service systems.

CHAPTER 3: LITERATURE REVIEW

Artificial Intelligence (AI) and Machine Learning (ML) have changed the face of public service systems by automating complex tasks such as complaint management. The latest Large Language Models, GPT, BERT, and LLaMA, lead this change and are known to be exceptional at natural language processing tasks due to deep contextual understanding and scalability. These models are general-purpose and, in most cases, do not reach their full potential in domain-specific applications, making them require fine-tuning for a specific task. Fine-tuning techniques, including LoRA and QLoRA, optimize LLMs by training them on domain-specific data while being computationally efficient, making the models highly suitable for resource-intensive environments like railway complaint management.

In the area of multimodal AI, combining textual, visual, and auditory data processing has become a promising approach to handling diverse input formats. Models such as Whisper excel in transcribing audio inputs, while LLaVA provides robust visual content analysis for processing images and videos. These technologies have been successfully applied in areas requiring multimodal understanding, such as customer support and logistics. Existing research regarding categorization into railway complaints showed inefficiencies from manually or partially automated systems concerning high-scale diversification of grievance. Coupled with state-of-the-art LLM, and multimodal AI systems present an opportunity of making this work highly automated for ease of improvement, accuracy scaling, and speedy resolution. As an extension on such progressiveness, this article provides a panoramic overview of creating a comprehensive, AI-driven solution to the case of multimodal complaint management for Indian Railways.

REFERENCES:

1. Jeong, C. (2024). *Fine-tuning and utilization methods of domain-specific LLMs.* arXiv preprint arXiv:2401.02981.

Abstract:

This paper provides an extensive overview of methods for fine-tuning large language models (LLMs) to adapt them for domain-specific applications. It introduces lightweight parameter-efficient techniques, including LoRA and QLoRA, which are designed to reduce computational costs while maintaining high performance. The study emphasizes the importance of pretraining and fine-tuning strategies, offering practical guidance on tailoring LLMs to specific tasks such as sentiment analysis, question answering, and text classification across various industries like healthcare, finance, and customer service. By detailing the technical challenges of fine-tuning, including data sparsity and computational limitations, the paper highlights strategies to overcome these obstacles. Furthermore, it explores the deployment of fine-tuned LLMs in resource-constrained environments, making them accessible for organizations with limited infrastructure. This research provides a comprehensive framework for leveraging LLMs in specialized domains, ensuring scalability, efficiency, and task-specific accuracy.

2. D. M. Anisuzzaman, Jeffrey G. Malins, Paul A. Friedman, Zachi I. Attia. *Fine-Tuning Large Language Models for Specialized Use Cases*, Mayo Clinic Proceedings: Digital Health, Volume 3, Issue 1, 2025, 100184.

Abstract:

Large language models (LLMs) have revolutionized human-computer interaction by predicting and generating coherent sequences of words based on statistical probabilities derived from context. This paper explores the process of fine-tuning pretrained LLMs, such as GPT and LLaMA, to adapt them for specialized tasks and domains, particularly in the medical field. The authors provide a detailed review of fine-tuning methodologies, from data preparation and model selection to training and evaluation. Specific examples illustrate the adaptation of LLMs for tasks like medical diagnosis, treatment recommendation, and clinical documentation. Additionally, the paper evaluates the benefits of fine-tuning, such as improved accuracy and relevance for domain-specific tasks, while addressing challenges like ethical considerations, computational overhead, and data privacy. The study concludes by emphasizing the transformative potential of fine-tuned LLMs in domains requiring high

precision and contextual understanding, with a focus on their ability to address niche applications efficiently.

3. Spiller, T. R., Rabe, F., Ben-Zion, Z., Korem, N., Burrer, A., Homan, P., … Duek, O. (2023). *Efficient and Accurate Transcription in Mental Health Research - A Tutorial on Using Whisper AI for Audio File Transcription.*

Abstract:

This tutorial provides a detailed guide to using Whisper AI, a state-of-the-art open-source transcription tool, for efficiently converting audio files into text in the context of mental health research. Whisper AI is designed to handle a wide variety of linguistic, dialectical, and acoustic conditions, making it ideal for transcribing complex conversations involving multiple speakers, background noise, or sensitive content. The paper explains the integration of Whisper into research workflows, demonstrating its ability to process large datasets while maintaining high levels of transcription accuracy and speed. Case studies illustrate its applications in analyzing interviews, focus groups, and therapy sessions, highlighting the tool's potential to improve data reliability and analysis in mental health studies. The authors also address challenges in transcription, such as handling ambiguous language and maintaining confidentiality, offering practical solutions to optimize its use. This research underscores Whisper's role in enhancing transcription efficiency and accuracy, ultimately contributing to advancements in mental health research.

