RESOURCE ALLOCATION POLICIES FOR NEXT GENERATION NETWORKS USING REINFORCEMENT LEARNING AND LARGE LANGUAGE MODELS

MTech. Thesis

By

SHUBHAM SINHA



DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2026

RESOURCE ALLOCATION POLICIES FOR NEXT GENERATION NETWORKS USING REINFORCEMENT LEARNING AND LARGE LANGUAGE MODELS

A THESIS

Submitted in partial fulfilment of the
requirements for the award of the degree
of

Master of Technology

by

SHUBHAM SINHA

2402102006



DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

May 2026

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"Resource Allocation Policies for Next Generation Networks Using Reinforcement Learning and Large Language Models"** in partial fulfilment of the requirements for the award of the degree of **Master of Technology** and submitted in the **Department of Electrical Engineering, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the period from July 2024 to May 2026 under the supervision of **Dr. Dibbendu Roy**, Indian Institute of Technology Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Shubham Sinha
16/05/2026

Signature of the Student with Date

(SHUBHAM SINHA)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dibbendu Roy

15/05/2026.

Signature of Thesis Supervisor with Date

(Dr. Dibbendu Roy)

Shubham Sinha has successfully given his MTech. Oral Examination held on 27th April 2026.

Signature(s) of Supervisor(s) of MTech. Thesis.

Date: **15/05/2026**.

Convener, DPGC

Date: 15/05/2026

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, **Dr. Dibbendu Roy**, for his continuous guidance, valuable suggestions, and constant encouragement throughout the course of this research work. His technical insights, motivation, and support helped me successfully complete this thesis and deepen my understanding of intelligent wireless communication systems and machine learning based network optimization.

I am also thankful to all the faculty members and staff of the **Department of Electrical Engineering, Indian Institute of Technology Indore**, for providing an excellent academic and research environment during my MTech. Program.

I would also like to thank my friends and classmates for their support, discussions, and encouragement throughout this journey. Their cooperation and valuable interactions made the research experience more enriching and enjoyable.

I sincerely acknowledge the support provided by **IIT Indore** for offering the computational resources, laboratory facilities, internet access, and research infrastructure necessary for carrying out this work successfully.

Finally, I express my deepest gratitude to my family, especially my parents, for their unconditional love, encouragement, patience, and unwavering support throughout my academic journey. Their belief in me has always been my greatest source of strength and motivation.

SHUBHAM SINHA

Master of Technology

(Communication and Signal Processing)

Roll Number: 2402102006

IIT Indore

ABSTRACT

The rapid growth of next generation wireless communication systems has significantly increased the complexity of resource management in Radio Access Networks (RANs).

Modern 5G and future 6G networks require intelligent and adaptive resource allocation mechanisms capable of handling dynamic traffic conditions while satisfying diverse Quality of Service (QoS) requirements.

Conventional static and heuristic based approaches often fail to efficiently manage varying network demands, motivating the use of Artificial Intelligence and Reinforcement Learning (RL) based techniques for autonomous wireless network optimization.

This thesis presents an intelligent resource allocation framework for next generation networks using Reinforcement Learning and Large Language Models (LLMs).

Initially, a Knowledge Based Reinforcement Learning (KBRL) framework is studied for SLA aware network slicing and Physical Resource Block (PRB) allocation under multiple traffic scenarios.

The KBRL framework enables adaptive allocation decisions while balancing resource utilization and service quality constraints.

To improve scalability and decision interpretability, the work further extends the conventional KBRL framework using Large Language Models for policy learning and reasoning aware resource allocation.

Runtime trajectories and allocation decisions generated by the KBRL framework are utilized to construct a structured dataset for supervised learning and LLM fine tuning.

The developed LLM based framework learns state to allocation mappings and generates allocation decisions across different network conditions while approximating the behaviour of the expert KBRL policy.

The implementation includes scenario generation, KPI based state representation, supervised policy learning, and comparative performance evaluation using metrics such as PRB utilization, SLA violation behaviour, and prediction consistency.

Experimental results demonstrate that the proposed framework successfully approximates expert allocation policies while enabling scalable and interpretable intelligent resource management for future wireless communication systems.

The proposed work highlights the potential of integrating Reinforcement Learning and Large Language Models for AI native wireless networks and autonomous RAN optimization in future communication infrastructures.

TABLE OF CONTENTS

LIST OF FIGURES iv

LIST OF TABLES vi

1. INTRODUCTION 1

1.1 Background 1

1.2 Motivation 2

1.3 Problem Statement 3

1.4 Objectives of the Thesis 4

1.5 Contributions of the Thesis 4

2. LITERATURE REVIEW 6

2.1 Network Slicing and Intelligent RAN Architectures 6

2.2 Reinforcement Learning for Wireless Resource Allocation 7

2.3 Knowledge Based Reinforcement Learning Frameworks 8

2.4 Large Language Models for Wireless Resource Management 9

2.5 Research Gap and Motivation for Proposed Work 12

3. BASELINE KBRL FRAMEWORK AND SYSTEM MODEL 13

3.1 System Model and Network Slicing Environment 13

3.2 Knowledge Based Reinforcement Learning Framework 14

3.2.1 H-Learner 16

3.2.2 E-Learner 17

4. LLM BASED POLICY LEARNING FRAMEWORK 19

4.1 Large Language Models for Wireless Resource Allocation 19

4.2 Transformer Architecture and Self-Attention Mechanism 20

4.3 Positional Encoding and Multi-Head Attention 21

4.3.1 Positional Encoding 21

4.3.2 Multi-head Attention 22

4.4 LoRa Based Fine Tuning of Mistral-7B-Instruct 23

5. DATASET GENERATION AND DATA PREPROCESSING 25

5.1 Overview of Dataset Generation Framework 25
5.2 KBRL Trajectory Generation 26
5.3 State Representation and KPI Extraction 26
5.4 Data Preprocessing and Normalization 28
5.5 Prompt Construction and Instruction Formatting 28
5.6 Training and Validation Split 29
5.7 Final Dataset Structure 30
6. Experimental Setup 32
6.1 Experimental Environment 32
6.1.1 Hardware Configuration 32
6.1.2 Software Environment 33
6.2 LLM Training Configuration 33
6.3 Instruction Tuning and Inference Pipeline 35
6.4 Evaluation Metrics 36
6.5 Experimental Scenarios 36
6.6 Baseline Comparison Framework 37
7. Results and Performance Analysis 38
7.1 Overview of Experimental Evaluation 38
7.2 Baseline KBRL Performance Analysis 38
7.2.1 SLA Violation Behaviour 38
7.2.2 SLA Violation vs Resource Occupation Analysis 40
7.3 LLM Training Convergence Analysis 41
7.4 LLM-Based Policy Imitation Performance 43
7.5 Resource-Constrained Policy Learning 44
7.6 Generalization Across Multiple Scenarios 46
7.7 Structured Reasoning and Explainability 47

8. CONCLUSION AND FUTURE WORK 50

8.1 Conclusion 50

8.2 Limitations 51

8.3 Future Work 51

REFERENCES 53

LIST OF FIGURES

Figure 2.1: Dynamic RAN slicing and resource allocation framework 7

Figure 2.2: Comparison of optimization based, deep learning based, and LLM based resource allocation approaches 10

Figure 2.3: LLM empowered xApp framework for adaptive O-RAN resource management 11

Figure 2.4: Multi agent LLM framework for network slicing management and orchestration 11

Figure 3.1: Architecture of the proposed Kernel Based Reinforcement Learning framework 16

Figure 4.1: Transformer architecture used in Large Language Models 20

Figure 4.2: Scaled dot-product self-attention mechanism in transformer models 21

Figure 4.3: Multi-head self-attention mechanism in transformer models 22

Figure 4.4: Low-Rank Adaptation (LoRa) based parameter efficient fine-tuning framework 24

Fig7.1. Results for scenario_1. From left to right: SLA violations per stage, cumulative SLA violations, and total RBs allocated. Each plot shows the per-stage average of 30 simulation runs and its confidence interval with a 90% confidence level. 39

Fig7.2. Results for scenario_2. From left to right: SLA violations per stage, cumulative SLA violations, and total RBs allocated. Each plot shows the per-stage average of 30 simulation runs and its confidence interval with a 90% confidence level. 39

Fig7.3. Results for scenario_3. From left to right: SLA violations per stage, cumulative SLA violations, and total RBs allocated. Each plot shows the per-stage average of 30 simulation runs and its confidence interval with a 90% confidence level. 40

Figure 7.4: SLA violations per stage versus resource occupation for KBRL and model free reinforcement learning approaches 41

Figure 7.5: Training convergence behaviour of the LoRA fine tuned Mistral-7B-Instruct model during supervised instruction tuning 42

Figure 7.6: Generalization behaviour of the proposed LLM allocation framework across Scenario 0, Scenario 1, and Scenario 2. 46

Figure 7.7: Interactive transformer based PRB allocation and reasoning interpretation interface for RAN slicing evaluation 48

LIST OF TABLES

Table 5.1: Feature representation for 6-feature eMBB slice states 27

Table 5.2: Feature representation for 10-feature eMBB slice states 27

Table 5.3: Feature representation for mMTC slice states 27

Table 5.4: Example instruction tuning sample generated from KBRL allocation trajectories 29

Table 5.5: Final dataset configuration used for instruction tuning 30

Table 6.1: Local development system configuration 32

Table 6.2: Remote GPU workstation configuration 32

Table 6.3: Software libraries and frameworks used during implementation 33

Table 6.4: Hyperparameter configuration used for LoRA based instruction tuning 34

Table 7.1: Allocation prediction performance of the LoRA fine tuned LLM policy imitation framework 43

Table 7.2: PRB allocation comparison between KBRL and LLM generated allocation outputs 44

Table 7.3: Performance evaluation of the resource constrained LLM allocation framework 45

Table 7.4: Structured reasoning output generated by the proposed LLM allocation framework 47

Table 7.5: Structured reasoning severity classification used during allocation evaluation 47

CHAPTER 1

INTRODUCTION

1.1 Background

The increasing demand for high data rate communication, ultra reliable low latency services, and massive device connectivity has significantly transformed the design requirements of modern wireless communication systems. Fifth Generation (5G) and emerging Sixth Generation (6G) networks are expected to support heterogeneous applications including enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communication (URLLC), and massive Machine Type Communication (mMTC). These applications exhibit diverse Quality of Service (QoS) requirements in terms of throughput, latency, reliability, and resource utilization, thereby introducing substantial challenges in efficient Radio Access Network (RAN) resource management.

To address the increasing heterogeneity of wireless services, network slicing has emerged as a key enabling technology in next generation communication systems. Network slicing enables multiple logical networks to coexist over a shared physical infrastructure while providing customized resource allocation and Service Level Agreement (SLA) guarantees for different service categories. Efficient allocation of Physical Resource Blocks (PRBs) and radio resources among multiple network slices is therefore critical for maintaining network performance and user satisfaction under dynamic traffic conditions. Conventional heuristic and rule-based allocation approach often face limitations in adapting to rapidly changing network states and fluctuating traffic demands.

In recent years, Reinforcement Learning (RL) based techniques have gained considerable attention for autonomous wireless network optimization due to their ability to learn allocation policies directly from interactions with the environment. RL driven frameworks enable adaptive resource allocation by continuously observing network states and optimizing long term performance objectives. Among these approaches, Knowledge Based Reinforcement Learning (KBRL) frameworks have demonstrated strong capabilities in SLA aware network slicing and resource orchestration. The KBRL framework combines online reinforcement learning with knowledge driven policy learning to reduce unsafe exploration while improving allocation efficiency and service reliability.

