KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments	
Healthcare	IMAS	IMAS: A Comprehensive Agentic Approach to Rural Healthcare Delivery	IMAS proposes an agentic AI system designed to support semi-trained rural healthcare providers—such as Community Health Workers and rural practitioners—by offering context-aware, triaged, and appropriate medical guidance in local languages.The system's main objective is to improve healthcare access in underserved rural areas where formal practitioners are scarce. built as a pipeline of five AI agents: Translation Agent: Converts patient queries in local spoken languages (e.g., Telugu, Hindi) into the LLM's processing language using a fine-tuned Seamless-M4T model Medical Complexity Assessment Agent: Categorizes cases into low, medium, or high complexity, triggering escalation to specialists for high-risk cases Expert Network Integration Agent: Aggregates reasoning from multiple LLMs and medical knowledge bases (e.g., PubMedQA, MedQA) to collaboratively diagnose medium-complexity cases Final Advice Generation Agent: Leverages models like Llama-3 70B or GPT-4, fine-tuned on medical benchmarks, to produce accurate clinical recommendations Response Simplification Agent: Converts technical medical advice into clear, culturally sensitive, actionable guidance, while applying guardrails to avoid misinformation and ensure privacy It demonstrates clinical effectiveness on medical QA datasets like MedQA, PubMedQA, and JAMA	Agentic AI,LLM optionally MERN	Badrinath	https://drive.google.com/file/d/1qS5eIXtmdEeuVIpjxFfOU8RrWEaQ0OF/view?usp=drive_link	https://github.com/uheal/IMAS			
AI	AD-AGENT	AD-AGENT: A Multi-agent Framework for End-to-end Anomaly Detection	AD-AGENT simplifies anomaly detection (AD) for non-expert users by translating natural language into executable AD pipelines. This is achieved through an LLM-driven multi-agent framework coordinating specialized agents for: Intent Parsing: Understanding user requests. Data Preparation: Readying data for analysis. Library & Model Selection: Choosing appropriate AD tools. Documentation Mining: Extracting relevant information. Iterative Code Generation & Debugging: Building and refining scripts. The agents utilize a shared short-term workspace and a long-term cache to integrate popular AD libraries such as PyOD, PyGOD, and TSLib into a unified workflow. This architecture enables the system to produce reliable scripts and recommend competitive models across diverse AD challenges.	AI Agents,LLM	Badrinath	https://drive.google.com/file/d/1uvMKEIh2E49LjSF9g_WriX6Mhim_atsr/view?usp=drive_link	https://github.com/USC-FORTIS/AD-AGENT?tab=readme-ov-file			
AI	AgentMonitor	AgentMonitor: A Plug-and-Play Framework for Predictive and Secure Multi-Agent	The rapid rise of LLM-based multi-agent systems (MAS) highlights the challenge of pre-configuring them effectively and ensuring trustworthy responses. Unlike single LLM scaling laws, MAS performance is only known post-execution, and malicious agents pose significant security risks by spreading harmful content. AgentMonitor, a framework designed to integrate at the agent level within existing MAS. AgentMonitor captures step-by-step inputs and outputs, enabling two key functionalities: Performance Prediction: Transforming captured data into statistics to train a regression model (e.g., XGBoost) to predict downstream task performance. Experiments show high in-domain correlation and moderate out-of-domain correlation. On-the-Fly Corrections: Applying immediate interventions to mitigate negative impacts, reducing safety risks. This framework significantly improves the predictability and reliability of MAS.	Agentic AI,LLM	Badrinath	https://drive.google.com/file/d/1PzPRoPmYMl_iKCsEHqZLdQiZrV1mZPjP/view?usp=drive_link	https://github.com/chanchimin/AgentMonitor?tab=readme-ov-file			
AI	ALWUMS	Autonomous Legacy Web Application Upgrades Using a Multi-Agent System	Addressing the challenges of updating outdated web applications (security, reliability, cost), an LLM-based multi-agent system for autonomous legacy web application upgrades is proposed. The core architecture involves distributing the complex upgrade task across multiple specialized agents operating in distinct phases. This design allows the system to: Maintain context across diverse tasks and agents during the update process. Update all relevant files to their latest versions. Evaluations using Zero-Shot Learning (ZSL) and One-Shot Learning (OSL) prompts demonstrated the system's effectiveness. Compared to standalone LLMs, the multi-agent system delivered better solutions in some cases, significantly reducing errors in updated view files and successfully handling complex requirements. It provides a working foundation for future LLM-based code maintenance, showing high precision even with basic prompts.	Agentic AI,Web Technologies	Badrinath	https://drive.google.com/file/d/1YO30AymguHjWXB-lB9lQe7gXqISjulPqe/view?usp=drive_link	https://github.com/alasalm1/Multi-agent-pipeline			
AI	RLVO	Re-Aligning Language to Visual Objects with an Agentic Workflow	Language-based object detection (LOD) relies heavily on paired visual-language data, often scaled up using Vision-Language Models (VLMs) to generate human-like expressions. A critical issue, however, is VLM hallucinations, leading to inaccurate object descriptions and degraded vision-language (VL) alignment. To combat this, we propose Real-LOD, an agentic workflow controlled by an LLM designed to re-align language to visual objects. Real-LOD adaptively adjusts image and text prompts through a cyclic process involving: Planning: Automatic reasoning and action arrangement based on neural symbolic designs. Tool Use: Adjusting prompts and sending them to VLMs for object re-description. Reflection: An LLM analyzes refined expressions for feedback. This iterative refinement process significantly improves language descriptions. Real-LOD, despite using a smaller dataset , enabled a prevalent LOD model to outperform existing methods, demonstrating its potential to preserve data quality while scaling quantity.	AI,CV and VLM	Badrinath	https://drive.google.com/file/d/1Eq3f_sQXVlcLZ39fpFr-iRm9n3fPQaa8/view?usp=drive_link	https://arxiv.org/abs/2503.23508			
Defence Technol	MMLM-PDI	Multimodal Large Language Models for Phishing Webpage Detection and Identification	Traditional brand-based phishing detection, relying on costly and maintenance-heavy Computer Vision models, struggles with continuous data collection and managing reference lists. This project explores the efficacy of Large Language Models (LLMs), specifically multimodal LLMs, to overcome these limitations. The proposed system leverages LLMs' pre-trained understanding of webpage elements (logo, theme, favicon, etc.) in a two-phase architecture: Brand Identification: An LLM identifies the brand imitated by the webpage. Domain Verification: A second LLM compares the identified brand with the URL's domain to flag phishing. Evaluations on a new dataset demonstrate the LLM-based system's high detection rate at high precision, significant outperformance of state-of-the-art methods, and robustness against adversarial attacks. Crucially, it also provides interpretable evidence for its decisions, enhancing trust and understanding.	Multimodel LLMs	Badrinath	https://drive.google.com/file/d/1fsBPb-ndGqX1qsmjS5-cgMcTKpUgSiaO/view?usp=drive_link	https://github.com/jehleekr/multimodal_llm_phishing_detection			
AI	HuggingGPT	HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face	To tackle complex, multi-domain, and multi-modal AI tasks—a key step towards Artificial General Intelligence (AGI)—HuggingGPT proposes an innovative approach. It posits that Large Language Models (LLMs) can act as a central controller, leveraging language as a generic interface to orchestrate existing, specialized AI models. HuggingGPT's architecture integrates an LLM (e.g., ChatGPT) with a vast repository of AI models (e.g., Hugging Face) through a structured workflow: Task Planning: The LLM receives a user request and breaks it down into a sequence of subtasks. Model Selection: Based on function descriptions from Hugging Face, the LLM selects the most appropriate AI model for each subtask. Task Execution: The selected AI models execute their respective subtasks. Response Summarization: The LLM then synthesizes the results from all executed subtasks into a coherent response. This framework enables HuggingGPT to handle a wide range of sophisticated AI tasks across diverse modalities (language, vision, speech) and domains, demonstrating impressive results and paving a new path towards AGI.	Multimodel LLMs	Badrinath	https://drive.google.com/file/d/1s1uh4HzsX7oiD2aCpNF8m5V6BJCIbrlPk/view?usp=drive_link	https://github.com/microsoft/JARVIS/tree/main			

KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments
Defence Technol	MAMSCN	A Multi-Agent Monitoring System for Computer Networks	Traditional centralized monitoring systems face limitations in detecting faults and anomalies in dynamic computer networks due to their rigid structure, scalability issues, and reliance on predefined rules. A decentralized, intelligent architecture is needed—where distributed components monitor, analyze, and respond to network behavior autonomously and collaboratively.This project enhances the original multi-agent architecture by incorporating pre-trained AI models into specific agents. The goal is to transition from rule-based detection to intelligent, adaptive monitoring, allowing the system to recognize complex patterns, detect unseen anomalies, and evolve with changing network behavior. Agents and Their Enhanced Functions Capturing Agents (Non-AI) Role: Collect low-level system and network data (e.g., CPU usage, bandwidth, active connections). Tools: tshark, netstat, psutil, SNMP (optional). Enhancement: No AI integration; remains efficient and lightweight for real-time data capture. Analysis Agents (Enhanced with AI) Role: Detect anomalies or suspicious patterns in captured data. Enhancement: Use pre-trained AI models for anomaly detection and classification: Isolation Forest or Autoencoder models trained on normal network behavior Hugging Face tabular models or models from libraries like PyOD, scikit-learn, or torch Behavior: Can detect port scans, DDoS patterns, login anomalies using learned behavior instead of static rules. Communication Agents (Non-AI) Role: Coordinate and transmit events between agents across different nodes. Tools: Flask or FastAPI for REST APIs, or Kafka for scalable message streaming. Enhancement: Optionally add message prioritization AI (e.g., transformer-based classifiers to rank alerts) Decision Agents (Optional AI Enhancement) Role: Aggregate events from analysis agents and make final decisions. Enhancement: Use rule-based decision logic with optional Bayesian or rule-learning models to identify patterns over time and adjust severity scores.	Agentic AI, Network	Badrinath	https://drive.google.com/file/d/148a1XwCJmXR1m0igdeOyo9eaG5Az9Y7q/view?usp=drive_link	There is no signle source, however exploring the technologies based on the architecture it need to be implemeted. Once achieved it can be pushed to github public domain		
AI	Multi-Agent-VQA	Multi-Agent VQA: Exploring Multi-Agent Foundation Models in Zero-Shot Visual Question Answering	This project implements a modular, multi-agent system for Visual Question Answering (VQA) without any task-specific training. It integrates specialized agents such as object detectors, counters, and language models like GPT-4V or Gemini Pro Vision. The system dynamically decides which tools to activate based on the type of visual question, improving accuracy in complex scenarios like object counting or spatial reasoning. An orchestrator coordinates agent outputs and guides final answer generation through LLM-based voting. This approach demonstrates the power of agentic AI in real-world, zero-shot tasks. The project promotes modularity, explainability, and robustness across varied image-question pairs. It is ideal for exploring vision-language systems and agent-based AI design.	AI,CV and VQA	Badrinath	https://drive.google.com/file/d/1KvTLOg9L6G26JeVofrtGcUsOGjT99qrJ/view?usp=drive_link	https://github.com/bowen-upenn/Multi-Agent-VQA?utm_source=chatgpt.com		
AI	From-Simple-to-Professional	From Simple to Professional: A Combinatorial Controllable Image Captioning Agent	This project aims to generate rich, controllable image captions from simple user prompts using an agentic AI approach. Traditional captioning models lack customization, while CapAgent introduces two key agents: (1) the Instruction Evolving Agent, which refines vague user inputs using web search and image metadata, and (2) the Planner Agent, which retrieves relevant examples and generates tool-calling Python code. The planner coordinates specialized tools—VQA, object detector, sentiment controller, keyword extractor, and caption rewriter—through a thought → action → observation loop. The final output is a polished caption tailored to sentiment, emphasis, or domain needs. The system is modular, explainable, and open-source—suitable for students exploring Agentic AI, vision-language tasks, or intelligent content generation.	AI,CV and Image Caption	Badrinath	https://drive.google.com/file/d/1ZzG2Euk8TEuM7SbxCvH8t_kDMEJmlqKh/view?usp=drive_link	https://github.com/xin-ran-w/CapAgent		
Healthcare	**CareBot**	AI Agents for Conversational Patient Triage	A simulation-based evaluation using Electronic Health Record (EHR) data to assess conversational triage agent performance. Focus is on response quality, empathy, and accuracy of medical advice.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1-xp4Tp3ye8ArSiaKz7x1cXncPNHo8hwX/view?usp=drive_link			https://www.researchgate.net/publication/392406828_AI_Agents_for_Conversational_Patient_Triage_Preliminary_Simulation-Based_Evaluation_with_Real-World_EHR_Data
Supply Chain	**FusionCast**	Supply Chain Demand Forecasting via an Explainable Multi-Channel Data Fusion Network Model	MCDFN is a hybrid deep-learning model combining CNN, LSTM, and GRU channels working in parallel to capture both spatial patterns (via CNN) and temporal dependencies (via LSTM/GRU) in time-series data. The fused output yields forecasts, and explainability is added through ShapTime and permutation feature importance	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1eKHFeWqBPzeZt-GLuNqYfMZpJa9KozUB/view?usp=drive_link			https://arxiv.org/pdf/2405.15598
efence Technolog	**CyberShieldRL**	Deep Reinforcement Learning for Adaptive Cyber Defense in Network Security	The paper explores a Deep Reinforcement Learning (DRL) framework to defend networked systems against cyber threats like malware, intrusions, and phishing. It applies DRL techniques—DQN, PPO, and TD3—to learn how to detect and respond to attacks in simulated network environments.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1-hWxcJ-WgZ2R_UFI5IXoZ9xuflvJoH8oM5/view?usp=drive_link			https://www.researchgate.net/publication/381859037_Deep_Reinforcement_Learning_for_Adaptive_Cyber_Defense_in_Network_Security
Legal System	**LegalEase**	LawPal : A Retrieval Augmented Generation Based System for Enhanced Legal Accessibility in India	Builds a legal chatbot using vector-store retrieval (FAISS) over Indian legal texts (constitution, statutes), answering user queries with accurate, context-rich responses and clear justifications.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1GK62pDvkAhq8HI5-umCqaoA1cH3aHG3O/view?usp=drive_link	https://github.com/AaryanGole26/LawPal		https://arxiv.org/abs/2502.16573
FinTech	**FraudSynth**	Utilizing GANs for Fraud Detection: Model Training with Synthetic Transaction Data	Studies the use of GANs to generate realistic fraudulent transaction samples, improving the performance of anomaly detectors—especially useful in class-imbalanced scenarios	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1wtc4tNuSuQNuOUtfyDtCdfagn0u2oZNCW/view?usp=drive_link			https://arxiv.org/abs/2402.09830
efence Technolog	**AutoSentinel**	Reinforcement Learning for Autonomous Resilient Cyber Defence	This paper explores the use of Deep Reinforcement Learning (DRL) to create autonomous cyber defenders that can respond to cyberattacks in real time — faster and more flexibly than static, rule-based systems.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1kpAI_Gevq70YwrQvIYCI7ZfSPNrASWU7/view?usp=drive_link			https://www.fnc.co.uk/media/mwcnckk/us-24-milesfarmer-reinforcementlearningforautonomousresilientcyberdefence-wp.pdf
Legal System	**LawMate AI**	Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model	Provides an explainable method to match legal cases using inverse optimal transport, extracting justification “rationales” between sentence pairs. Useful for research on transparent legal retrieval.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1YGRodaQd10cpY-JlVCuhFecxsHRh52Ou/view?usp=drive_link			https://arxiv.org/abs/2306.16092?