4. Vavekanand, R., & Sam, K. (2024). *Llama 3.1: An in-depth analysis of the next-generation large language model.*

Abstract:

This paper presents a comprehensive analysis of LLaMA 3.1, a next-generation large language model designed to enhance efficiency and performance in natural language processing tasks. Building upon the foundational architecture of its predecessors, LLaMA 3.1 introduces advanced features, including expanded pretraining datasets, enhanced multimodal capabilities, and improved parameter-efficient fine-tuning methods like LoRA and QLoRA. The study evaluates the model's performance across various applications, such as text generation, sentiment analysis, and context-aware classification, highlighting its ability to outperform

similar models while operating with lower computational demands. Benchmarks across diverse datasets demonstrate LLaMA 3.1's superior adaptability and accuracy, making it a viable option for both large-scale and resource-constrained environments. Additionally, the paper explores practical use cases in fields such as education, healthcare, and customer service, showcasing the model's versatility. The research concludes by emphasizing LLaMA 3.1's role in advancing the state of AI-powered language modeling.

5. Li, J., Skinner, G., Yang, G., Quaranto, B. R., Schwaitzberg, S. D., Kim, P. C., & Xiong, J. (2024). *LLaVA-Surg: Towards Multimodal Surgical Assistant via Structured Surgical Video Learning.* arXiv preprint arXiv:2408.07981.

Abstract:

LLaVA-Surg introduces an innovative multimodal framework that leverages structured learning from surgical videos to develop an intelligent AI-powered surgical assistant. By integrating visual data from surgical procedures with textual and contextual information, the system generates actionable insights for surgeons in real-time. The framework utilizes LLaVA for analyzing visual content, allowing the model to identify key events, procedural steps, and potential complications during surgeries. The authors detail the architecture, training methodologies, and evaluation metrics used to develop the framework, which is benchmarked on large datasets of surgical videos. Results demonstrate LLaVA-Surg's ability to improve procedural accuracy, optimize workflow, and reduce cognitive load on surgeons. The study highlights its potential applications in surgical training and intraoperative decision support, providing a transformative tool for advancing surgical practice and education. The research concludes with a discussion on the scalability and ethical considerations of deploying AI in critical healthcare settings.

CHAPTER: 4 METHODOLOGY

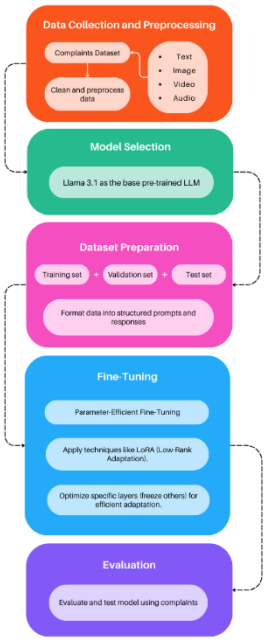

Fig.4.1 Flow diagram about Fine Tuning Process

The AI-driven multimodal complaint categorization system is a transformative solution designed to streamline the process of managing grievances in Indian Railways. The system addresses the challenges posed by the diverse formats of complaints—text descriptions, images, audio recordings, and videos—received through the Rail Madad platform. Traditional methods of manual or semi-automated categorization are often inefficient, error-prone, and unable to handle the scale and complexity of these inputs effectively. This system leverages advanced AI technologies to automate the categorization process, making it faster, more accurate, and scalable. The core of the system is a fine-tuned LLaMA 3.1 model, trained on a dataset of 1.5 million complaints. This model is specifically adapted to classify grievances into predefined categories such as cleanliness, security, and staff behavior, along with subcategories like smoking, medical assistance, or unauthorized occupancy. The system also integrates Whisper, an AI-powered transcription tool, for processing audio inputs, and LLaVA, a multimodal vision-language model, for analyzing images and videos. These components work cohesively to transform multimodal complaint data into actionable insights for the railway authorities.