Although RL based approaches provide adaptive and autonomous decision-making capabilities, practical deployment of such systems in large scale wireless networks remains challenging due to issues related to scalability, computational complexity, and interpretability of learned policies. At the same

time, recent advancements in Large Language Models (LLMs) have demonstrated strong reasoning and pattern learning capabilities across multiple domains. Emerging research in AI native wireless networking and Open RAN architectures has highlighted the potential of integrating LLMs with intelligent network controllers for dynamic resource allocation and autonomous network orchestration.

Recent studies have explored the use of LLM driven orchestration frameworks and intelligent xApps for network slicing, RAN control, and intent based wireless resource management. These frameworks aim to improve network automation and decision interpretability by enabling AI assisted policy generation and adaptive control mechanisms.

Motivated by these developments, this thesis investigates intelligent resource allocation policies for next generation wireless networks using Reinforcement Learning and Large Language Models. The work focuses on studying a Knowledge Based Reinforcement Learning framework for network slicing and PRB allocation and further extends it through LLM based policy learning for scalable and reasoning aware RAN resource optimization.

1.2 Motivation

Existing Reinforcement Learning based resource allocation frameworks primarily focus on maximizing network performance metrics such as throughput, resource utilization, and SLA satisfaction. Although these approaches demonstrate strong adaptability in dynamic wireless environments, the learned policies often behave as black box decision systems where the reasoning behind allocation decisions is not explicitly interpretable. In practical Radio Access Network (RAN) management scenarios, understanding why a particular allocation action is selected under a given network state is important for policy validation, debugging, reliability analysis, and operator trust.

Knowledge Based Reinforcement Learning (KBRL) frameworks improve the stability and efficiency of wireless resource allocation by incorporating knowledge driven policy learning and reducing unsafe exploration during online optimization. However, despite improved allocation performance, the generated policies remain difficult to interpret in highly dynamic network conditions. As wireless systems evolve towards AI native and autonomous network management architectures, there is an increasing requirement for resource allocation frameworks that are not only adaptive and scalable but also capable of providing understandable and reasoning aware decision support.

Recent advances in transformer architectures and Large Language Models (LLMs) have demonstrated strong capabilities in contextual learning, sequential reasoning, and structured response generation. Transformer based models can

learn complex relationships from expert generated data and producing human understandable outputs based on contextual information. These capabilities make LLMs particularly interesting for approximating and interpreting allocation behaviour generated by expert Reinforcement Learning frameworks.

Motivated by these observations, this thesis explores the use of Large Language Models for learning and approximating expert KBRL allocation policies in wireless network slicing scenarios. Instead of directly replacing Reinforcement Learning based optimization, the objective is to investigate whether transformer-based models can learn state to allocation mappings from KBRL generated trajectories while also enabling more interpretable and reasoning aware resource allocation decisions. The proposed framework therefore aims to bridge adaptive Reinforcement Learning based optimization with scalable and human understandable policy learning for intelligent RAN resource management.

1.3 Problem Statement

Current Reinforcement Learning based resource allocation frameworks for wireless communication systems primarily optimize network performance metrics such as throughput, latency, and resource utilization. Although these approaches demonstrate strong adaptability in dynamic Radio Access Network (RAN) environments, the learned allocation policies are generally represented as complex mathematical or neural decision functions that are difficult to interpret from an operational perspective. As a result, understanding the relationship between network conditions and generated allocation actions becomes challenging, particularly in multi slice and heterogeneous traffic scenarios.

In intelligent RAN management, allocation decisions are often required to satisfy Service Level Agreement (SLA) constraints across multiple network slices while maintaining efficient utilization of limited Physical Resource Blocks (PRBs). Existing Knowledge Based Reinforcement Learning (KBRL) frameworks improve policy stability and reduce unsafe exploration; however, the generated policies still lack explainability and human understandable reasoning support. Consequently, network operators and intelligent controllers have limited visibility into why specific allocation actions are selected under network states.

The increasing complexity of AI native wireless networks further amplifies this challenge. Future autonomous network architectures are expected to support scalable, interpretable, and reasoning aware decision systems capable of assisting real time network orchestration and policy validation. While transformer based Large Language Models (LLMs) have recently demonstrated strong contextual learning and sequence modelling capabilities, their

application for approximating and interpreting Reinforcement Learning based wireless allocation policies remains insufficiently explored.

Therefore, there exists a need for a framework capable of learning expert Reinforcement Learning allocation behaviour while also improving interpretability and reasoning awareness in wireless resource management. This thesis addresses this problem by investigating whether transformer based Large Language Models can learn and approximate KBRL generated allocation policies for intelligent network slicing and PRB allocation in next generation wireless communication systems.

1.4 Objectives of the Thesis

The primary objectives of this thesis are summarized as follows:

• To study Reinforcement Learning based resource allocation techniques for intelligent Radio Access Network (RAN) management and network slicing.

• To analyse the Knowledge Based Reinforcement Learning (KBRL) framework for Service Level Agreement (SLA) aware Physical Resource Block (PRB) allocation under varying traffic conditions.

• To generate structured datasets from KBRL based allocation trajectories for supervised policy learning.

• To investigate the capability of transformer based Large Language Models (LLMs) in learning and approximating expert Reinforcement Learning allocation behaviour.

• To explore interpretable and reasoning aware resource allocation using LLM driven policy generation.

• To evaluate and compare the performance of the proposed framework under multiple network scenarios using relevant network performance metrics.

1.5 Contributions of the Thesis

The work presented in this thesis focuses on extending intelligent wireless resource allocation using Reinforcement Learning and Large Language Models. The major contributions of the thesis are summarized as follows:

• Investigated the Knowledge Based Reinforcement Learning (KBRL) framework for SLA aware network slicing and Physical Resource Block (PRB) allocation in dynamic wireless network environments.

• Developed a simulation and evaluation pipeline for generating resource allocation trajectories and network state information under multiple traffic scenarios.

• Constructed structured state action datasets from KBRL generated allocation policies for supervised policy learning and transformer-based modelling.

• Explored the capability of transformer based Large Language Models to learn and approximate expert Reinforcement Learning allocation behaviour from KBRL generated trajectories.

• Studied the potential of LLM driven policy generation for improving interpretability and reasoning awareness in intelligent Radio Access Network (RAN) resource management.

• Performed comparative analysis between KBRL based allocation behaviour and LLM generated allocation policies using network performance and allocation consistency metrics.

CHAPTER 2

LITERATURE REVIEW

2.1 Network Slicing and Intelligent RAN Architectures

Network slicing has emerged as one of the key enabling technologies for Fifth Generation (5G) and beyond wireless communication systems. Unlike conventional cellular architectures where radio resources are shared using static and uniform allocation policies, network slicing enables the creation of multiple logical networks over a common physical infrastructure while supporting heterogeneous service requirements. Each slice can be configured independently according to specific Quality of Service (QoS) constraints, latency requirements, throughput demands, and reliability objectives. This flexibility allows network operators to efficiently support diverse applications such as enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communication (URLLC), and massive Machine Type Communication (mMTC) within the same Radio Access Network (RAN) infrastructure.

In sliced RAN architectures, efficient allocation of radio resources among multiple slices becomes a critical optimization problem. Physical Resource Blocks (PRBs) or Resource Blocks (RBs) must be dynamically distributed among slices while ensuring Service Level Agreement (SLA) compliance and efficient spectrum utilization. Since wireless traffic conditions vary continuously with user mobility, channel quality, and service demand fluctuations, static allocation mechanisms often fail to provide efficient long term resource management. As a result, intelligent and adaptive RAN control frameworks have gained considerable attention for next generation wireless systems.

Recent Open RAN and intelligent RAN architectures further extend this concept by introducing programmable and software driven control mechanisms capable of supporting autonomous network optimization. These architectures enable intelligent controllers to continuously monitor network Key Performance Indicators (KPIs), observe slice level traffic conditions, and dynamically adjust resource allocation policies in real time. Such adaptive resource management frameworks are particularly important in dense and heterogeneous wireless environments where network conditions change rapidly over time.

Several recent studies have investigated dynamic RAN slicing frameworks for SLA aware resource allocation using intelligent control policies. These frameworks typically model wireless resource allocation as a sequential decision-making problem in which the controller observes the current network state and allocates available RBs among multiple slices to maximize long term

network performance while minimizing SLA violations. The use of intelligent slicing architectures therefore provides an important foundation for applying machine learning and Reinforcement Learning based optimization techniques in next generation wireless communication systems.

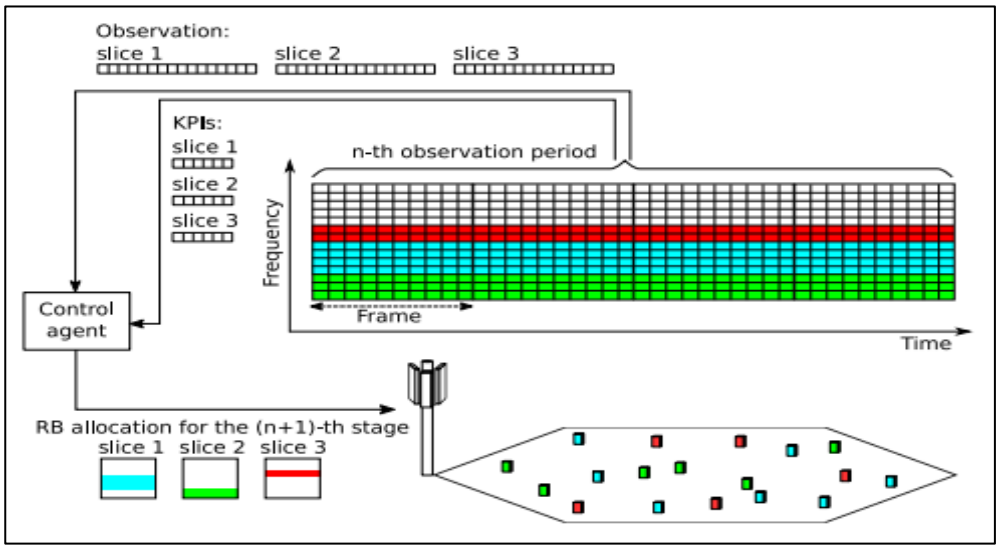


Figure 2.1: Dynamic RAN slicing and resource allocation framework

2.2 Reinforcement Learning for Wireless Resource Allocation

The dynamic and stochastic nature of wireless communication environments makes efficient radio resource allocation a challenging optimization problem. In practical Radio Access Network (RAN) deployments, traffic demand, channel conditions, user mobility, and slice level Quality of Service (QoS) requirements continuously vary over time. Conventional optimization and heuristic based allocation approach often require predefined traffic assumptions and static decision policies, which limit their adaptability in highly dynamic wireless scenarios.

To address these limitations, Reinforcement Learning (RL) based techniques have gained significant attention for autonomous wireless resource management. RL frameworks model resource allocation as a sequential decision-making problem in which an intelligent agent interacts with the wireless environment, observes network states, and learns allocation policies through continuous feedback. Unlike supervised learning approaches that require labelled datasets, RL agents improve their decision policies using reward driven interactions with the environment, making them suitable for adaptive and online wireless optimization.

Several studies have explored the application of RL techniques for dynamic spectrum allocation, scheduling, power control, traffic management, and network slicing in wireless communication systems. Model free RL approaches

such as Q Learning and Deep Q Networks (DQN) have demonstrated the capability to learn adaptive allocation strategies under varying network conditions. More advanced policy optimization techniques including Deep Deterministic Policy Gradient (DDPG), Trust Region Policy Optimization (TRPO), and Proximal Policy Optimization (PPO) have also been investigated for continuous and large-scale wireless optimization problems.