efence Technolog	**SeaGuard MARL**	Multi-Agent Reinforcement Learning for Maritime Operational Technology Cyber Security	Operational Technology (OT) systems aboard maritime vessels (like ship control systems) are increasingly targeted by cyberattacks but lack robust defense mechanisms. These systems differ from regular IT—they’re brittle, legacy-driven, and cannot simply adopt IT-based defenses. The paper explores using Multi-Agent Reinforcement Learning (MARL) to automate cyber defense for OT environments using a novel simulated environment called IPMSRL (Integrated Platform Management System for RL)	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1AlI23bbLSqTzLpgwiRuxnQBORPhHgLJo/view?usp=drive_link			https://arxiv.org/abs/2401.10149
AI Agriculture	**AgroVision**	Towards more efficient agricultural practices via transformer-based crop type classification	Machine learning has great potential to increase crop production and resilience to climate change. Accurate maps of where crops are grown are a key input to a number of downstream policy and research applications. In this proposal, we present preliminary work showing that it is possible to accurately classify crops from time series derived from Sentinel 1 and 2 satellite imagery in Mexico using a pixel-based binary crop/non-crop time series transformer model. We also find preliminary evidence that meta-learning approaches supplemented with data from similar agro-ecological zones may improve model performance. Due to these promising results, we propose further development of this method with the goal of accurate multi-class crop classification in Jalisco, Mexico via meta-learning with a dataset comprising similar agro-ecological zones.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1fKASfAb6sA3gn4XeuiOSi_Aq83hNGKM/view?usp=drive_link			https://arxiv.org/abs/2411.02627
efence Technolog	**SkyTactix**	Explaining Strategic Decisions in Multi-Agent Reinforcement Learning for Aerial Combat Tactics	This paper implements explainability in MARL—specifically for air-combat simulation. After agents are trained on tactical scenarios, explainability techniques highlight which factors influenced each strategic decision, bridging black-box models and human interpretability.	MERN with AI	Nidhi Srivastav	https://drive.google.com/file/d/1zC-zIPf48g_I-MmOHfgy9EJwp8Sxhke/view?usp=drive_link			https://arxiv.org/abs/2505.11311
Supply Chain	**Speak2Rover**	**Speak2Rover**: LLM-Powered Task Execution	**Problem statement** Design and implement an end-to-end embodied agent system that integrates the LLM-Planner approach with the SDK of a rover (simulated or real). The system must enable the rover to understand and execute natural language instructions by planning and performing complex navigation and manipulation tasks in a visually-perceived environment. **Example Task** A user instructs: “Go to the kitchen, pick up the red cup from the counter, and bring it to the dining table.” The system processes the instruction, plans the sequence of actions using the LLM (e.g., navigate to kitchen → locate counter → identify red cup → pick up → navigate to dining table → place cup), grounds the plan based on the current visual input (e.g., detects that the cup is on a different table and updates the plan), and executes the actions by controlling the rover through its SDK. **Learning Outcomes** By the end of this project, students will be able to: 1. Explain the role and challenges of LLM-based planning in embodied agents and robotics. 2. Integrate natural language processing, computer vision, and robotic control in a unified system. 3. Work with a rover SDK (e.g., ROS-based, TurtleBot, or any simulated rover) to programmatically control movement and actions. 4. Implement LLM-based few-shot planners with grounding mechanisms that adapt to live environmental data. 5. Design and execute experiments to evaluate task performance and sample efficiency in embodied tasks. 6. Develop skills in multi-modal AI system engineering and gain hands-on experience with robotics and AI integration.	LLMs, ROS Hardware: KMIT’s Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1PJlFFnznlZFGcCaz43WZ0u5p3nlbEVkk/view?usp=drive_link https://arxiv.org/pdf/2212.04088	https://github.com/OSU-NLP-Group/LLM-Planner https://github.com/zeetwit/IlmRover	**KMIT’ Self-Driving rover platform demo:** https://youtube.com/shorts/qNh9c_iTQcA?si=0P8MbUL21VaF_4go	

KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments	
Defence Technol	**FalconEye**	**FalconEye:** Intelligent Tracking Robot	**Problem Statement** Develop an end-to-end robotic system that can detect, track, and follow any object of interest in real time, using an open-vocabulary, multimodal approach. The system should accept queries in text, image, or click form, detect and segment the target object in video frames, and robustly track it even under occlusions or reappearances. The solution must run on commodity hardware (laptop with a mid-range GPU) and achieve interactive frame rates. **Learning Outcomes** By completing this project, students will be able to: 1. Understand and implement open-vocabulary object detection and segmentation using foundation models. 2. Design multimodal query interfaces (text, image, click) for specifying objects of interest. 3. Apply tracking algorithms to maintain object identity across video frames, handling occlusions and reappearance. 4. Integrate perception and control in a real-time robotics loop for autonomous object following. 5. Optimize and deploy computer vision models for real-time inference on resource-constrained hardware. 6. Gain experience with open-source robotics and vision frameworks, and demonstrate their system on a real or simulated robotic platform.	Multimodal LLMs, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1m3C_F3s6b_2GxPOmaU5rKvJzDKz_vkRx/view?usp=drive_link https://arxiv.org/pdf/2308.05737	https://github.com/alaamaalouf/followanything	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbULz1VaF_4qp		
Healthcare	**BuzzKill**	**BuzzKill**: Autonomous Mosquito Breeding Ground Eliminator	**Problem statement** Design and develop an end-to-end autonomous mosquito breeding place detection and eradication rover that can autonomously navigate a defined area, detect potential mosquito breeding sites using computer vision and obstacle detection, and mitigate them by spraying a chemical agent. The system should use a custom-trained models for detection, implement autonomous navigation and obstacle avoidance, and demonstrate improved efficiency over manual control methods. **Learning Outcomes** By completing this project, students will be able to: 1. Apply computer vision techniques to build and train a custom object detection model. 2. Develop autonomous navigation algorithms with obstacle detection and path following. 3. Evaluate and compare the efficiency of autonomous mitigation systems against manual/traditional methods. 4. Collaborate using version control platforms (e.g., GitHub) to manage and document project code. **Real time application** Andhra Pradesh Pioneers AI-Powered Mosquito Control With Smart Surveillance System	CV, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/14V37HeHqq7CI6HCFPHfem_64_sm3CSzN/view?usp=drive_link https://arxiv.org/abs/2409.08078	https://github.com/faiyazabdullah/MosquitoMiner?tab=readme-ov-file	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbULz1VaF_4qp		
Supply Chain	**AutoPilot VLM**	**AutoPilot VLM**: VLMs Behind the Wheel	**Problem Statement** Develop an end-to-end autonomous driving agent using a vision-language model (VLM)–based framework inspired by LightEMMA. The project aims to explore the potential and limitations of VLMs for safe and reliable vehicle control in a **simulated environment.** Students will build, integrate, and evaluate lightweight multimodal driving agents using state-of-the-art VLMs and benchmark their performance (inference time, computational cost, predictive accuracy) on a standard dataset such as nuScenes. Students must implement the full pipeline: data preparation, model integration, agent construction, evaluation, and result analysis, while critically examining VLMs' strengths and weaknesses in autonomous driving tasks. **Learning Outcomes** By completing this project, students will be able to: 1. Explain the role of vision-language models in multimodal AI tasks. 2. Design and implement a lightweight end-to-end autonomous driving framework using VLMs. 3. Integrate commercial and open-source VLMs into a unified system. 4. Evaluate AI models based on inference time, computational cost, and prediction accuracy using standard datasets. 5. Analyze and articulate the limitations of current VLMs in real-world autonomous driving scenarios. 6. Develop reproducible and modular code following open-source practices.	VLMs, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1EI2UlYGjrv-qmZttag5B1V5SYEyB-JWa/view?usp=drive_link https://arxiv.org/pdf/2505.00284v1	https://github.com/michigan-traffic-lab/lightemma	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbULz1VaF_4qp		
Supply Chain	**ChatPilot**	**ChatPilot**: Conversational Autonomous Driving Agent	**Problem Statement** Design and implement an end-to-end autonomous driving agent that **integrates a video-language foundation model to enable natural language communication between humans and an autonomous vehicle.** The agent should operate in a simulated driving environment, demonstrating its ability to perceive the environment, handle long-horizon navigation tasks, interact via free-form dialogue, and adapt to unexpected changes in the environment or task. The system must also analyze and highlight the current limitations of such models in terms of inference latency, visual understanding, multi-turn dialogue, and real-time adaptability. **Learning Outcomes** By completing this project, students will be able to: 1. Understand the architecture and principles of video-language foundation models (FMs) in embodied AI. 2. Implement a simulation-based autonomous driving system with integrated human-agent dialogue. 3. Develop and evaluate an agent's ability to handle dynamic environments and unforeseen task changes. 4. Analyze and critique the limitations of current FM-based approaches in terms of response time, dialogue quality, and situational awareness. 5. Gain hands-on experience with multi-modal data (video and language) and real-time system constraints. 6. Practice end-to-end software development, from model integration and simulation setup to testing and reporting results.	VLMs, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1Zwjy3OmM-yL2KSjE9CR81TYGhgHlBQTJ/view?usp=drive_link https://arxiv.org/abs/2406.03008	https://github.com/sled-group/driVLMe?tab=readme-ov-file	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbULz1VaF_4qp		