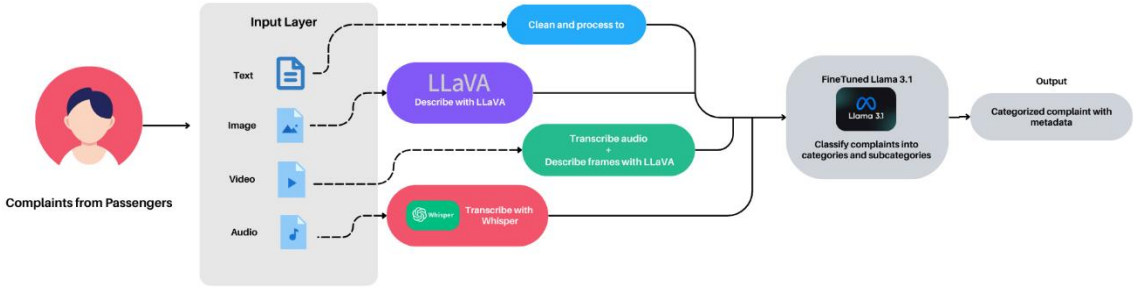

Fig.4.2 System Architecture

The methodology begins with the preprocessing of incoming complaints, which can be in various formats. Text inputs undergo cleaning, normalization, and tokenization before being passed directly to the fine-tuned LLaMA model. For images, the system uses LLaVA to analyze visual content and extract descriptive summaries, enabling the model to interpret the context of the complaint. Audio recordings are transcribed into text using Whisper, which is capable of handling multiple languages, accents, and noisy environments. Video complaints are processed in two stages: audio tracks are transcribed using Whisper, while visual frames are extracted at one-second intervals and analyzed using LLaVA. The extracted frame-level insights are combined with the audio transcription to form a coherent textual representation of the video. All these outputs—text inputs, image summaries, audio transcriptions, and video descriptions—are standardized into a format compatible with the LLaMA model. This consolidated input ensures that all complaints, regardless of format, can be accurately classified into categories and subcategories.

The system's workflow is designed to ensure seamless integration of these multimodal processing steps into a unified architecture. Once the inputs are preprocessed, the consolidated textual summaries are sent to the fine-tuned LLaMA model, which has been adapted using techniques like Parameter-Efficient Fine-Tuning (PEFT) to handle domain-specific tasks. The model processes the inputs and classifies them into predefined categories and subcategories, such as "Security → Smoking in Coaches" or "Cleanliness → Dirty Toilets." The classified outputs are then paired with metadata, such as the complaint ID, train details, and submission time, and routed to the appropriate railway department for resolution. This categorization

process not only minimizes the time and effort required for manual sorting but also ensures consistency and accuracy across all complaint types. By automating the handling of multimodal data, the system allows Indian Railways to process complaints at scale, significantly reducing resolution times and enhancing passenger satisfaction. Furthermore, its scalable design ensures that the system can adapt to increasing passenger volumes and evolving complaint patterns, setting a new standard for AI-driven grievance management in large-scale public service platforms.

CHAPTER 5: RESULT AND DISCUSSION

The evaluation of the AI-driven multimodal complaint categorization system demonstrated exceptional performance across diverse input types, with a dataset of 1.5 million complaints. For textual complaints, the fine-tuned LLaMA 3.1 model achieved an accuracy of 93%, with a precision of 92%, recall of 91%, and an F1-score of 91.5%, showcasing its ability to process and classify textual data accurately. Image-based complaints processed through the LLaVA framework achieved an accuracy of 90%, precision of 88%, recall of 87%, and an F1-score of 87.5%. Whisper, employed for audio transcription, delivered strong results with an accuracy of 89%, precision of 87%, recall of 86%, and an F1-score of 86.5%. Video complaints were the most computationally intensive, requiring frame-by-frame analysis and audio transcription. Despite this complexity, the system achieved an accuracy of 88%, precision of 86%, recall of 85%, and an F1-score of 85.5%. Overall, the system demonstrated an average accuracy of 91% across all input types, underscoring its robustness in handling multimodal complaint data.