Despite their advantages, conventional model free RL techniques often require large amounts of interaction data and extensive exploration before converging to stable allocation policies. In wireless communication systems, unsafe exploration during training may lead to Service Level Agreement (SLA) violations, inefficient spectrum utilization, and unstable network performance. Furthermore, large state action spaces and continuously varying network conditions significantly increase the complexity of online policy learning.

To improve sample efficiency and policy stability, recent research has investigated model based and knowledge driven Reinforcement Learning frameworks for intelligent wireless resource allocation. These approaches incorporate prior knowledge, environment modelling, or policy guidance mechanisms to improve learning efficiency and reduce unsafe exploration during resource optimization. Such developments have motivated the design of Knowledge Based Reinforcement Learning frameworks for SLA aware RAN slicing and adaptive resource management in next generation wireless communication systems.

2.3 Knowledge Based Reinforcement Learning Frameworks

Most existing Reinforcement Learning based resource allocation approaches for network slicing rely on model free learning techniques in which the agent learns allocation policies directly through repeated interaction with the environment. Although such approaches can adapt to varying traffic conditions, their practical deployment in operating wireless networks remains challenging due to low sample efficiency and extensive exploration requirements. During the learning process, the agent may repeatedly select inefficient allocation actions before converging to a stable policy, leading to temporary degradation in network performance and increased SLA violation rates.

To address these limitations, recent research has explored model-based Reinforcement Learning strategies for wireless resource allocation. In the work proposed by Alcaraz et al., the resource allocation problem is formulated as a constrained sequential decision process in which the controller allocates Resource Blocks (RBs) among multiple RAN slices while minimizing SLA violations and improving spectrum efficiency. Instead of learning value functions directly, the framework estimates whether a particular allocation

action can satisfy the SLA constraints associated with a slice under a given network observation.

The proposed framework incorporates kernel based online learning for classification using the Projection algorithm. The learning process maintains classifiers corresponding to slice level allocation behaviour and continuously updates them using online observations collected from the network environment. A key feature of the framework is the introduction of sample augmentation and self-assessment mechanisms that improve learning efficiency while reducing unsafe exploration during policy updates. The controller further uses prediction confidence estimates and security margins to select reliable allocation actions under varying traffic conditions.

Compared to conventional model free RL approaches such as DQN and PPO, the KBRL framework focuses on online learning efficiency, reduced computational overhead, and controlled SLA violation behaviour during training. These characteristics make model based KBRL approaches particularly relevant for practical deployment in intelligent RAN slicing environments where allocation stability and operational reliability are critical requirements.

2.4 Large Language Models for Wireless Resource Management

Recent advances in Large Language Models (LLMs) have motivated their application beyond conventional natural language processing tasks toward intelligent wireless network management and resource orchestration. Unlike task specific deep learning models that require dedicated training pipelines for individual optimization problems, LLMs provide a more generalized reasoning framework capable of interpreting complex system states, generating adaptive control decisions, and supporting intent driven network management. This capability has recently attracted significant interest in next generation wireless communication systems, particularly in the context of Radio Access Network (RAN) slicing, Open RAN (O-RAN) orchestration, and autonomous resource management.

Recent studies have investigated the feasibility of applying LLMs to wireless resource allocation problems involving spectrum efficiency, energy efficiency, and adaptive resource provisioning. Lee and Park proposed an LLM empowered resource allocation framework in which the language model acts as a

generalized optimization engine for wireless communication systems. The framework highlights the capability of LLMs to perform resource allocation without requiring task specific model retraining while supporting flexible adaptation to varying communication objectives and wireless environments. The study further discusses the advantages of multimodal reasoning and sequential decision generation for future AI native wireless system.

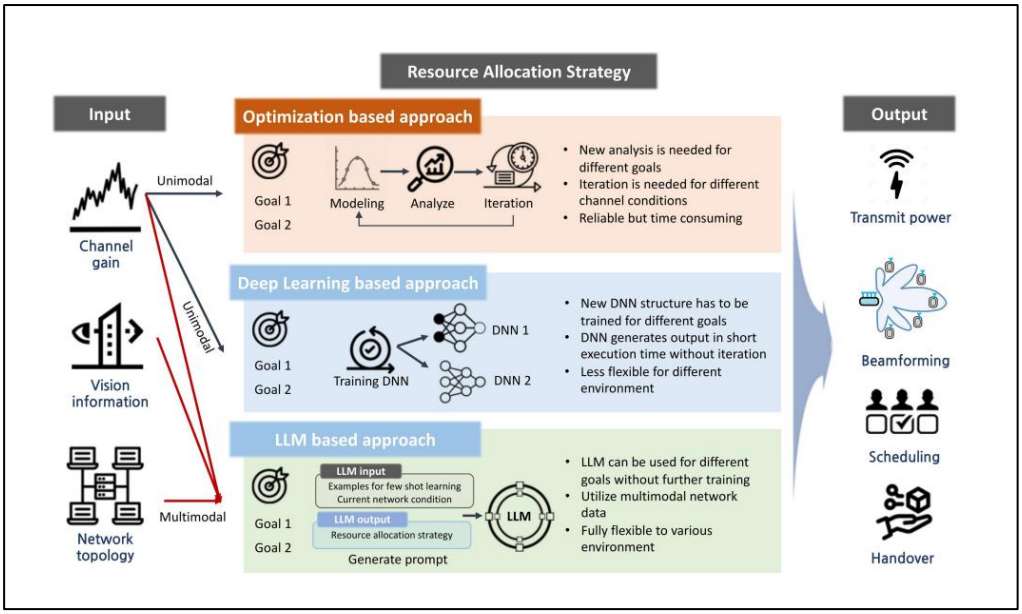


Figure 2.2: Comparison of optimization based, deep learning based, and LLM based resource allocation approaches

The integration of LLMs into Open RAN architectures has also emerged as an important research direction for intelligent network orchestration. In O-RAN systems, the disaggregation of network functions across distributed units, centralized units, and RAN Intelligent Controllers (RICs) introduces additional complexity in dynamic slice management and real time resource coordination. To address these challenges, the paper proposed an LLM-xApp, a LLM empowered xApp framework for adaptive radio resource management in O-RAN environments.

The proposed framework utilizes LLM agents to interpret user Quality of Service (QoS) requirements and dynamically optimize Physical Resource Block (PRB) allocation across multiple network slices.

Experimental evaluation on the OpenAI Cellular platform demonstrated improvements in slice reliability and adaptive resource provisioning under varying traffic conditions.

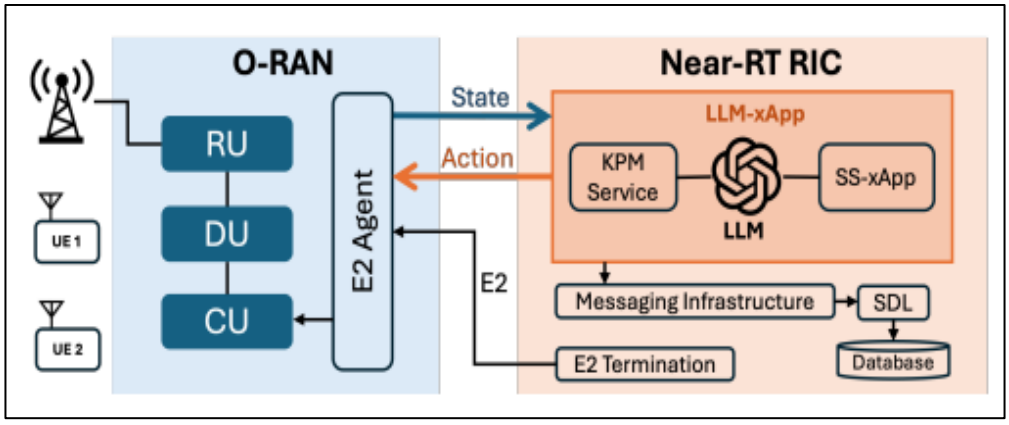


Figure 2.3: LLM empowered xApp framework for adaptive O-RAN resource management

Beyond resource allocation, recent studies have further explored the role of LLMs in network slicing management and orchestration.

The proposal of a multi agent LLM based framework for end-to-end network slice orchestration across heterogeneous and multi administrative domain environments is one of the most researched surveys in the LLM domain.

The proposed architecture utilizes collaborative LLM agents for translating user intents into technical slicing requirements, coordinating slice lifecycle management, and enabling dynamic resource orchestration across distributed network domains.

The framework also highlights the potential of intent driven and reasoning aware network management for future self-driving wireless networks.

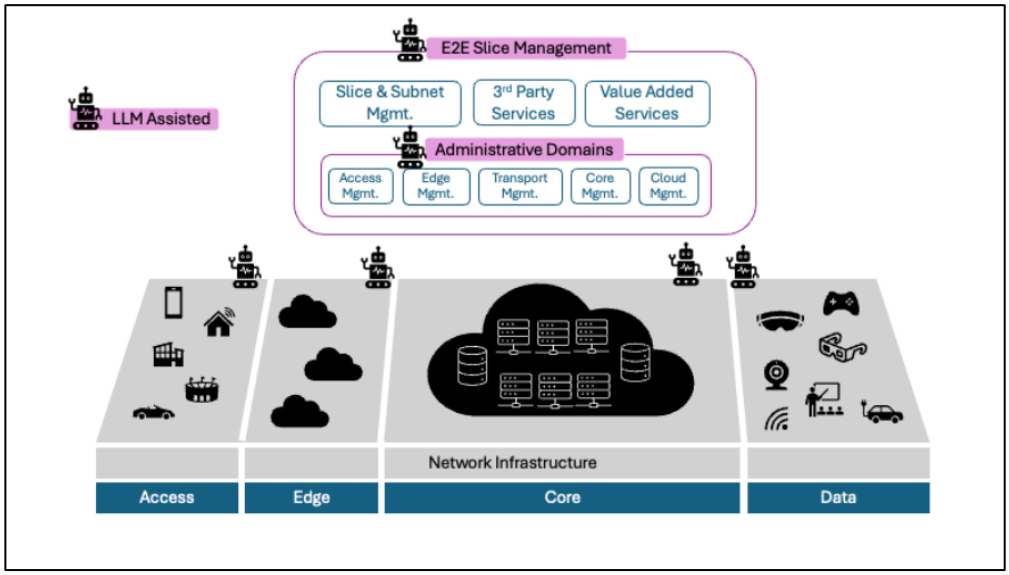


Figure 2.4: Multi agent LLM framework for network slicing management and orchestration

The growing integration of LLMs into wireless communication systems demonstrates a transition from conventional rule based and task specific optimization approaches toward more adaptive, reasoning aware, and intent driven network management frameworks. Although existing studies have demonstrated the feasibility of LLM assisted orchestration and resource allocation, the application of transformer-based models for learning and approximating Reinforcement Learning driven allocation behaviour in intelligent RAN slicing environments remains relatively unexplored. This motivates further investigation of LLM driven policy learning and interpretable resource allocation for next generation AI native wireless communication systems.

2.5 Research Gap and Motivation for Proposed Work

Based on the existing literature, the following research gaps are identified in current wireless resource allocation and network slicing frameworks:

• Existing Reinforcement Learning based resource allocation approaches primarily focus on throughput optimization and SLA satisfaction, while limited attention is given to interpretability and reasoning aware allocation behaviour.