Supply Chain	**SeeND**	**SeeND**:Autonomous Delivery Rover with Vision-based Drop Point Detection	**Problem Statement** This project involves the design and development of an autonomous ground rover that can perform last-mile delivery tasks by navigating its environment, identifying drop-off locations using Computer Vision (CV), and confirming successful delivery through AI-based validation techniques. The system utilizes real-time image processing to detect features such as QR codes, house numbers, or visual markers to locate the intended drop point. It combines path-planning algorithms, object detection models (e.g., YOLO), and OCR tools (e.g., Tesseract) to autonomously navigate and perform deliveries without human intervention. Additionally, the rover can generate intelligent status reports using Generative AI (e.g., GPT) for real-time feedback to users or operators. These include delivery confirmation messages, error reporting, and location-based summaries. **Learning Outcomes** By completing this project, learners will be able to: 1. Apply computer vision techniques such as object detection and OCR to recognize delivery drop points. 2. Train and deploy real-time object detection models (e.g., YOLOv5/YOLOv8) for embedded systems. 3. Design and implement autonomous navigation systems with obstacle avoidance using sensors and camera input. 4. Integrate and control hardware components including motors, cameras, ultrasonic sensors, and GPS modules. 5. Develop intelligent behavior logic for autonomous task execution (e.g., detect → verify → deliver). 6. Generate natural language delivery updates using Generative AI models (e.g., OpenAI GPT). 7. Implement cloud or IoT integration for live tracking, remote updates, or data logging. 8. Analyze and optimize system performance metrics such as detection accuracy and navigation stability. 9. Solve real-world problems using interdisciplinary knowledge from AI, robotics, and embedded systems. 10. Collaborate effectively in a technical team, documenting and presenting the project in a professional manner.	CV, OCR, GenAI, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1pDmhYIlW3vCn2rD-vwPcpOqq-U27T3D/view https://arxiv.org/pdf/2412.09627v1	https://github.com/wzzheng/doe	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbULz1VaF_4qp		
Supply Chain	**RoboJournalist Rover**	**RoboJournalist Rover**: AI-Powered Autonomous Field Reporter	**Problem Statement** The RoboJournalist Rover is an AI-enabled mobile robot designed to autonomously navigate environments, capture multimedia content (images, videos, or audio), and generate context-aware news reports or summaries using Generative AI. It uses Computer Vision to detect and interpret events (e.g., crowds, activities, objects), and leverages Language Models (like GPT) to convert this data into coherent textual reports or real-time social media updates. This system is especially useful for campus events, field data logging, construction monitoring, or emergency situations. **Learning Outcomes** By completing the RoboJournalist Rover project, students will be able to: 1. Design an autonomous mobile robot capable of capturing visual and audio data in dynamic environments. 2. Apply computer vision techniques for event detection, object tracking, and scene understanding. 3. Train and deploy models for human activity recognition and crowd analysis. 4. Use Generative AI (LLMs) to convert structured and unstructured sensory data into natural language summaries. 5. Implement multimodal AI pipelines combining image, audio, and text inputs for real-time reporting. 6. Integrate hardware components including cameras, microphones, motors, and GPS with embedded platforms. 7. Develop navigation and control systems for rovers using sensor fusion and path-planning algorithms. 8. Optimize edge-based AI processing for real-time multimedia analytics. 9. Build cloud-connected interfaces for remote monitoring and report publishing. 10. Collaborate in interdisciplinary teams and document technical work clearly for public or academic presentation.	VQA, GenAI, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1QO7DR2P4S7cqWiNUgvAawdlVvKUZafvl/view?usp=drive_link https://arxiv.org/pdf/2302.14115	https://github.com/google-research/scenic/tree/main/scenic/projects/vid2seq	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbULz1VaF_4qp		

KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments	
Healthcare	**MediBot**	**MediBot:** Medication Delivery Rover	**Problem Statement** The Medication Delivery Rover is an indoor autonomous ground robot designed to transport and deliver medications within hospitals or elderly care centers. It navigates through hallways and rooms, identifies patients using computer vision, and verifies delivery using face recognition, QR codes, or NFC tags. The rover integrates Computer Vision for patient/room recognition and Generative AI to provide reminders, verbal guidance, and automated logs. This reduces the workload on nurses, minimizes human error, and improves medicine delivery reliability. **Learning Outcomes** 1. Implement indoor robot navigation using CV-based SLAM or marker tracking 2. Learn to integrate face recognition and QR/NFC-based identification 3. Use generative AI to automate report generation and patient interaction 4. Handle real-time decision-making with sensor fusion (camera + distance sensors) 5. Design secure, human-centric HCI systems in healthcare settings 6. Explore ethical considerations in patient data handling and automation 7. Develop an integrated full-stack robotics system: perception → decision → action 8. Build user interfaces for real-time monitoring and control 9. Understand hospital workflow and how automation can improve delivery efficiency 10. Prototype a real-world use case with strong research and commercial potential	CV, GenAI, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/12IS_CG90LaB6PzT0eYTGIUPRSByIeS7X/view?usp=drive_link https://arxiv.org/pdf/2307.00666	https://github.com/manglanisagar/vision-search-navigation	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbJLz1VaF_4qp		
Healthcare	**Emobot**	**Emobot:** Emotion Expressive Rover	**Problem Statement** The Emotion Expressive Rover is a mobile robot that can detect human emotions using facial expressions and respond with matching or comforting expressions, speech, gestures, or movements. It aims to build empathy and connection between humans and machines — useful in elder care, pediatric wards, education, or therapy. **Learning Outcomes** 1. Understand facial expression recognition using CV/ML 2. Design emotionally expressive robotic interfaces (face, sound, gesture) 3. Apply GenAI to emotion-aware conversations 4. Combine perception, cognition, and expression in a real-time system 5. Explore ethics in affective HCI (e.g., privacy, emotional manipulation) 6. Work with servo/motor synchronization and emotion-to-movement mapping 7. Evaluate emotional responses from real users (quantitative + qualitative)	CV, GenAI, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1ZTxmgYYwBPQpOOq_CHoRO0EAQacaM-Lk/view?usp=drive_link https://arxiv.org/pdf/2407.21315v4	https://github.com/zehuiwu/SpeechCueLLM	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbJLz1VaF_4qp		
Supply Chain	**DevoRo**	**DevoRo:** Divine Seva Rover	**Problem Statement** The Divine Seva Rover is a mobile robot designed to assist in daily seva rituals in a temple, such as offering flowers, lighting aarti lamps, presenting water, incense waving, and chant playing — all with Computer Vision, synchronized motion, and generative spiritual interaction. It respects tradition, while automating repetitive and timed rituals with utmost precision, cleanliness, and spiritual alignment. The provided reference research work must be adapted to this problem statement. While most of the robots are deployed in hospitality domain, this is a unique attempt to cater to the needs specific to a community. **Learning Outcomes** 1. Apply AI in culturally sensitive and sacred applications 2. Design expressive motion using servo control and arm kinematics 3. Use CV for symbolic object detection and shrine positioning 4. Implement time-triggered ritual logic (like temple schedules) 5. Work with multilingual GenAI to produce Sanskrit or vernacular chants 6. Understand ethical design in religious robotics 7. Use HRI (Human-Robot Interaction) for devotees to trigger custom sevas	CV, GenAI, ROS Hardware: KMIT's Self-Driving Rover Platform (live demo link in the last column)	SAIKRISHNA	https://drive.google.com/file/d/1eKx1Pe0YRyk4gupwLOgwdy-suNwauYYX/view?usp=drive_link https://arxiv.org/pdf/2411.18539v1	https://github.com/dillonloh/adavln	**KMIT' Self-Driving rover platform demo:** https://youtube.com/shorts/pNh9c_iTQcA?si=0P8MbJLz1VaF_4qp		
FinTech	**StockCast**	A hybrid CNN-LSTM + XGBoost pipeline for predicting short-term stock trends using daily OHLCV data.	This paper introduces AttCLX, a hybrid model combining ARIMA, attention-enhanced CNN-LSTM, and XGBoost to improve stock price prediction accuracy. It uses ARIMA for data preprocessing, then applies CNN-LSTM with attention mechanisms to capture deep temporal patterns. XGBoost further refines predictions by modeling residual nonlinearities. The method significantly outperforms traditional models like ARIMA, LSTM, and XGBoost alone, demonstrated on Bank of China stock data.	MERN with AI	Sireesha	https://drive.google.com/file/d/1hVl8SCikVTu23uxOA9QiLwn8xS7gqkD/view?usp=sharing	https://github.com/zshicode/Attention-CLX-stock-prediction		https://arxiv.org/abs/2204.02623	
FinTech	**FinTweet**	Mining tweets for sentiment analysis using VADER/TextBlob and comparing with stock movement.	This paper demonstrates that public mood derived from Twitter can predict stock market movements. Analyzing tweets from 2008 using sentiment analysis tools, it identifies mood dimensions, especially "Calm," as significant predictors of Dow Jones Industrial Average changes. A fuzzy neural network model achieves high accuracy (~87.6%) in forecasting stock direction based on mood data. The study highlights social media sentiment as a valuable indicator for financial forecasting.	MERN with AI	Sireesha	https://drive.google.com/file/d/14xx319zhbBgFCkxDpWeOzo1olJLpvW-S/view?usp=sharing	https://www.catalyzex.com/paper/twitter-mood-predicts-the-stock-market/code		https://arxiv.org/abs/1010.3003	