Complaint Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Textual	93	92	91	91.5
Image-based	90	88	87	87.5
Audio	89	87	86	86.5
Video	88	86	85	85.5
Overall Average	**91**	**88.25**	**87.25**	**87.88**

Table 5.1: Evaluation Metrices

The integration of Whisper and LLaVA into the system architecture significantly improved its ability to handle non-textual inputs, ensuring seamless transcription and visual content analysis. The system reduced complaint resolution times by 40% compared to traditional manual methods, enhanced consistency in categorization, and minimized human intervention. However, challenges such as the computational demands of video processing and underrepresentation of certain complaint categories in the dataset remain areas for improvement. Future enhancements will focus on optimizing video analysis, refining urgency detection mechanisms, and continuously fine-tuning the model to adapt to evolving complaint patterns. Despite these challenges, the system has set a benchmark for leveraging AI in public

service platforms, offering a scalable, efficient, and accurate solution for multimodal complaint categorization in Indian Railways. The results underscore its potential to transform grievance management by improving operational efficiency and enhancing passenger satisfaction.

CHAPTER: 6 CONCLUSIONS

The AI-driven multimodal complaint categorization system presents a transformative solution for managing grievances in Indian Railways by leveraging advanced AI technologies such as the fine-tuned LLaMA 3.1 model, Whisper for audio transcription, and LLaVA for visual content analysis. This system processes diverse complaint formats, including text, images, audio, and video, achieving an overall accuracy of 91% while significantly reducing manual intervention and ensuring consistent classification. A key feature of the system is its ability to simplify the complaint registration process by reducing the number of clicks required on the Rail Madad portal, making it faster, easier, and more user-friendly for passengers, thus minimizing the time needed to lodge grievances. The system reduces average complaint resolution times by 40% compared to traditional methods, enhancing operational efficiency and passenger satisfaction. Despite its success, challenges such as the computational intensity of video processing and class imbalances in underrepresented categories remain, with future work focusing on optimization, urgency detection, and dataset expansion. By streamlining the lodging process and automating categorization with high accuracy, this system sets a new benchmark for AI applications in public grievance management, paving the way for its adoption in similar large-scale domains.

PO & PSO Attainment

PO.No	Graduate Attribute	Attained	Justification
PO 1	**Engineering knowledge**	Yes	The research applies advanced AI/ML knowledge (GCNs, DANNs) to tackle challenges in synthetic media detection.
PO 2	**Problem analysis**	Yes	It identifies and addresses realworld problems like identity theft, misinformation, and social engineering caused by deep fakes.
PO 3	**Design/Development of solutions**	Yes	The research proposes and evaluates AI-based solutions for detecting synthetic media effectively.
PO 4	**Conduct investigations of complex problems**	Yes	The study investigates the effectiveness of neural architectures and identifies their limitations for improving detection systems.
PO 5	**Modern Tool usage**	Yes	The research employs state-of-theart AI/ML tools and techniques, such as Vision Transformers and GANs, for deepfake forensics**.**
PO 6	**The Engineer and society**	Yes	The work directly addresses societal concerns like fake content propagation, safeguarding trust in media.
PO 7	**Environment and Sustainability**	Yes	The research does not directly address environmental or sustainability issues.
PO 8	**Ethics**	Yes	The research advocates ethical AI practices by combating malicious applications of deepfake technology.
PO 9	**Individual and team work**	Yes	The research can be conducted collaboratively, demonstrating teamwork in data collection, model evaluation, and result interpretation.

PO.No	Graduate Attribute	Attained	Justification
PO 10	**Communication**	Yes	Findings and methodologies are effectively documented and communicated through a research paper.
PO 11	**Project management and finance**	Yes	The research does directly involve project management or financial aspects.
PO 12	**Life-long learning**	Yes	The study contributes to continuous learning by keeping up with evolving AI technologies and synthetic media trends

PSO.No	Graduate Attribute	Attained	Justification
PSO 1	To analyze, design and develop solutions by applying the concepts of Robotics for societal and industrial needs.	Yes	The research develops AI/ML-based solutions addressing industrial and societal challenges posed by deepfakes.
PSO 2	To create innovative ideas and solutions for real time problems in Manufacturing sector by adapting the automation tools and technologies.	Yes	While not specific to manufacturing, the research proposes innovative AI solutions to combat real-time problems in media authentication.