• Model free Reinforcement Learning techniques often require extensive exploration and large-scale interaction data, making stable policy learning difficult in dynamic wireless network environments.

• Although Knowledge Based Reinforcement Learning (KBRL) frameworks improve allocation stability and learning efficiency, the generated policies still behave largely as black box decision systems from the perspective of network operators.

• Recent Large Language Model based networking frameworks mainly focus on intent translation, orchestration, and high-level network automation rather than direct learning of wireless resource allocation policies.

• Existing studies on LLM assisted O-RAN management do not sufficiently investigate whether transformer-based models can learn and approximate expert Reinforcement Learning allocation behaviour from network trajectories.

• The application of LLMs for interpreting KBRL generated allocation decisions and supporting reasoning aware RAN resource management remains relatively unexplored in current literature.

Motivated by these observations, this thesis investigates the integration of Knowledge Based Reinforcement Learning and transformer based Large Language Models for intelligent and interpretable wireless resource allocation in network slicing environments.

CHAPTER 3

BASELINE KBRL FRAMEWORK AND SYSTEM MODEL

3.1 System Model and Network Slicing Environment

The considered wireless communication system consists of a sliced Radio Access Network (RAN) environment in which multiple network slices share a finite pool of available Resource Blocks (RBs). Each slice is associated with specific Quality of Service (QoS) and Service Level Agreement (SLA) requirements depending on the supported service category. The controller dynamically allocates RBs among slices according to the observed network state and traffic conditions at every scheduling interval.

Let the set of network slices be represented as:

S = {1,2, 3,...,N} where N denotes the total number of active slices within the RAN environment.

At every scheduling interval t, the controller observes the network state consisting of traffic demand, queue conditions, resource utilization, and SLA related information.

The observed state can be represented as $s_t = q_t, u_t, rho_t$ where q_t represents queue information, $u_t$denotes resource utilization, and rho_t corresponds to traffic load observations at time t.

Based on the observed state, the controller selects an allocation action from the feasible allocation space.

The action determines the number of Resource Blocks assigned to each network slice and can be represented as $a_t = r_1(t), r_2(t), \dots, r_N(t)$ where $r_i(t)$ denotes the number of RBs allocated to slice i during scheduling interval t.

The objective of the resource allocation framework is to maximize long term allocation efficiency while minimizing SLA violation behaviour across slices.

During the learning process, the agent may repeatedly select inefficient allocation actions before converging to a stable policy, leading to temporary degradation in network performance and increased SLA violation rates.

The optimization problem can therefore be represented as:

$$\sum_{t=0}^{T} \quad R(s_t, a_t)$$

s.t the resource allocation constraint

$$\sum_{i=1}^{N} \quad r_i(t) \leq R_{total};$$

R_{total} = the total number of available RBs within the system.

The reward function can therefore be expressed as:

$$R(s_t, a_t) = \alpha * U_t - \beta * V_t$$

α , β are weighting parameters used to balance allocation efficiency and SLA violation behaviour.

The reward associated with a selected allocation action depends on both network utility and SLA satisfaction performance.

U_t=Allocation utility

V_t=SLA violation penalty at scheduling interval t.

The overall resource allocation process is therefore modelled as a constrained sequential decision-making problem in which the controller continuously interacts with the wireless environment and updates allocation behaviour according to observed system performance.

This formulation provides the basis for applying Knowledge Based Reinforcement Learning for adaptive and SLA aware resource management in dynamic RAN slicing environments.

3.2 Knowledge Based Reinforcement Learning Framework

The proposed Kernel Based Reinforcement Learning (KBRL) framework adopts a model-based Reinforcement Learning approach for dynamic radio resource allocation in RAN slicing environments.

Unlike conventional model free RL approaches, the proposed framework learns a predictive model capable of identifying SLA compliant allocation actions while reducing unsafe exploration during online learning.

The KBRL architecture consists of three major operational modules namely the controller, the H-learner, and the E-learner.

At every decision stage, the controller observes the current slice conditions and determines the RB allocation policy according to the learned system model.

The H-learner performs online classification for SLA prediction, whereas the E-learner estimates the prediction reliability associated with the learned classifier.

The interaction between the controller and the wireless environment is represented using the observation action pair (s_n, a_n) where s_n denotes the observed system state and a_n represents the RB allocation action selected at decision stage n.

`For each network slice i` $\in s_n^{(i)}$K, the controller receives the slice observation vector $s_n^{(i)}$ which contains traffic statistics, queue measurements, resource occupation information, and SLA related indicators collected during the observation interval.

The overall allocation policy is represented by $\pi: s_n \rightarrow a_n$ where the policy π maps the observed network state to the corresponding RB allocation action.

The H-learner operates as an online kernel-based classifier responsible for predicting whether the SLA of a slice will be fulfilled for a selected allocation action.

The classifier prediction function is represented as $\widehat{y_n} = h_n(x_n)$ where $\widehat{y_n}$ denotes the predicted SLA fulfilment outcome for the observed state action pair.

The E-learner module estimates the classification reliability associated with the H-learner predictions.

The prediction error probability for slice i is represented as $e^{(i)}\left(h_n^{(i)}, s_n^{(i)}, a_n^{(i)}\right)$ which corresponds to the probability of obtaining an SLA violation even when the classifier predicts successful SLA satisfaction.

The controller utilizes the outputs of both the H-learner and the E-learner to determine reliable allocation actions with reduced SLA violation probability.

The resulting framework therefore enables adaptive and SLA aware RB allocation while maintaining improved learning efficiency during online operation.

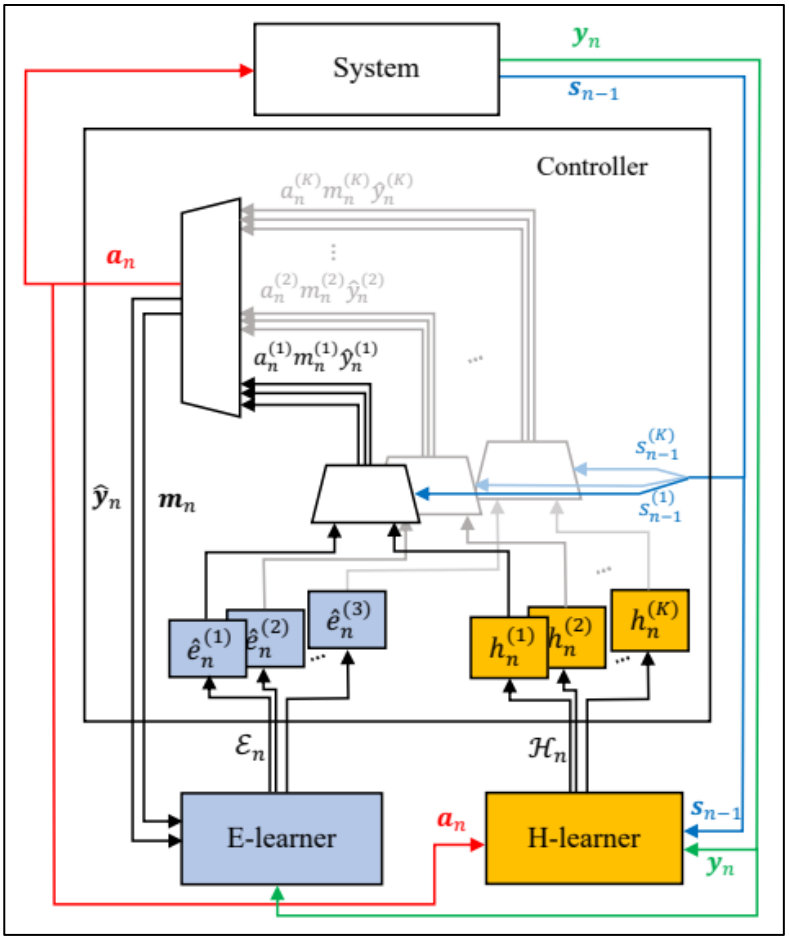


Figure 3.1: Architecture of the proposed Kernel Based Reinforcement Learning framework

3.2.1 H-Learner

The H-learner module operates as an online kernel-based classifier responsible for predicting whether the Service Level Agreement (SLA) associated with a network slice will be satisfied for a selected Resource Block (RB) allocation action. The primary objective of the H-learner is to identify SLA compliant and non-compliant allocation regions during online interaction with the wireless environment.

For each network slice $i \in K$, the H−learner receives the observation action pair $\left(s_n^{(i)}, a_n^{(i)}\right)$ where $s_n^{(i)}$ denotes the slice observation vector and $a_n^{(i)}$ represents the RB allocation action corresponding to slice i at decision stage n.

The prediction model generated by the H-learner is represented using a kernel-based hypothesis function

$$h_n(x) = \Sigma \alpha_i \kappa(x_i, x)$$

where $X_N$denotes the support vector set generated during online learning, α_i represents the classifier coefficient associated with support sample x_i, and $\kappa(x_i, x)$ denotes the kernel similarity function between the observed input sample x and the stored support vector x_i

The H-learner performs online binary classification in which the predicted allocation label is represented as:

$$\widehat{y_n} = sign\big(h_n(x_n)\big)$$

where $\widehat{y_n}$=1 denotes an SLA violation prediction and $\widehat{y_n}$=0 indicates successful SLA satisfaction for the selected allocation action.

During online operation, the H-learner continuously updates the classifier using newly observed allocation samples obtained from the wireless environment. Whenever a prediction error occurs, the classifier parameters are updated using the Projectron online learning mechanism. The Projectron algorithm performs support vector projection to control the growth of the support vector space while maintaining classification accuracy.

The hinge loss associated with the online classification process is represented as:

$$l\big((x_n, y_n); h_n\big) = \big(0{,}1 - y_n h_n(x_n)\big)$$

where y_n denotes the true allocation label obtained after observing the actual SLA outcome.

To improve learning efficiency, the H-learner incorporates a sample augmentation mechanism based on SLA monotonicity assumptions. If an SLA is satisfied for a given allocation level, then higher RB allocations are also considered SLA compliant. Similarly, if an SLA violation occurs for a selected allocation level, lower RB allocations are also treated as non-compliant allocation samples.

The online learning process therefore enables the H-learner to progressively identify feasible allocation regions while reducing unsafe exploration during dynamic resource allocation.

3.2.2 E-Learner

The E-learner module is responsible for estimating the prediction reliability associated with the H-learner classifier during online resource allocation. The objective of the E-learner is to reduce unsafe exploration by estimating the probability of SLA violation for a selected allocation action.

For each network slice i ∈ K, the prediction error probability is represented as:

$$e^{(i)}\left(h_n^{(i)}, s_n^{(i)}, a_n^{(i)}\right)$$

where $e^{(i)}(.)$ denotes the probability of obtaining an SLA violation for slice i even when the H-learner predicts successful SLA satisfaction for the selected allocation action.

The E-learner further introduces a security margin associated with each allocation action to estimate the reliability of the generated prediction. The security margin is defined as

$$m_n^{(i)} = a_n^{(i)} - a : h_n^{(i)}\left(s_n^{(i)}, a\right) = 0$$

where $m_n^{(i)}$ represents the distance between the selected allocation action and the minimum RB allocation satisfying the SLA constraints. Larger security margins correspond to lower prediction error probability and more reliable allocation decisions.

The reliability estimator is continuously updated using the observed allocation outcomes during online interaction with the wireless environment.