FinTech	**CausalCast**	FinSen dataset that revolutionizes financial market analysis by integrating economic and financial news articles from 197 countries with stock market data	This paper presents FinSen, a large-scale dataset of financial news and global stock data to enhance market predictions. It uses causality-driven sentiment features with LSTM models to improve accuracy. A new loss function, Focal Calibration Loss, ensures better alignment between predicted probabilities and actual outcomes. The approach delivers more reliable and calibrated financial forecasts for high-stakes decisions.	MERN with AI	Sireesha	https://drive.google.com/file/d/1q5RqGJKWaV3B2nvYttL2FEKQjvoZxc1h/view?usp=sharing	https://github.com/EagleAdelaide/FinSen_Dataset		https://arxiv.org/abs/2408.01005	
FinTech	**EntityPulse**	Entity-level sentiment analysis for accurately assessing the sentiment directed toward a specific financial entity.	The paper introduces FinEntity, a dataset for entity-level sentiment classification in financial texts. It labels sentiment for specific entities using BILOU tagging and benchmarks models like FinBERT-CRF, which outperforms ChatGPT. Results show that entity-level sentiment improves stock price forecasting, especially for cryptocurrencies. The dataset and tools are publicly available for financial analysis and research.	MERN with AI	Sireesha	https://drive.google.com/file/d/1T_TzPv_rTbtHINGhNZfVtVmPM76tQxI5/view?usp=sharing	https://github.com/yixuantt/FinEntity		https://arxiv.org/abs/2310.12406	
Legal System	**CasePulse**	A web tool where you enter a legal research question and it retrieves the top 5 relevant cases and then generates a concise "analysis brief" grounded in those cases.	The paper introduces CLERC, a large dataset of U.S. federal court cases designed for legal case retrieval and analysis generation. It enables two key tasks: retrieving relevant case citations and generating legal analysis with those citations. Benchmarks show current models struggle, especially with long-context understanding and hallucinations. CLERC serves as a challenging benchmark to advance legal AI research.	MERN with AI	Sireesha	https://drive.google.com/file/d/1s7NE2U7ICemmlRI7HAf5C2EaDYJfaf55/view?usp=sharing	https://github.com/bohanhou14/CLERC		https://arxiv.org/abs/2406.17186	
Legal System	**LexChat**	A multi-turn "legal advice" chat where users ask follow-up questions about a topic	The paper presents LexRAG, a benchmark for retrieval-augmented generation in multi-turn legal consultations. It includes over 1,000 dialogues and 17,000 annotated legal documents. The study evaluates both legal document retrieval and accurate response generation. Results show existing RAG systems struggle with legal dialogue, highlighting the need for improved legal AI tools.	MERN with AI	Sireesha	https://drive.google.com/file/d/1LHWNqHBfeLPRMpHs5W1cAC5qeiltR6m/view?usp=sharing	https://github.com/CSHaitao/LexRAG		https://arxiv.org/abs/2502.20640	
Legal System	**BenchBot**	Legal Q&A portal	The paper introduces LegalBench-RAG, a benchmark focused on evaluating retrieval quality in legal retrieval-augmented generation (RAG) systems. It includes 6,858 expert-annotated query–snippet pairs mapped to a large legal corpus. A smaller variant, LegalBench-RAG-mini, supports quick testing. Results show existing retrievers struggle with snippet-level accuracy, highlighting challenges in legal RAG.	MERN with AI	Sireesha	https://drive.google.com/file/d/1j1fttEI9xIM_Sm3N-vgDvFL6cXgV9b7I/view?usp=drive_link	https://github.com/zeroentropy-cc/legalbenchrag		https://arxiv.org/abs/2408.10343	
Supply Chain	**SupplyGraph**	Forcasting in Supply Chain Mgmt	The paper introduces SupplyGraph, a real-world dataset modeling supply chain networks with temporal and relational data. It uses Graph Neural Networks (GNNs) to capture complex interactions between production nodes, warehouses, and product flows. GNN-based models outperform traditional baselines in demand forecasting tasks. This approach enhances supply chain efficiency through better planning and prediction.	MERN with AI	Sireesha	https://drive.google.com/file/d/1XnYgcijmcK5KQd8ictQVRM1_hd_CWe6/view?usp=sharing	https://github.com/cioi-researchlab/SupplyGraph		https://arxiv.org/pdf/2501.06221	
Supply Chain	**Generative Chain Sim**	Supply Chain Optimization via Generative Simulation and Iterative Decision Policies	This paper introduces Sim-to-Dec, a data-driven framework for optimizing supply chains using generative simulation and iterative decision policies. It learns supply chain behavior through autoregressive simulation, avoiding handcrafted rules. A dual-aware decision model interacts with the simulator to refine transport strategies. The method enhances responsiveness and efficiency in dynamic supply chain environments.	MERN with AI	Sireesha	https://drive.google.com/file/d/17QS1YsNtYXf4PXVE7ZR9eqeKlYHq-sO/view?usp=sharing	https://arxiv.org/abs/2507.07355		https://arxiv.org/abs/2507.07355	
Defence Technology	**GeoShield** AI-Driven Surveillance of Defense Zones	**Project Idea** With increasing border threats and covert installations in remote regions, traditional surveillance methods struggle to scale. This project aims to build an AI-powered surveillance and defence system that uses zero-shot classification of satellite images to detect unusual movements, constructions, or land-use changes -- especially in data-scarce zones. Powered by Microsoft's GeoVision Labeler (GVL), students will leverage vLLMs and LLMs for modular, interpretable image analysis with no labeled training data needed. **Development Tasks** 1. Understand the Domain and Tools 2. Set Up the Environment and Data Pipeline 3. Implement Zero-Shot Classification Pipeline 4. Modular and Interpretable Analysis Using LLMs 5. Visualize Results 6. Test, Evaluate & Document	VLM, AI Agents, MERN, FastAPI, Docker	Sripooja	Geo Vision Labeler https://arxiv.org/pdf/2505.24340 https://drive.google.com/file/d/1VfF-bLnapgWZbuV5dRVUBaHssyFLirzp/view?usp=share_link	https://github.com/microsoft/geo-vision-labeler			
Legal System	**JusticeBot** Your Reliable AI Legal Advisor	**Project Idea** Build a smart legal assistant that gives more accurate and trustworthy advice by combining multiple expert models, legal knowledge, and real law firm workflows—reducing the risk of giving wrong or misleading answers. **Development Tasks** 1. Curate and Structure Legal Knowledge Base 2. Select and Evaluate Pretrained LLMs 3. Design the Expert System Logic 4. Build a Fact-Checking and Validation Layer 5. Implement Workflow Automation (SOP) 6. Develop User Interface and Test the System	MoE, Agentic workflows, MERN, FastAPI, Docker	Sripooja	ChatLaw https://arxiv.org/pdf/2306.16092v2 https://drive.google.com/file/d/1AQStJfvfyve5J5vpBHfWpakfEnsGBH77/view?usp=share_link	https://github.com/pku-yuangroup/chatlaw			
Healthcare	**MediVision** Multimodal AI for Real-time Chest X-ray Diagnosis	**Project Idea** Current AI tools for chest X-ray analysis work well individually but fail to support doctors in real clinical settings where multiple types of information must be combined. This project will build a smart assistant that answers complex medical questions by understanding both images and text, making CXR interpretation more practical and useful for healthcare professionals. **Development Tasks** 1. Design and implement the system architecture 2. Develop the query processing and orchestration module 3. Integrate pre-trained CXR analysis and LLM models 4. Build a unified interactive chat user interface 5. Implement response synthesis and explanation generation 6. Evaluate and benchmark the system	AI Agents, MERN, FastAPI, Docker	Sripooja	MedRAX https://arxiv.org/pdf/2502.02673v2 https://drive.google.com/file/d/1OZmlOfDogLOR2uxvvbUKjKNadkSNuQg6/view?usp=share_link	https://github.com/bowang-lab/medrax			

KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments	
Defence Technology	**DroneHunter**	**DroneHunter** Autonomous Detection and Tracking of Unknown UAVs	**Project Idea** Design and implement an AI-based system for detecting and tracking unidentified drones in real-time, without any prior knowledge of their appearance, using thermal video data from complex outdoor environments. **Development Tasks** 1. Design the system architecture integrating detection and tracking modules. 2. Preprocess and annotate thermal video sequences for evaluation. 3. Implement a real-time UAV detection algorithm for complex scenes. 4. Develop a UAV tracking module to handle appearance and reappearance. 5. Integrate detection and tracking outputs into a seamless pipeline. 6. Evaluate system performance using standard metrics and benchmark scenarios.	Agentic workflows, MERN, FastAPI, Docker	Sripooja	Anti-UAV https://arxiv.org/pdf/2306.15767v2 https://drive.google.com/file/d/1tsIm2mo991GKvGY9IKCvIHGlUg3CByeP/view?usp=share_link	https://github.com/ZhaoJ9014/Anti-UAV			
Healthcare	**CounselAI**	**CounselAI** A Proactive, Expressive Voice Assistant for Mental Health	**Project Idea** The aim of the project is to develop a voice-based AI assistant for mental health counseling that provides real-time, emotionally expressive, and personalized support to users. The system will engage in natural, empathetic conversations by recognizing emotional cues and adapting its tone, rhythm, and persona to suit individual needs. Designed to be proactive, responsive, and privacy-conscious, the assistant seeks to improve accessibility and trust in psychological care through seamless human-like interactions. **Development Tasks** 1. Design the system architecture integrating Voila's API with a secure healthcare-grade backend. 2. Implement user persona and voice customization based on text-defined parameters for therapy styles. 3. Develop real-time emotion detection and adaptive tone modulation in conversations. 4. Build a privacy-compliant user session management and logging module with anonymization. 5. Create an intuitive user interface for patients to access, schedule, and interact with the AI counselor. 6. Conduct usability testing and refine the dialogue flow for empathetic and context-aware responses.	Voice-based AI, Agentic workflows, MERN, FastAPI, Docker	Sripooja	Voila (VoiceFM) https://arxiv.org/pdf/2505.02707 https://drive.google.com/file/d/1MOh_un6LE3Vdzq0_GBpRGu-z5iIOxiBv/view?usp=share_link	https://github.com/maitrix-org/voila			