The estimator update equation is represented as:

$$\hat{\mathrm{e}}_n + 1^{(i)}(m) = (1-\beta)\hat{\mathrm{e}}_n^{(i)}(m) + \beta y_n^{(i)}$$

where β denotes the learning rate parameter, $\hat{\mathrm{e}}_n^{(i)}(m)$ represents the estimated prediction error probability associated with security margin m, and $y_n^{(i)}$denotes the observed allocation outcome for slice i.

The E-learner therefore assists the controller in selecting allocation actions with lower SLA violation probability while improving allocation reliability and reducing unstable exploration during online policy learning.

CHAPTER 4

LLM BASED POLICY LEARNING FRAMEWORK

4.1 Large Language Models for Wireless Resource Allocation

Recent advancements in transformer based Large Language Models (LLMs) have demonstrated strong capabilities in contextual learning, sequential reasoning, and structured decision generation across multiple application domains.

Transformer architectures process input sequences using self-attention mechanisms that enable the model to capture long range contextual dependencies without relying on recurrent processing structures. This capability has significantly improved the performance of modern language models in tasks involving sequence modelling, reasoning, and generative decision making.

In wireless communication systems, intelligent Radio Access Network (RAN) management requires adaptive resource allocation under dynamic traffic conditions, heterogeneous Quality of Service (QoS) constraints, and varying Service Level Agreement (SLA) requirements. Conventional Reinforcement Learning based resource allocation frameworks provide adaptive optimization capabilities; however, these approaches often require extensive online interaction and generate policies that are difficult to interpret operationally.

The proposed framework investigates the use of transformer based Large Language Models for learning and approximating expert resource allocation behaviour generated by the Knowledge Based Reinforcement Learning (KBRL) framework.

Instead of performing direct online exploration, the proposed model learns state to allocation mappings using structured allocation trajectories generated from the expert KBRL controller.

The proposed approach focuses on instruction tuning and parameter efficient fine tuning of the Mistral-7B-Instruct model using LoRa based adaptation techniques.

The generated training samples contain network state observations, slice level KPI measurements, Resource Block (RB) allocation actions, and SLA related indicators collected during KBRL based simulation under multiple traffic scenarios.

The objective of the proposed framework is to approximate expert allocation behaviour while enabling scalable and reasoning aware resource allocation for future AI native wireless communication systems.

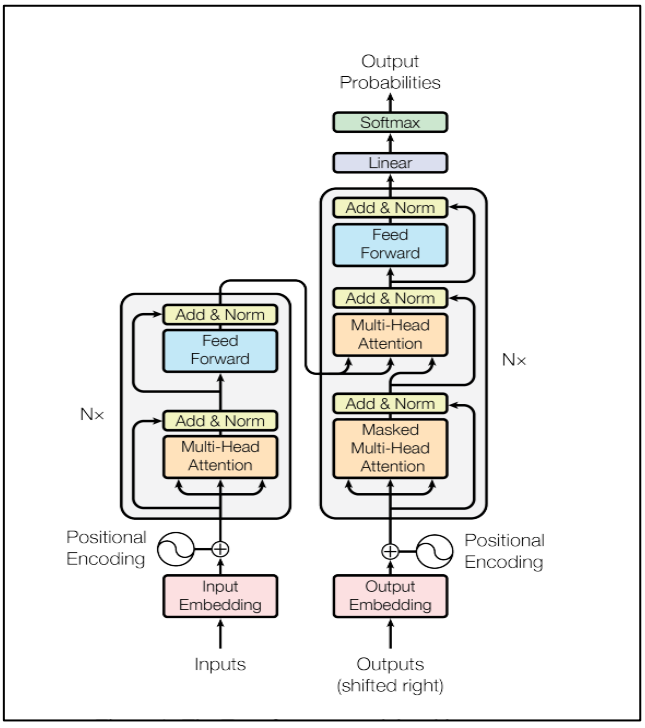


Figure 4.1: Transformer architecture used in Large Language Models

4.2 Transformer Architecture and Self-Attention Mechanism

Transformer architectures process sequential data using stacked attention-based learning blocks that operate on tokenized input representations. Unlike recurrent neural networks, transformers process all tokens in parallel using self-attention mechanisms, thereby improving computational efficiency and long-range dependency modelling.

For an input sequence represented by token embeddings X, the transformer generates Query (Q), Key (K), and Value (V) matrices using learnable projection matrices. The corresponding transformations are represented as:

$$Q = XW_Q$$

$$K = XW_K$$

$$V = XW_V$$

where W_Q, W_K, and W_V denote the learnable parameter matrices associated with query, key, and value projections respectively.

The scaled dot-product attention mechanism is computed using:

$$Attention(Q, K, V) = soft\big((QK^T)/sqrt(d_k)\big)\, V$$

where d_k is the dimensionality of the key vectors.

The attention mechanism enables the model to estimate contextual dependencies among input tokens by computing pairwise similarity scores between query and key representations.

Modern Large Language Models further utilize multi-head self-attention in which multiple attention heads independently learn different contextual relationships from the input sequence. The outputs generated by all attention heads are concatenated and projected to form the final contextual representation.

In the proposed framework, transformer based contextual learning enables the model to capture relationships between network states, traffic conditions, slice level KPI observations, and corresponding RB allocation decisions generated by the expert KBRL controller.

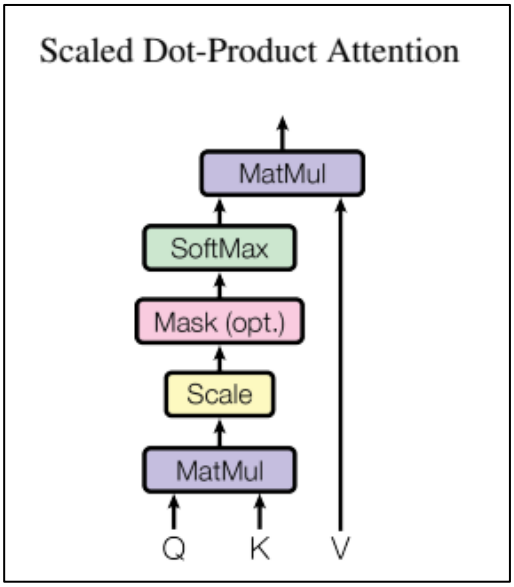


Figure 4.2: Scaled dot-product self-attention mechanism in transformer models

4.3 Positional Encoding and Multi-Head Attention

4.3.1 Positional Encoding

Since transformer architectures process all input tokens in parallel, positional information must be explicitly incorporated into the token representations to preserve sequential relationships within the input sequence. This is achieved using positional encoding mechanisms that inject token position information into the embedding space.

For a token position pos and embedding dimension i, the sinusoidal positional encoding is represented as:

$$PE(pos, 2i) = sin\ sin\left(pos/10000^{(2i/d_{model})}\right)$$

$$PE(pos, 2i+1) = cos\ cos\left(pos/10000^{(2i/d_{model})}\right)$$

where d_{model} denotes the embedding dimension of the transformer model. The positional encoding mechanism enables the transformer to capture token ordering and sequential context during attention computation.

4.3.2 Multi-head Attention

Modern Large Language Models further utilize multi-head self-attention mechanisms in which multiple attention heads independently learn different contextual relationships from the input sequence. Instead of computing a single attention representation, the transformer computes multiple parallel attention outputs represented as:

$$head_i = Attention(Q_i, K_i, V_i)$$

The outputs generated by all attention heads are concatenated and projected to produce the final contextual representation according to:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W_O$$

where h denotes the number of attention heads and W_O represents the output projection matrix.

The use of multi-head attention enables the transformer to simultaneously capture multiple contextual dependencies and semantic relationships from the input sequence. In the proposed framework, these mechanisms assist the model in learning complex relationships between network observations, KPI measurements, traffic conditions, and corresponding Resource Block allocation actions generated by the KBRL framework.

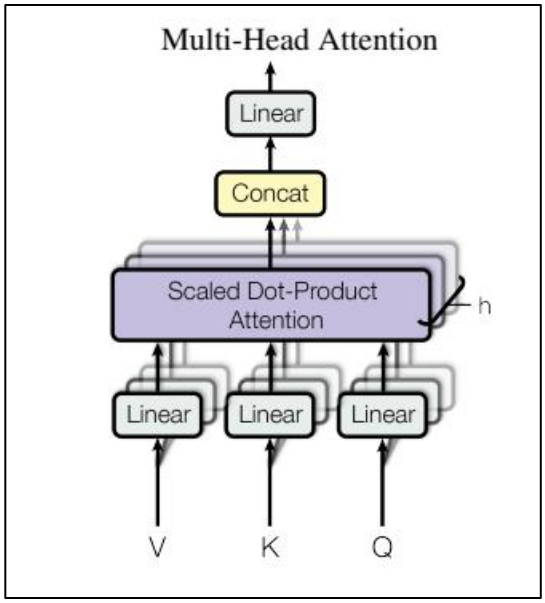


Figure 4.3: Multi-head self-attention mechanism in transformer models

4.4 LoRa Based Fine Tuning of Mistral-7B-Instruct

Large Language Models typically contain billions of trainable parameters, making full parameter fine tuning computationally expensive for domain specific learning tasks. To reduce training complexity and memory requirements, the proposed framework utilizes Low Rank Adaptation (LoRA) based parameter efficient fine tuning for adapting the Mistral-7B-Instruct model to wireless resource allocation tasks.

LoRa introduces trainable low rank matrices into selected transformer layers while keeping most pretrained model parameters frozen. Instead of directly updating the original parameter matrix W, the adapted parameter update is represented as:

$$W = W_0 + \Delta W$$

where W_0 denotes the pretrained parameter matrix and ΔW corresponds to the low rank adaptation update.

The low rank decomposition is represented as:

$$\Delta W = BA$$

where B and A denote trainable low rank matrices with significantly smaller dimensionality compared to the original parameter space. This approach substantially reduces the number of trainable parameters during fine tuning while preserving the general reasoning capabilities of the pretrained language model.

The proposed framework performs supervised instruction tuning of the Mistral-7B-Instruct model using KBRL generated allocation trajectories. During training, structured prompts containing network observations and KPI information are provided as input, whereas the target output corresponds to the RB allocation action generated by the expert KBRL controller.

The resulting fine-tuned model therefore learns to approximate expert allocation behaviour while enabling scalable and interpretable allocation decision generation for intelligent RAN slicing environments.

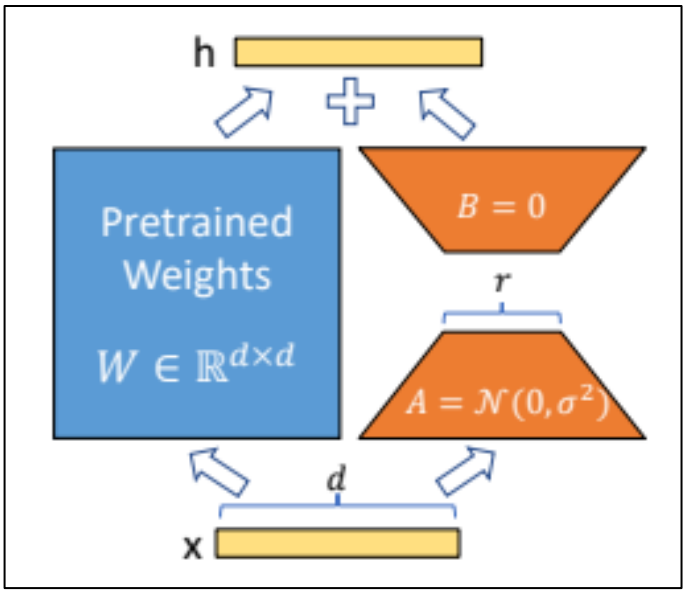


Figure 4.4: Low-Rank Adaptation (LoRa) based parameter efficient fine-tuning framework

CHAPTER 5

DATASET GENERATION AND DATA PREPROCESSING

5.1 Overview of Dataset Generation Framework

This chapter presents the dataset generation, and preprocessing pipeline developed for training and evaluating the proposed Large Language Model (LLM) based resource allocation framework. The dataset is constructed using allocation trajectories generated by the baseline Knowledge Based Reinforcement Learning (KBRL) controller under multiple Radio Access Network (RAN) slicing scenarios.

The primary objective of the dataset generation framework is to construct structured state action mappings capable of representing expert resource allocation behaviour under dynamic wireless network conditions. During simulation, the KBRL controller continuously interacts with the wireless environment and generates allocation trajectories consisting of network observations, Resource Block (RB) allocation actions, Service Level Agreement (SLA) outcomes, and network performance indicators.

Unlike conventional natural language datasets, the generated dataset represents structured sequential decision-making information collected from wireless network simulations. Each allocation sample therefore captures the relationship between the observed network state and the corresponding allocation decision generated by the expert KBRL policy.

The overall dataset generation process evolved incrementally throughout the development of the proposed framework. Initial trajectory generation primarily focused on recording RB allocation decisions and reward observations. In later stages, the dataset pipeline was extended to include slice level KPI measurements, traffic statistics, allocation consistency information, and multi scenario network observations suitable for transformer based supervised policy learning.

The resulting dataset therefore provides a structured representation of expert allocation behaviour and serves as the foundation for instruction tuning and policy approximation using transformer based Large Language Models.

5.2 KBRL Trajectory Generation

The dataset generation process begins with trajectory collection from the baseline Knowledge Based Reinforcement Learning (KBRL) framework presented in Chapter 3.

During simulation, the KBRL controller continuously observes the current network state and generates Resource Block allocation actions for active network slices operating within the Radio Access Network environment.

At every scheduling interval, the simulation environment records the observed state vector, selected allocation action, allocation reward, SLA satisfaction behaviour, and associated network performance metrics.

The generated trajectories therefore represent sequential state action interactions collected during online resource allocation.

The generated dataset contains allocation samples collected under multiple traffic conditions and network scenarios to improve policy generalization during Large Language Model training.

Different traffic intensities, slice utilization patterns, and SLA conditions are introduced during simulation to ensure variability in the generated allocation trajectories.

The recorded allocation trajectories are exported in structured CSV based formats for further preprocessing and supervised learning preparation. Each trajectory entry therefore represents a complete allocation observation consisting of network state information and the corresponding expert allocation decision generated by the KBRL controller.

5.3 State Representation and KPI Extraction

The generated dataset utilizes structured slice level state representations collected from the KBRL simulation environment. Each state vector contains traffic observations, queue statistics, Signal-to-Noise Ratio (SNR) measurements, and service related KPI indicators associated with active network slices.

The proposed framework considers both enhanced Mobile Broadband (eMBB) slices and massive Machine Type Communication (mMTC) slices. Different feature representations are utilized depending on the slice category and traffic configuration generated during simulation.

For eMBB slices, the generated state representation contains either 6-feature or 10-feature vectors depending on the simulation configuration.

The extracted features include Constant Bit Rate (CBR) traffic observations, Variable Bit Rate (VBR) traffic measurements, queue backlog information, and SNR based channel quality indicators.

eMBB Slice Features (6-Feature Representation)

Slice	Feature Name	Description
s [0]	CBR Traffic Load	Constant Bit Rate traffic demand
s [1]	CBR Queue Backlog	Queue occupancy for CBR traffic
s [2]	CBR SNR	Channel quality measurement
s [3]	CBR Queue Secondary	Secondary queue observation
s [4]	CBR SNR Secondary	Secondary SNR measurement
s [5]	VBR Traffic Load	Variable Bit Rate traffic demand

Table 5.1: Feature representation for 6-feature eMBB slice states

eMBB Slice Features (10-Feature Representation)

Slice	Feature Name	Description
s [0]	CBR Traffic Load	Constant Bit Rate traffic demand
s [1]	CBR Queue Backlog	Queue occupancy for CBR traffic
s [2]	CBR SNR	Channel quality measurement
s [3]	CBR Queue Secondary	Secondary queue observation
s [4]	CBR SNR Secondary	Secondary SNR measurement
s [5]	VBR Traffic Load	Variable Bit Rate traffic demand
s [6]	VBR Queue Backlog	Queue occupancy for VBR traffic
s [7]	VBR SNR	Variable Bit Rate channel quality
s [8]	VBR Queue Secondary	Secondary VBR queue measurement
s [9]	VBR SNR Secondary	Secondary VBR SNR measurement

Table 5.2: Feature representation for 10-feature eMBB slice states

mMTC Slice Features (3- feature representation)

Slice	Feature Name	Description
s [0]	Device Ratio	Ratio of active IoT devices
s [1]	Repetition Factor	Number of transmission repetitions
s [2]	Delay	Delay measurement in milliseconds

Table 5.3: Feature representation for mMTC slice states

The extracted state representations are subsequently used for trajectory generation, supervised policy learning, and instruction tuning of the proposed Large Language Model based allocation framework. The feature vectors therefore provide contextual information required for learning state to allocation mappings under dynamic RAN slicing conditions.

5.4 Data Preprocessing and Normalization

The generated allocation trajectories undergo multiple preprocessing stages before being utilized for supervised instruction tuning and transformer-based policy learning. The major preprocessing operations performed on the dataset are summarized as follows:

• Removal of incomplete and invalid trajectory samples generated during simulation.

• Identification and separate handling of inactive slices represented using zero initialized or empty state vectors.

• Normalization of KPI features including traffic load, queue backlog, SNR measurements, and PRB utilization indicators.

• Consistent formatting of eMBB and mMTC slice representations across different traffic scenarios.

• Conversion of state action mappings into structured prompt response pairs suitable for instruction tuning.

• Export of processed trajectories into CSV and JSON based formats for supervised training and inference generation.

• Shuffling and organization of allocation samples to improve training generalization across multiple network conditions.

The resulting pre-processed dataset provides a structured and scalable representation of expert KBRL allocation behaviour suitable for transformer-based policy learning and allocation prediction.

5.5 Prompt Construction and Instruction Formatting

The pre-processed allocation trajectories are converted into structured prompt response pairs for supervised instruction tuning of the proposed Large Language Model based allocation framework. The objective of the prompt construction stage is to transform network state observations and allocation labels into textual instruction-based representations suitable for transformer-based learning.

Each training sample consists of an input prompt containing slice level state information, KPI measurements, traffic indicators, and network observations collected during KBRL simulation. The corresponding output response contains the expert Resource Block (RB) allocation action generated by the KBRL controller.

The generated prompts are designed to preserve contextual information associated with network slicing conditions while enabling the language model to learn state to allocation mappings from sequential allocation trajectories.

Different slice categories including eMBB and mMTC traffic scenarios are incorporated into the prompt structure to improve generalization across heterogeneous network conditions.

SAMPLE PROMPT TABLE

Component	Example
Input Prompt	Slice State: [0.021, 1.571, 9.291, 0.275]
KPI Information	PRB Utilization: 0.74
Slice Type	eMBB
Target Response	Allocate 55 PRBs

Table 5.4: Example instruction tuning sample generated from KBRL allocation trajectories

The instruction tuning format enables the transformer model to learn relationships between network observations and expert allocation decisions using supervised sequence modelling. The generated prompt response pairs are subsequently tokenized and utilized for LoRa based fine tuning of the Mistral-7B-Instruct model.

The resulting instruction dataset therefore provides a scalable and structured representation of expert allocation behaviour suitable for transformer-based policy approximation and reasoning aware resource allocation.

5.6 Training and Validation Split

The generated instruction tuning dataset is partitioned into training and validation subsets for supervised fine tuning of the proposed Large Language Model based allocation framework. Dataset partitioning is performed to evaluate the generalization capability of the fine-tuned model under previously unseen network conditions and traffic scenarios.

The generated allocation samples are shuffled before partitioning to reduce ordering bias associated with sequential trajectory generation. A larger portion of the dataset is utilized for training, whereas the remaining samples are reserved for validation and inference evaluation.

The dataset split is performed while preserving variability across eMBB, mMTC, and inactive slice conditions to ensure balanced representation of heterogeneous network scenarios during training. This improves the robustness of the learned allocation policy and reduces overfitting toward specific traffic conditions.

The resulting training and validation datasets are subsequently utilized for LoRa based supervised instruction tuning of the Mistral-7B-Instruct model.

5.7 Final Dataset Structure

The final instruction tuning dataset was generated using runtime allocation trajectories collected from the baseline Knowledge Based Reinforcement Learning (KBRL) framework during simulation.

The dataset generation process was designed to capture diverse network conditions and allocation behaviours across multiple Radio Access Network (RAN) slicing scenarios.

The generated allocation trajectories were collected from three independent traffic scenarios namely scenario_0, scenario_1, and scenario_2 to improve allocation diversity and policy generalization during transformer based supervised learning.

Multiple simulation runs were performed for each scenario to further increase trajectory variability and allocation coverage across heterogeneous traffic conditions.

DATASET

Parameter	**Value**
Dataset Source	KBRL runtime simulations
Number of Scenarios	3
Scenarios	scenario_0, scenario_1, scenario_2
Runs per Scenario	5
Total Samples	756,000
Slices per Timestep	5
Dataset Format	CSV / Structured prompts

Table 5.5: Final dataset configuration used for instruction tuning

Each timestep in the generated dataset contains state representations associated with five active network slices.

Depending on the generated traffic conditions and slice category, different feature representations are utilized for eMBB and mMTC slices.

The generated eMBB slice representations contain either 6-feature or 10-feature state vectors consisting of Constant Bit Rate (CBR) traffic observations, Variable Bit Rate (VBR) traffic measurements, queue statistics, and Signal-to-Noise Ratio (SNR) indicators.

Similarly, mMTC slices are represented using device activity indicators, repetition factors, and delay related measurements collected during simulation.

Inactive slices are represented using empty state vectors indicating the absence of active traffic demand during the corresponding scheduling interval.

Consequently, the combined network state representation maintains a variable length structure instead of a strictly fixed dimensional representation.

The generated dataset therefore captures diverse allocation behaviours across heterogeneous slice conditions and provides a scalable instruction tuning corpus for transformer-based policy learning and allocation prediction.

Chapter 6

EXPERIMENTAL SETUP

6.1 Experimental Environment

The proposed LLM based RAN slicing controller was developed and evaluated using a hybrid experimental setup consisting of a local development system and a GPU enabled remote workstation. The experimental environment was designed to support baseline KBRL simulation, large scale dataset generation, LoRa based Large Language Model fine tuning, and closed loop allocation evaluation under multiple network slicing scenarios.

6.1.1 Hardware Configuration

The hardware configuration utilized for implementation and evaluation is divided into two stages namely local development and remote GPU based training.

Local Development System

The initial development, debugging, and baseline KBRL simulations were performed on a local development machine with the following configuration:

Component	Specification
Device	MacBook Air M2
Processor	Apple M2
Purpose	KBRL simulation, dataset logging, preprocessing, debugging

Table 6.1: Local development system configuration

Remote GPU Workstation

Large scale LLM training and transformer-based allocation evaluation were performed on a remote GPU enabled Linux workstation.