AI	**SciFusion**	**SciFusion** AI-Driven Innovation in Scientific Research	**Project Idea** Design an AI-powered assistant for accelerating hypothesis generation and validation in scientific research projects by leveraging NovelSeek's multi-agent framework. The system should autonomously propose novel research ideas, run simulations or baseline experiments, and incorporate expert feedback in a closed-loop cycle to iteratively improve results across at least one selected domain (e.g., bioinformatics, materials science, or environmental modeling). **Development Tasks** 1. Design and implement the user interface for researchers to input goals and review AI-generated hypotheses. 2. Integrate NovelSeek's multi-agent framework as the backend engine for hypothesis generation and validation. 3. Develop a feedback loop mechanism to capture and incorporate expert corrections into the process. 4. Implement logging and reporting modules to track experiments, outcomes, and performance metrics. 5. Optimize the workflow orchestration for scalability across different scientific tasks or domains. 6. Conduct rigorous testing of the end-to-end system for usability, accuracy, and efficiency under realistic scenarios.	Multi-Agent, MERN, FastAPI, Docker	Sripooja	NovelSeek https://arxiv.org/pdf/2505.16938v1 https://drive.google.com/file/d/12Q66ly3QQrpYDE5xX0oFV2q_s8nP3XDQ/view?usp=share_link	https://github.com/alpha-innovator/noveleek			
AI	**MultiTalk**	**MultiTalk** Generating Lifelike Multi-Speaker Conversational Avatars	**Project Idea** Existing audio-driven animation models struggle to generate realistic multi-person conversational videos due to incorrect audio-person associations and poor instruction-following. This project aims to build a system that accurately binds multiple audio streams to corresponding animated characters. It will leverage techniques like Label Rotary Position Embedding (L-RoPE) and multi-task training to enhance synchronization and dialogue coherence. **Development Tasks** 1. Design the system architecture for multi-person audio-driven video generation with instruction-following capability. 2. Implement audio-person binding using Label Rotary Position Embedding (L-RoPE). 3. Develop the multi-stream audio injection and synchronization module. 4. Integrate an instruction parser to guide dialogue flow and visual actions. 5. Build a video rendering pipeline to produce high-quality conversational animations. 6. Evaluate the system's performance on benchmark datasets and analyze results.	Audio-driven animation, MERN, FastAPI, Docker	Sripooja	LetThemTalk https://arxiv.org/pdf/2505.22647v1 https://drive.google.com/file/d/1AsQA6KOlmCIbvYiGDO0SAJwYAINAI1wk/view?usp=share_link	https://github.com/meigen-ai/multitalk			
FinTech	**FinSight Agents**	**FinSight Agents** Exploring Market Behavior Through AI Agents	**Project Idea** This project aims to build a multi-agent system using Large Language Models (LLMs) to mimic investor actions in a virtual stock market. By studying how these agents respond to real-world news and market events, we can uncover patterns in trading behavior and evaluate how such factors influence market outcomes—without relying on future data or test leakage. The system can be used to explore smarter investment strategies and improve financial decision-making tools. **Development Tasks** 1. Design and implement a multi-agent simulation framework to mimic diverse investor profiles and trading strategies. 2. Integrate external factor inputs (e.g., macroeconomic indicators, policy changes, global events) into the simulation environment. 3. Develop a mechanism to prevent test set leakage and ensure agents operate without prior knowledge of evaluation data. 4. Build a dashboard to visualize trading behaviors, stock price fluctuations, and agent performance metrics. 5. Implement analysis modules to identify patterns and correlations between external factors and trading outcomes. 6. Evaluate and compare different LLM-based agent configurations within the simulation and document insights.	Multi-Agent, MERN, FastAPI, Docker	Sripooja	StockAgent https://arxiv.org/pdf/2407.18957 https://drive.google.com/file/d/1vW0IcBEsEuWKusBRtHoj-2zdYNKG2LUx/view?usp=share_link	https://github.com/mingyuj666/stockagent			
Material Science	**CrystalGen**	**CrystalGen** Revolutionizing Materials Design with Diffusion-Based Generative Models	**Project Idea** The discovery of new functional materials with specific properties—such as stability, magnetism, or conductivity—is crucial for innovations in energy, catalysis, and sustainability. Traditional computational and experimental methods are slow and resource-intensive, while existing generative models struggle to design stable crystalline structures or meet multiple property constraints. This project aims to explore and enhance MatterGen, a diffusion-based generative model capable of producing diverse and stable inorganic materials tailored to user-defined chemical, structural, and functional requirements, pushing the boundaries of AI-driven materials design. **Development Tasks** 1. Implement the diffusion-based generative pipeline to iteratively refine atomic structures and lattice parameters. 2. Develop adapter modules to enable fine-tuning of generated materials towards specific property constraints. 3. Integrate stability and property evaluation modules to assess generated structures against physical criteria. 4. Design a user interface for specifying desired chemical, structural, and functional property inputs. 5. Conduct benchmarking of generated materials against existing models for novelty, stability, and property satisfaction. 6. Demonstrate multi-property generation by producing and validating materials with combined target properties (e.g., magnetic density and supply-chain safety).	Generative models, MERN, FastAPI, Docker	Sripooja	MatterGen https://arxiv.org/pdf/2312.03687v2 https://drive.google.com/file/d/17JEkSfci-y-49rZqTMlh3ScJW9H3Eego/view?usp=share_link	https://github.com/microsoft/mattergen			
Defence Technology	**IntelTrace**	**IntelTrace** Reasoning-Based Segmentation for Tactical Defence Surveillance using Multimodal LLMs	**Project Idea** In modern defence scenarios, identifying critical visual elements (e.g., camouflaged threats, suspicious patterns, or hidden installations) often requires contextual understanding beyond explicit labels. This project aims to develop a reasoning-based segmentation assistant using multimodal LLMs that can interpret natural language defence queries and generate segmentation masks in complex, high-resolution imagery (e.g., satellite or drone feeds). The system will enable zero-shot detection and segmentation of semantically complex objects using implicit instructions, enhancing situational awareness and threat identification in real time. **Development Tasks** 1. Design the system architecture integrating multimodal LLM with a segmentation inference pipeline. 2. Implement the reasoning-aware query parsing and context interpretation module. 3. Develop the image pre-processing and mask post-processing components for high-resolution defence imagery. 4. Integrate zero-shot inference capability with real-time surveillance feed handling. 5. Build an intuitive user interface for submitting natural language queries and visualizing segmentation results. 6. Conduct quantitative and qualitative evaluation of system performance on benchmark scenarios.	Multimodal LLMs, AI Agents, MERN, FastAPI, Docker	Sripooja	LISA https://arxiv.org/pdf/2308.00692 https://drive.google.com/file/d/1KS-2UmV1Ab4B1jCDH0Bhnc3JUopKtc0b/view?usp=share_link	https://github.com/dvlab-research/LISA			

KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments	
FinTech	Finbloom	FinBloom: Knowledge Grounding Large Language Model with Real-time Financial Data	This paper introduces a knowledge-grounding approach called Financial Agent for Large Language Models (LLMs) to address financial queries using real-time text and tabular data. Problem Addressed: LLMs generally excel at generating human-like responses but struggle with interactive tasks that require access to real-time information. This limitation is particularly challenging in finance, where up-to-date data like recent news or price movements are crucial for decision-making. Traditional methods of frequently training or fine-tuning LLMs on high-velocity financial data are costly and time-consuming, and it's difficult to ascertain what the LLM has learned from new data. Retrieval-Augmented Generation (RAG) models also face limitations in finance due to difficulties in representing structured financial data, challenges in retriever accuracy for specific financial queries, and an inability to handle high-velocity, high-volume data Proposed Solution - FinancialAgent: To overcome these challenges, the authors propose an alternative approach where the core LLM remains ""frozen"" (weights are not changed or fine-tuned), and instead, external modules are added for knowledge grounding. This framework involves a Data Module that provides access to real-time financial information and a Financial Agent that interprets user queries, extracts relevant data from the Data Module, and constructs the necessary financial context. The system architecture involves the user inputting a query (x), the Financial Agent processing it and retrieving relevant news and tabular data (d) from the Data Module. This data is then converted into a text format (c(d)) and appended to the user query (x), forming a combined input (c(d), x) for a larger LLM (e.g., GPT 3.5 or GPT 4). This process reduces latency and eliminates the need for users to manually provide accurate data	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1bRKfzeelusosnWezId-mStAXOGjuxPjn/view?usp=drive_link	https://huggingface.co/Chaitanya14/FinBloom_7B https://huggingface.co/Chaitanya14/Financial_Agent			
AI Agriculture	Agrollm	AgroLLM: Connecting Farmers and Agricultural Practices through Large Language Models for Enhanced Knowledge Transfer and Practical Application	This text introduces AgroLLM, an innovative chatbot designed to enhance agricultural knowledge transfer and practical application for farmers through Large Language Models (LLMs). The system utilizes a Retrieval Augmented Generation (RAG) framework, drawing information from a comprehensive dataset of agricultural resources, including textbooks and research articles. A comparative study of leading LLMs such as Gemini 1.5 Flash, ChatGPT-4o Mini, and Mistral-7B-Instruct-v0.2 within the AgroLLM framework revealed that ChatGPT-4o Mini with RAG achieved the highest accuracy at 93%. While RAG significantly improves accuracy across models, it also increases response time compared to using FAISS (Facebook AI Similarity Search) for direct retrieval. The document emphasizes AgroLLM's potential as a valuable educational and decision-support tool in agriculture, bridging theoretical knowledge with practical farming needs.	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1_DIeFMeYZhtQBMalCugkg4jgNnqQRN0O/view?usp=sharing	https://drive.google.com/file/d/1_DIeFMeYZhtQBMalCugkg4jgNnqQRN0O/view?usp=sharing			