Component	Specification
GPU	NVIDIA GeForce RTX 5090
GPU Memory	32 GB VRAM
Operating System	Linux
Purpose	LoRA fine tuning, inference, large scale dataset processing

Table 6.2: Remote GPU workstation configuration

The GPU based environment significantly accelerated the fine tuning and inference stages associated with the Mistral-7B-Instruct model during transformer-based allocation learning.

6.1.2 Software Environment

The implementation pipeline was developed using Python based deep learning and networking libraries. The software environment utilized for simulation, dataset generation, and transformer fine tuning is summarized in Table 6.3.

Library / Framework	Purpose
Python 3.10	Core programming environment
NumPy	Numerical computation
Pandas	Dataset processing
Gym	Simulation environment
PyTorch	Deep learning framework
Transformers	LLM loading and inference
PEFT	LoRA based fine tuning
Accelerate	GPU optimized training
Datasets	Dataset handling
Matplotlib	Visualization
tqdm	Progress monitoring

Table 6.3: Software libraries and frameworks used during implementation

The Hugging Face Transformers ecosystem was utilized for loading, fine tuning, and evaluating the Mistral-7B-Instruct model. Parameter efficient fine tuning was performed using the PEFT framework with LoRA based adaptation to reduce GPU memory consumption and training complexity.

Mixed precision training and gradient accumulation were further utilized during LoRA fine tuning to improve computational efficiency and accelerate transformer training on large scale allocation datasets.

6.2 LLM Training Configuration

The proposed allocation framework utilizes supervised instruction tuning of the Mistral-7B-Instruct-v 0.2 model using structured allocation trajectories generated from the baseline Knowledge Based Reinforcement Learning (KBRL) framework.

Parameter efficient fine tuning was performed using the Low Rank Adaptation (LoRA) framework to reduce computational complexity and GPU memory requirements during transformer training.

The generated instruction tuning dataset consisted of structured prompt response pairs containing slice level state observations, KPI measurements,

traffic indicators, and corresponding expert Resource Block (RB) allocation actions generated by the KBRL controller.

During training, the Mistral tokenizer was utilized for sequence tokenization and formatting of instruction-based allocation samples.

Parameter	**Value**
Base Model	Mistral-7B-Instruct-v0.2
Fine-Tuning Method	LoRA
Task Type	Causal Language Modelling
Learning Rate	5e-5
Batch Size	2
Gradient Accumulation Steps	4
Effective Batch Size	8
Epochs	0.5
Maximum Sequence Length	2048
Precision	BF16
LoRA Rank (r)	16
LoRA Alpha	32
LoRA Dropout	0.05
Target Modules	q_p, k_p, v_p, o_p
Logging Steps	20
Checkpoint Save Steps	200

Table 6.4: Hyperparameter configuration used for LoRA based instruction tuning

LoRA based adaptation was applied to the transformer attention projection layers while keeping many pretrained model parameters frozen during training.

The target adaptation modules included the query, key, value, and output projection layers of the transformer attention mechanism.

Mixed precision BF16 training and gradient checkpointing were utilized to reduce GPU memory consumption and improve computational efficiency during transformer fine tuning.

Gradient accumulation was further incorporated to achieve stable optimization with larger effective batch sizes under GPU memory constraints.

The training pipeline utilized supervised causal language modelling in which the instruction prompt was provided as the input sequence, while the corresponding expert allocation action generated by the KBRL framework acted as the target output response.

Prompt tokens associated with the input instruction were masked during loss computation to ensure that optimization was performed only on the generated allocation response tokens.

The resulting fine-tuned model therefore learned to approximate expert KBRL allocation behaviour across heterogeneous RAN slicing conditions while maintaining reduced computational overhead during training and inference generation.

6.3 Instruction Tuning and Inference Pipeline

The proposed training pipeline utilizes supervised instruction tuning for learning Resource Block allocation behaviour from structured KBRL generated allocation trajectories.

The generated dataset consists of instruction-based prompt response pairs in which network state observations and KPI measurements are provided as input prompts, whereas expert allocation actions generated by the KBRL controller are utilized as target responses.

During preprocessing, each allocation sample is converted into an instruction formatted sequence compatible with the Mistral-7B-Instruct tokenizer.

The generated prompt contains slice level state information, traffic indicators, and network observations collected during simulation.

The corresponding output response contains the PRB allocation action associated with the observed network state.

The generated instruction sequences are tokenized using the Mistral tokenizer and subsequently processed using supervised causal language modelling.

During training, prompt tokens associated with the input instruction are masked from loss computation, whereas optimization is performed only on the generated allocation response tokens.

The resulting fine-tuned model can generate allocation predictions for previously unseen network conditions during inference.

During evaluation, the trained model receives structured slice state observations as input and generates PRB allocation predictions for active network slices operating within the RAN slicing environment.

The generated allocation outputs are subsequently evaluated using PRB allocation accuracy, allocation consistency, SLA violation behaviour, and resource utilization related performance metrics.

6.4 Evaluation Metrics

The performance of the proposed LLM based allocation framework was evaluated using multiple allocation prediction and resource utilization metrics. The selected evaluation metrics were designed to measure allocation accuracy, policy consistency, SLA aware behaviour, and operational reliability during inference generation.

• Mean Absolute Error (MAE): Measures the difference between predicted PRB allocations and expert KBRL allocation actions.

• Slice-wise MAE: Evaluates allocation prediction accuracy individually for each network slice.

• Correlation Coefficient: Measures similarity between predicted and expert allocation behaviour.

• Mean PRB Utilization: Evaluates average PRB resource usage generated by the allocation model.

• Invalid Output Rate: Measures the percentage of invalid allocation outputs generated during inference.

• Over-Limit Allocation Rate: Measures the percentage of allocation outputs exceeding PRB constraints.

• Structured Reasoning Evaluation: Evaluates allocation consistency, over-allocation behaviour, under-allocation conditions, and inactive slice resource wastage during inference generation.

The selected evaluation metrics collectively provide quantitative and qualitative assessment of transformer-based allocation performance across heterogeneous RAN slicing conditions.

6.5 Experimental Scenarios

The proposed framework was evaluated under multiple Radio Access Network slicing scenarios generated using the baseline KBRL simulation environment.

Multiple traffic configurations were introduced to evaluate the generalization capability of the fine-tuned Large Language Model under heterogeneous network conditions.

The generated dataset consists of allocation trajectories collected from three independent traffic scenarios namely scenario_0, scenario_1, and scenario_2.

Each scenario represents different traffic distributions, slice utilization patterns, and network load conditions associated with eMBB and mMTC slices.

Multiple simulation runs were performed for each scenario to improve trajectory diversity and allocation variability during instruction tuning.

This enabled the proposed transformer-based allocation framework to learn resource allocation behaviour across varying traffic dynamics and slice level KPI conditions.

The generated scenarios therefore provided a diverse evaluation environment for analysing allocation consistency, PRB utilization behaviour, and SLA aware allocation performance during inference generation.

6.6 Baseline Comparison Framework

The performance of the proposed Large Language Model based allocation framework was compared against the baseline Knowledge Based Reinforcement Learning controller presented in Chapter 3.

The KBRL controller acted as the expert allocation policy responsible for generating the instruction tuning trajectories utilized during supervised learning.

During evaluation, the fine-tuned transformer model received structured slice state observations as input and generated PRB allocation predictions for active network slices.

The generated allocation outputs were subsequently compared against the expert KBRL allocation actions using allocation accuracy, resource utilization, and SLA related evaluation metrics.

The comparison framework therefore enabled quantitative evaluation of the capability of the proposed transformer-based model to approximate expert allocation behaviour across multiple RAN slicing scenarios.

Chapter 7

RESULTS

7.1 Overview of Experimental Evaluation

This chapter presents the performance evaluation and experimental analysis of the proposed Large Language Model based resource allocation framework for Radio Access Network slicing environments.

The objective of the evaluation process is to analyse the capability of the fine-tuned transformer model to approximate expert Knowledge Based Reinforcement Learning (KBRL) allocation behaviour under multiple network slicing scenarios.

The proposed framework was evaluated using allocation trajectories generated from heterogeneous traffic scenarios consisting of eMBB, mMTC, and inactive slice conditions.

The generated allocation outputs were compared against expert KBRL allocation actions using allocation accuracy, resource utilization, and structured reasoning-based evaluation metrics.

The experimental analysis focuses on transformer convergence behaviour, PRB allocation prediction quality, allocation consistency, SLA aware resource allocation, and inference reliability under dynamic network conditions.

Both quantitative and qualitative evaluation approaches were utilized to analyse the effectiveness of the proposed LoRA fine-tuned Mistral-7B-Instruct allocation framework.

The following sections present detailed analysis of training convergence, allocation prediction performance, structured reasoning behaviour, and comparative evaluation against the baseline KBRL controller.

7.2 Baseline KBRL Performance Analysis

7.2.1 SLA Violation Behaviour

The SLA violation behaviour of the baseline KBRL framework was evaluated across multiple network slicing scenarios and compared against conventional

model free reinforcement learning approaches including PPO, TRPO, A2C, SAC, TD3, and NAF.

The obtained results demonstrate that the proposed KBRL controller consistently achieves lower SLA violation rates during online learning compared with most model free reinforcement learning baselines.

The reduced violation behaviour indicates improved allocation safety and better SLA aware resource allocation under dynamic traffic conditions.

The experimental observations additionally indicate that model free reinforcement learning methods exhibit unstable exploration behaviour during training, resulting in larger fluctuations in SLA violation rates across different scenarios.

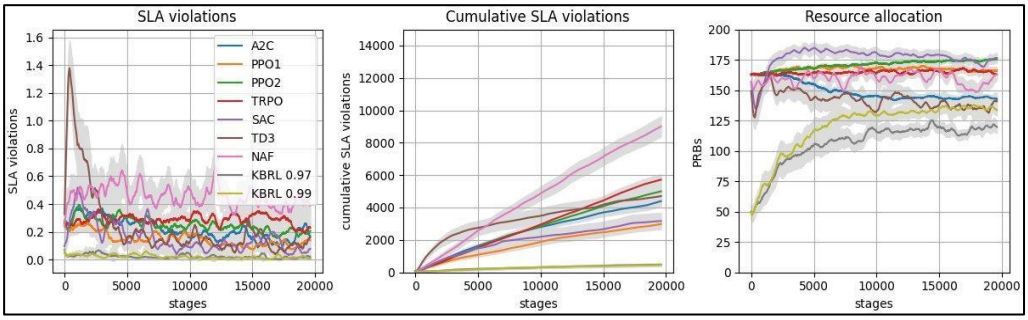


Fig7.1. Results for scenario_1. From left to right: SLA violations per stage, cumulative SLA violations, and total RBs allocated. Each plot shows the per-stage average of 30 simulation runs and its confidence interval with a 90% confidence level.

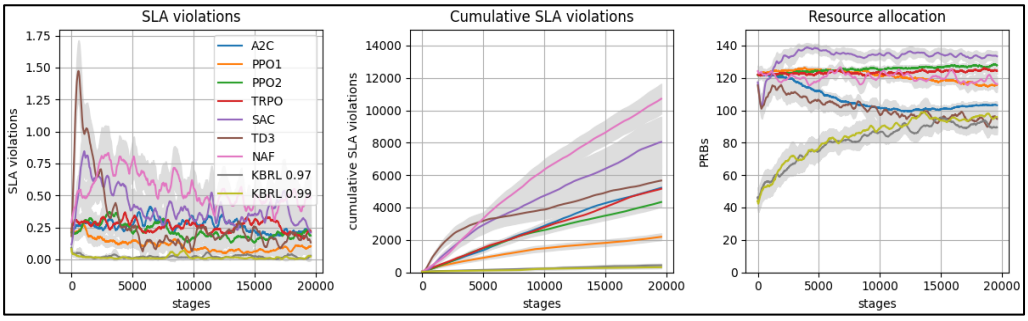


Fig7.2. Results for scenario_2. From left to right: SLA violations per stage, cumulative SLA violations, and total RBs allocated. Each plot shows the per-stage average of 30 simulation runs and its confidence interval with a 90% confidence level.