Legal System	Legalarggen	Mitigating Manipulation and Enhancing Persuasion: A Reflective Multi-Agent Approach for Legal Argument Generation	Large Language Models (LLMs) are increasingly explored for legal argument generation, yet they pose significant risks of manipulation through hallucination and ungrounded persuasion, and often fail to utilize provided factual bases effectively or abstain when arguments are untenable. This paper introduces a novel reflective multi-agent method designed to address these challenges in the context of legally compliant persuasion. Our approach employs specialized agents—a Factor Analyst and an Argument Polisher—in an iterative refinement process to generate 3-ply legal arguments (plaintiff, defendant, rebuttal). We evaluate Reflective Multi-Agent against single-agent, enhanced-prompt single-agent, and non-reflective multi-agent baselines using four diverse LLMs (GPT-4o, GPT-4o-mini, Llama-4-Maverick-17b-128e, Llama-4-Scout-17b-16e) across three legal scenarios: "arguable", "mismatched", and "non-arguable". Results demonstrate Reflective Multi-Agent's significant superiority in successful abstention (preventing generation when argumentscannot be grounded), marked improvements in hallucination accuacy (reducing fabricated and misattributed factors), particularly in "non-arguable" scenarios, and enhanced factor utilization recall (improving the use of provided case facts). These findings suggest that structured reflection within a multi-agent framework offers a robust computable method for fostering ethical persuasion and mitigating manipulation in LLM-based legal argumentation systems, a critical step towards trustworthy AI in law.	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1tUixc_DoxHLn6s8EM3nQX0ayvMjUxqji/view?usp=sharing	https://lizhang-aiandlaw.github.io/A-Reflective-Multi-Agent-Approach-for-Legal-Argument-Generation/ https://github.com/lizhang-AIandLaw/Mitigating-Manipulation-and-Enhancing-Persuasion			
AI Agriculture	CoffeeDisease Diagnosis	Vision Meets Language: A RAG-Augmented YOLOv8 Framework for Coffee Disease Diagnosis and Farmer Assistance	As a social being, we have an intimate bond with the environment. A plethora of things in human life, such as lifestyle, health, and food are dependent on the environment and agriculture. It comes under our responsibility to support the environment as well as agriculture. However, traditional farming practices often result in inefficient resource use and environmental challenges. To address these issues, precision agriculture has emerged as a promising approach that leverages advanced technologies to optimise agricultural processes. In this work, a hybrid approach is proposed that combines the three different potential fields of model AI: object detection, large language model (LLM), and Retrieval-Augmented Generation (RAG). In this novel framework, we have tried to combine the vision and language models to work together to identify potential diseases in the tree leaf. This study introduces a novel AI-based precision agriculture system that uses Retrieval Augmented Generation (RAG) to provide context-aware diagnoses and natural language processing (NLP) and YOLOv8 for crop disease detection. The system aims to tackle major issues with large language models (LLMs), especially hallucinations and allows for adaptive treatment plans and real-time disease detection. The system provides an easy-to-use interface to the farmers, which they can use to detect the different diseases related to coffee leaves by just submitting the image of the affected leaf the model will detect the diseases as well as suggest potential remediation methodologies which aim to lower the use of pesticides, preserving livelihoods, and encouraging environmentally friendly methods. With an emphasis on scalability, dependability, and user-friendliness, the project intends to improve RAG-integrated object detection systems for wider agricultural applications in the future.	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/16jn81_HLP-sQMjkoxa9fOE2gS6aP2X0E/view?usp=sharing	hishekvchaudhari/FarmTalk-Nexus-Vision-LLM-RAG e is CODE that integrates YOLOv8 with T-3.5 Turbo using a RAG framework {related work} Data Set: ta.mendeley.com/datasets/yy2k5y8mxg/1			
Legal System	Ragplus	RAG+: Enhancing Retrieval-Augmented Generation with Application-Aware Reasoning	The integration of external knowledge through Retrieval-Augmented Generation (RAG) has become foundational in enhancing large language models (LLMs) for knowledge-intensive tasks. However, existing RAG paradigms often overlook the cognitive step of applying knowledge, leaving a gap between retrieved facts and task-specific reasoning. In this work, we introduce RAG+, a principled and modular extension that explicitly incorporates application-aware reasoning into the RAG pipeline. RAG+ constructs a dual corpus consisting of knowledge and aligned application examples, created either manually or automatically, and retrieves both jointly during inference. This design enables LLMs not only to access relevant information but also to apply it within structured, goal-oriented reasoning processes. Experiments across mathematical, legal, and medical domains, conducted on multiple models, demonstrate that RAG+ consistently outperforms standard RAG variants, achieving average improvements of 3–5%, and peak gains up to 7.5% in complex scenarios. By bridging retrieval with actionable application, RAG+ advances a more cognitively grounded framework for knowledge integration, representing a step toward more interpretable and capable LLMs.	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1YjmTEAKwLeKOzmZPpszOfebSi-4IGQ7g/view?usp=sharing	Experimental setup, dataset and prompt templates are provided in the paper http://wenshu.court.gov.cn/			
AI Agriculture	Pllama	PLLaMa: An Open-source Large Language Model for Plant Science	Large Language Models (LLMs) have exhibited remarkable capabilities in understanding and interacting with natural language across various sectors. However, their effectiveness is limited in specialized areas requiring high accuracy, such as plant science, due to a lack of specific expertise in these fields. This paper introduces PLLaMa, an open-source language model that evolved from LLaMa-2. It's enhanced with a comprehensive database, comprising more than 1.5 million scholarly articles in plant science. This development significantly enriches PLLaMa with extensive knowledge and proficiency in plant and agricultural sciences. Our initial tests, involving specific datasets related to plants and agriculture, show that PLLaMa substantially improves its understanding of plant science-related topics. Moreover, we have formed an international panel of professionals, including plant scientists, agricultural engineers, and plant breeders. This team plays a crucial role in verifying the accuracy of PLLaMa's responses to various academic inquiries, ensuring its effective and reliable application in the field. To support further research and development, we have made the model's checkpoints and source codes accessible to the scientific community.	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1EUus7a32pb3Mw-5qdtYNq-ZOY4oSdVxG/view?usp=sharing	https://github.com/Xianjun-Yang/PLLaMa. https://huggingface.co/datasets/togethercomputer/RedPajama-Data-1T-Sample https://github.com/deepspeedai/DeepSpeed https://huggingface.co/docs/accelerate/usage_guides/fsdp			
Healthcare	**Myriad**	**MIRIAD: Augmenting LLMs with millions of medical query-response pairs**	Large language models (LLMs) are bound to transform healthcare with advanced decision support and flexible chat assistants. However, LLMs are prone to generate inaccurate medical content. In order to ground LLMs in high-quality medical knowledge, LLMs have been equipped with external knowledge sources via retrieval augmented generation (RAG), where unstructured medical knowledge is split into small chunks of text that can be selectively retrieved and integrated into the LLMs context. Yet, existing RAG pipelines rely on raw, unstructured medical text, which can be noisy, uncurated, and difficult for LLMs to effectively leverage. Systematic approaches to organize medical knowledge and to best surface it to LLMs are generally lacking. To address these challenges, here, we introduce MIRIAD, a large-scale, curated corpus of 5,821,948 medical instruction-response pairs, each rephrased from and grounded in a passage from peer-reviewed medical literature using a semi-automated pipeline combining LLM generation, filtering, grounding, and human annotation. Unlike prior medical corpora, which rely on unstructured text, MIRIAD encapsulates rich and web-scale medical knowledge in an operationalized query-response format, which enables more targeted retrieval. Experiments on challenging medical question-answering benchmarks show that augmenting LLMs with MIRIAD improves accuracy up to 6.7% compared to unstructured RAG baselines with the same source corpus and with the same amount of retrieved text. Moreover, MIRIAD improved the ability of LLMs to detect medical hallucinations by 22.5 to 37% (increase in F1 score). We further introduce MIRIAD-Atlas, an interactive semantic map of MIRIAD spanning 56 medical disciplines, enabling clinical users to visually explore, search, and refine medical knowledge. MIRIAD promises to unlock a wealth of down-stream applications, including medical information retrievers, enhanced RAG applications, and knowledge-grounded chat interfaces, which ultimately enables more reliable LLM applications in healthcare	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1MPEnhNEsgnl2XCl6ks2ZSy8UCn_2nLEO/view	GitHub - eth-medical-ai-lab/MIRIAD: MIRIAD is a million scale Medical Instruction and Retrieval Dataset https://huggingface.co/miriad			

KMIT 3-1 PS1 Projects

Domain	Short Name	Paper Title	Paper Description	Tech Stack	Mentor	Research Paper Link	Github/Code link	Slide Deck	Additional Comments
Healthcare	**Apollo**	**Apollo: A Lightweight Multilingual Medical LLM towards Democratizing Medical AI to 6B People**	Despite the vast repository of global medical knowledge predominantly being in English, local languages are crucial for delivering tailored healthcare services, particularly in areas with limited medical resources. To extend the reach of medical AI advancements to a broader population, we aim to develop medical LLMs across the six most widely spoken languages, encompassing a global population of 6.1 billion. This effort culminates in the creation of the ApolloCorpora multilingual medical dataset and the XMedBench benchmark. In the multilingual medical benchmark, the released Apollo models, at various relatively-small sizes (i.e.,0.5B, 1.8B, 2B, 6B, and 7B), achieve the best performance among models of equivalent size. Especially, Apollo-7B is the state-of-the-art multilingual medical LLMs up to 70B. Additionally, these lite models could be used to improve the multi-lingual medical capabilities of larger models without fine-tuning in a proxy-tuning fashion. We will open-source training corpora, code, model weights and evaluation benchmark	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/18KWuTJEbWp1Q3fUZoqwnujhYNbBni_iF/view?usp=sharing	https://github.com/FreedomIntelligence/Apollo		
Healthcare	**Moodangels**	**MoodAngels: A Retrieval-augmented Multi-agent Framework for Psychiatry Diagnosis**	The application of AI in psychiatric diagnosis faces significant challenges, including the subjective nature of mental health assessments, symptom overlap across disorders, and privacy constraints limiting data availability. To address these issues, we present MoodAngels, the first specialized multi-agent framework for mood disorder diagnosis. Our approach combines granular-scale analysis of clinical assessments with a structured verification process, enabling more accurate interpretation of complex psychiatric data. Complementing this framework, we introduce MoodSyn, an open-source dataset of 1,173 synthetic psychiatric cases that preserves clinical validity while ensuring patient privacy. Experimental results demonstrate that MoodAngels outperforms conventional methods, with our base-line agent achieving 12.3% higher accuracy than GPT-4o on real-world cases, and our full multi-agent system delivering further improvements. Evaluation in the MoodSyn dataset demonstrates exceptional fidelity, accurately reproducing both the core statistical patterns and complex relationships present in the original data while maintaining strong utility for machine-learning applications. Together, these contributions provide both an advanced diagnostic tool and a critical research resource for computational psychiatry, bridging important gaps in AI-assisted mental health assessment	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1Q0B71EAwdlI4RmjM-mUHlyRA_SPMwy14J/view?usp=sharing	https://github.com/elsa66666/MoodAngels		
Legal System	**Lawgpt**	**LAWGPT: KNOWLEDGE-GUIDED DATA GENERATION AND ITS APPLICATION TO LEGAL LLM**	Large language models (LLMs), both proprietary and open-source, have demonstrated remarkable capabilities across various natural language processing tasks. However, they face significant limitations in legal reasoning tasks. Proprietary models introduce data privacy risks and high inference costs, while open-source models underperform due to insufficient legal domain training data. To address these limitations, we study data generation for legal reasoning to improve the legal reasoning performance of open-source LLMs with the help of proprietary LLMs. This is challenging due to the lack of legal knowledge in proprietary LLMs and the difficulty in verifying the generated data. We propose KGDG, a knowledge-guided data generation framework for legal reasoning. Our frame-work enables leveraging legal knowledge to enhance generation diversity and introduces a refinement and verification process to ensure the quality of generated data. Moreover, we expand the generated dataset to further enhance the LLM reasoning capabilities. Using KGDG, we create a synthetic legal reasoning dataset containing 50K high-quality examples. Our trained model LAWGPT outperforms existing legal-specific LLMs and achieves performance comparable to proprietary LLMs, demonstrating the effectiveness of KGDG and LAWGPT.	Gen AI, Mern Stack	Ashok Sharma	https://drive.google.com/file/d/1Qdft92we8Td59K61RneeCGF5dYeaXjhW/view?usp=sharing	https://github.com/LAMDASZ-ML/Knowledge-Guide-Data-Generation		
Healthcare	GenAI_Surgery	GenAI_Surgery: GenAI based Multi model on surgical video	Surgical scene segmentation is the process of identifying and delineating different anatomical structures, surgical tools, organs, or regions of interest within surgical videos or images. It is a key task in computer-assisted surgery and surgical data science, enabling applications such as real-time surgical assistance,autonomous robotic surgery,surgical training & simulation and post-operative analysis.This project aims to implement GenAI based multimodel for surgical instrument segmentation,surgical action recognition and surgical scene understanding with vision transformer on videos obtained during surgical procedures.	GenAI, MERN, Docker, AWS	Dr.Devika		genAI_surgery.pdf	https://github.com/BCV-Uniandes/TAPIR	
Remote Ser	SAR_Marine	SAR_Marine_Surveillance: Marine Surveilance for maritime object and oil spils using GenAI	SAR_Marine_Surveillance refers to using Synthetic Aperture Radar (SAR) for monitoring marine environments, including vessel detection, tracking, and illegal activity recognition (like piracy, oil spills, or illegal fishing). This MERN App applies GENAI models on SAR images for for maritime objects (like ship) detection and Oil Spill Detection.	GenAI, MERN, Docker, AWS	Dr.Devika		SAR_Marine_Surveillance.pdf SAR_Marine_Surveillance_oil.pdf	https://github.com/fundamentalvision/Deformable-DETR/tree/main https://github.com/Beckschen/TransUNet/tree/main	
Healthcare	GenAI_Ophl	GenAI_Ophthalmologist:Glaucoma treatment and VQA for different imaging modalities of Opthamology	Glaucoma is a chronic, progressive eye disease that leads to o usually associated with increased intraocular pressure (IOP). It causes of irreversible blindness globally.Optic disc (OD) segme field (VF) progression prediction using fundus images is a critic detection and monitoring.This flutter App implement GenAI mo tasks along with VQA model for differnt imaging modalities of o	GenAI, flutter/firebase,Docker, AWS	Dr.Devika		GenAI_Ophthalmologist.pdf	https://github.com/Michi-3000/EyeCLIP	
Video comm	GenAI_Zoor	GenAI_Zoom:Video collaboration platform with conversational genai agent/experts	A Conversational GenAI agent is an intelligent, AI-powered ass interact naturally via text or voice. Celebrity GenAI agent mimic personality, or visual likeness of a celebrity using generative AI entertainment, knowledge sharing, or expert suggestion/evalua develops a realtime video collabration platform with conversatio and celebrity genAI agent.	GenAI, webRTC, MERN, socket.io	Dr.Devika		GenAI_Zoom.pdf	https://github.com/aquorio15/path_vqa	
Cybersecuri	Network_Gu	Network_Guard:Real time monitoring tool for network packet analysis	Network monitoring tools are essential for tracking, analyzing, network performance, availability, and security in real-time. This network monitoring tool for packet capture and inspection,real t visualization,genAI based network traffic/forensic analysis.	GenAI, Prometheus + Grafana/wireshark React, Docker/K8s	Dr.Devika		Network_Guard.pdf	https://github.com/linwhitehat/ET-BERT	
Cybersecuri	DB_Guard	DB_Guard:Database monitoring tool for transactional analysis	Database monitoring tools are essential for ensuring availability, performance, query optimization, and resource utilization of database systems. They help to detect bottlenecks, slow queries, memory leaks, and downtime before they impact users or applications.This project develop a multimodal DB monitoring tool that integrates SQL text, execution plans, logs, and key performance indicators (KPIs) to identify and rank root causes of a DB based on expected performance improvement	Prometheus + Grafana, MySQL/ MongoDB, Docker/K8s	Dr.Devika		DB_Guard.pdf	https://github.com/decisionintelligence/RCRank	
Drug Desigr	GenAI_Insulin_Drug:	GenAI_Insulin_Drug:Synthetic insulin drug design	Synthetic insulin analogs are engineered forms of insulin that have been chemically or genetically modified to improve their absorption, duration of action, stability, or receptor binding profile compared to native human insulin.They are crucial in diabetes treatment.This project generates synthetic insulin analog sequences using genAI models,predict its 3D structure and perform autodocking to INSR	GenAI, MERN, Docker, federated learning, AWS	Dr.Devika		GenAI_Insulin_Drug.pdf	https://github.com/salesforce/progen/tree/main/progen2	
Media	GenAI_YouT	GenAI_YouTube: GenAI based video/audio content creation	Video content generation involves using tools and technologies including AI/GenAI to create engaging videos with minimal manual effort. This is revolutionizing marketing, education, training, social media, and even entertainment production.This project develop multimodal generative model that allows "any-to-any" input-output combinations—such as text → video+audio, image+audio → video within a single unified framework .	GenAI, fluter/firebase, TTS, federated learning, AWS	Dr.Devika		GenAI_YouTube.pdf	https://codi-gen.github.io/	
Genomics	CRISPR_AI	CRISPR_AI: Gene-Editing tool for Off-targets activities prediction	CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a gene-editing tool that allows scientists to cut and modify DNA at specific locations in a genome, to correct disease-causing mutations at the DNA level. CRISPR can accidentally bind to similar, unintended DNA sequences (off-targets), causing unwanted genetic changes such as Cancer, Cell toxicity and Unpredictable gene function disruptions. Off-target prediction in CRISPR/Cas9 is critically important because it ensures precision, safety, and reliability in genome editing. This project helps to predict and interpret CRISPR/Cas9 off-target activities, particularly, When there are mismatches and indels (insertions/deletions) between the guide RNA (gRNA) and the target DNA sequence using BERT.	GenAI, MERN, Docker, federated learning, AWS	Dr.Devika		CRISPR_AI.pdf	https://github.com/BrokenStringx/CRISPR-BERT	
EdTech	GenAI_Moc	GenAI_Mock_interview: Framework for Human speech and facial expression analysis	Building a GenAI-based mock interview platform is a powerful idea that combines natural language understanding, LLMs, and voice/vision models to simulate realistic interview experiences across roles (tech, HR, MBA, clinical, etc.). This project develop a GenAI based framework to understand, generate and analyze human speech and facial expressions, which is needed for realistic virtual interviewers' tools.	GenAI, MERN, Docker/K8s, federated learning, AWS	Dr.Devika		GenAI_Mock_interview.pdf	https://github.com/facebookresearch/av_hubert	