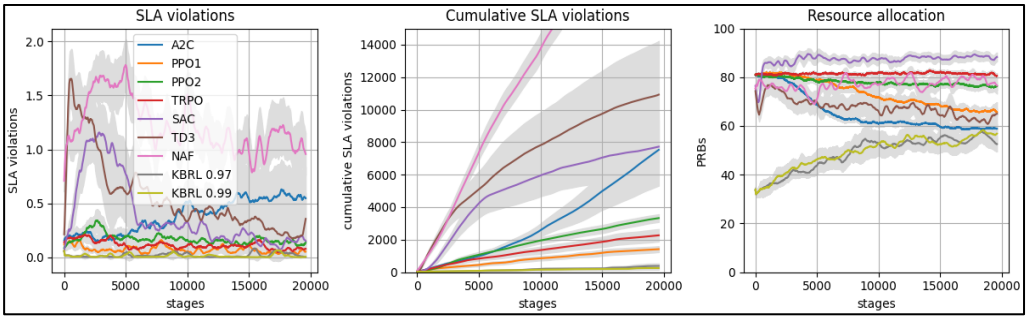


Fig7.3. Results for scenario_3. From left to right: SLA violations per stage, cumulative SLA violations, and total RBs allocated. Each plot shows the per-stage average of 30 simulation runs and its confidence interval with a 90% confidence level.

7.2.2 Comparative SLA Violation vs Resource Occupation Analysis

This section presents the baseline performance analysis of the Knowledge Based Reinforcement Learning (KBRL) framework utilized for Radio Access Network slicing and Resource Block allocation.

The objective of this analysis is to evaluate the capability of the baseline KBRL controller to achieve efficient PRB allocation while minimizing Service Level Agreement (SLA) violations under dynamic traffic conditions.

The KBRL framework was evaluated under multiple network slicing scenarios consisting of heterogeneous traffic conditions and varying slice level KPI observations.

The baseline controller utilizes kernel based online learning for SLA prediction and constrained optimization-based allocation decision making during each scheduling interval.

The experimental evaluation focused on analysing the trade-off between PRB resource utilization and SLA satisfaction behaviour during online learning. Multiple baseline reinforcement learning approaches including PPO, TRPO, A2C, SAC, TD3, DQN, and NAF were compared against the proposed KBRL framework in terms of SLA violation rate and resource occupation characteristics.

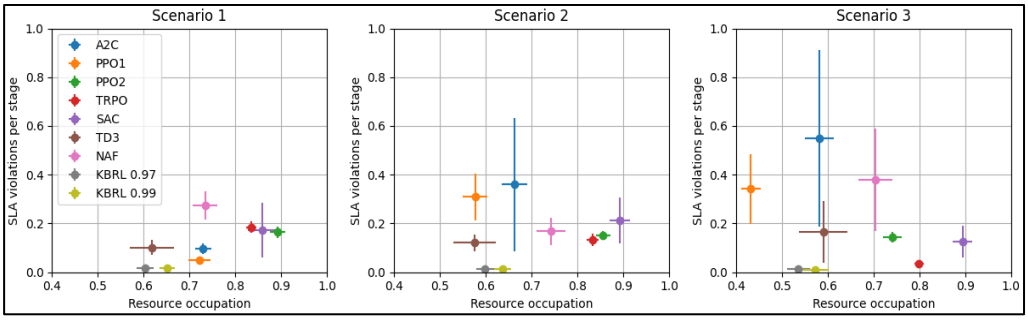


Figure 7.4: SLA violations per stage versus resource occupation for KBRL and model free reinforcement learning approaches

The obtained results demonstrate that the KBRL framework achieves significantly lower SLA violation rates while maintaining controlled PRB resource occupation compared with conventional model free reinforcement learning approaches.

The baseline KBRL controller therefore provides a stable and resource efficient expert allocation policy suitable for generating high quality allocation trajectories for transformer based supervised learning.

The observed experimental behaviour further highlights the dependency between PRB allocation efficiency and SLA satisfaction under dynamic network slicing conditions.

Lower resource occupation generally increases the probability of SLA violations, whereas excessive allocation results in inefficient spectrum utilization.

The KBRL controller attempts to balance this trade off using online kernel-based prediction and constrained allocation optimization.

The baseline analysis additionally demonstrated stable online learning behaviour and reduced unsafe exploration compared with model free reinforcement learning methods.

Consequently, the KBRL framework was selected as the expert policy generator for constructing the instruction tuning dataset utilized during Large Language Model based policy learning and allocation approximation.

7.3 LLM Training Convergence Analysis

This section presents the training convergence behaviour of the proposed LoRA fine tuned Mistral-7B-Instruct model during supervised instruction tuning for Resource Block allocation learning.

The objective of the convergence analysis is to evaluate the stability of transformer optimization and analyse the impact of dataset diversity on allocation learning performance.

Initial experiments were conducted using allocation trajectories generated from a single network slicing scenario consisting of one simulation run.

The generated dataset contained structured prompt response pairs derived from KBRL allocation trajectories and corresponding expert PRB allocation actions.

Supervised instruction tuning was subsequently performed using LoRA based adaptation of the Mistral-7B-Instruct model.

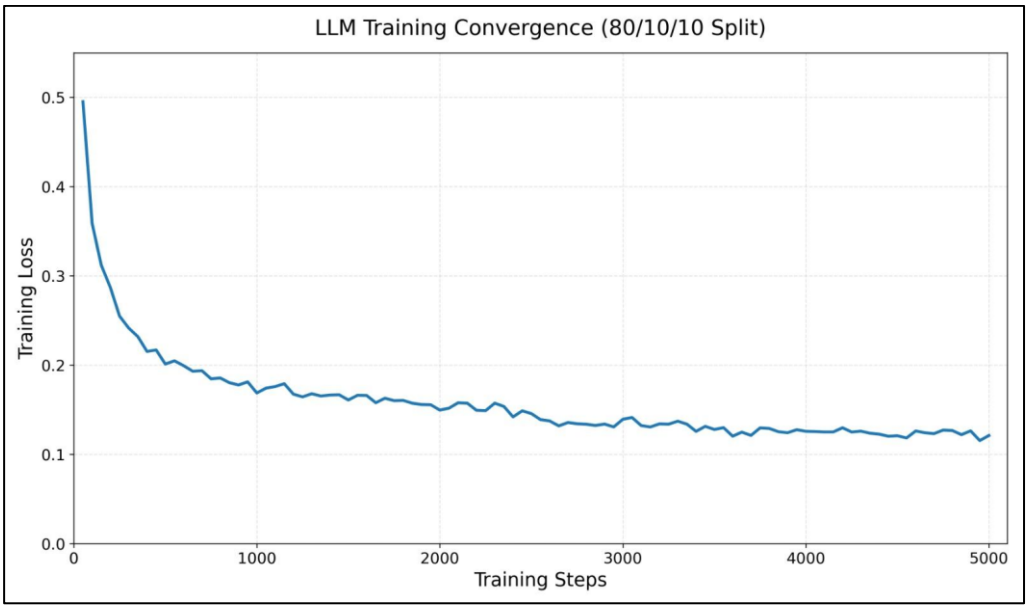


Figure 7.5: Training convergence behaviour of the LoRA fine tuned Mistral-7B-Instruct model during supervised instruction tuning

The obtained training results demonstrate stable convergence behaviour during transformer fine tuning.

The training loss decreases rapidly during the initial optimization stages and gradually stabilizes as the model approaches convergence.

The observed validation loss of approximately 0.12 further indicates successful supervised adaptation of the transformer model to the allocation prediction task.

The initial single scenario training experiments confirmed the capability of the proposed transformer-based framework to learn state to allocation mappings from structured KBRL allocation trajectories.

However, the limited diversity associated with single scenario training restricted the generalization capability of the learned allocation policy under heterogeneous traffic conditions.

In the final implementation phase, the dataset generation framework was extended to include multiple traffic scenarios and repeated simulation runs to improve allocation diversity and trajectory coverage during instruction tuning.

The final dataset therefore contained allocation samples generated across heterogeneous eMBB, mMTC, and inactive slice conditions collected from multiple KBRL simulation environments.

The introduction of multi scenario allocation trajectories significantly improved training robustness, allocation consistency, and transformer generalization capability during inference generation.

The resulting fine tuned model demonstrated improved convergence behaviour and reduced allocation prediction error across previously unseen network slicing conditions.

7.4 LLM-Based Policy Imitation Performance

This section presents the allocation prediction performance of the proposed LoRA fine tuned Mistral-7B-Instruct framework during supervised policy imitation of the baseline KBRL controller.

The objective of the evaluation process is to analyse the capability of the transformer model to approximate expert PRB allocation behaviour under dynamic Radio Access Network slicing conditions.

The generated allocation outputs were evaluated using Mean Absolute Error (MAE), slice wise allocation deviation, correlation analysis, resource utilization behaviour, and inference reliability related performance metrics.

The evaluation process compared the generated transformer allocation outputs against expert allocation actions generated by the baseline KBRL framework.

Metric	**Value**
MAE	4.95 PRBs
Slice-wise MAE	3.88 – 5.52 PRBs
Correlation	0.8249
Mean PRB Allocated (KBRL)	145.42
Mean PRB Allocated (LLM)	142.50
Invalid Output Rate	0%
Over-limit Rate	1.07%

Table 7.1: Allocation prediction performance of the LoRA fine tuned LLM policy imitation framework

The obtained results demonstrate that the proposed transformer based allocation framework successfully learns expert allocation behaviour from KBRL generated allocation trajectories.

The observed Mean Absolute Error of approximately 4.95 PRBs indicates that the generated transformer allocation outputs remain closely aligned with expert KBRL allocation actions during inference generation.

The obtained correlation coefficient of 0.8249 further demonstrates strong similarity between transformer generated allocation behaviour and expert KBRL allocation decisions.

Additionally, the generated allocation outputs maintain PRB utilization characteristics comparable to the baseline KBRL framework while preserving low invalid output generation behaviour.

The observed invalid output rate of 0% indicates that the fine tuned transformer model consistently generates structurally valid allocation responses during inference.

Similarly, the low over limit allocation rate demonstrates that the generated allocation outputs generally satisfy PRB resource constraints under heterogeneous network slicing conditions.

Slice	KBRL Allocation	LLM Allocation
Slice 1	17	15
Slice 2	12	11
Slice 3	8	9
Slice 4	5	6
Slice 5	6	7

Table 7.2: PRB allocation comparison between KBRL and LLM generated allocation outputs

The obtained experimental results therefore validate the capability of the proposed LoRA fine tuned transformer framework to approximate expert KBRL allocation behaviour using supervised instruction tuning over structured allocation trajectories.

7.5 Resource-Constrained Policy Learning

To further improve allocation stability and reduce excessive PRB utilization during inference generation, the instruction tuning framework was extended with resource constrained allocation learning.

The objective of this stage was to encourage the transformer model to generate allocation outputs that remain closely aligned with expert KBRL behaviour while simultaneously respecting PRB utilization constraints under dynamic traffic conditions.

Unlike the earlier policy imitation experiments, the resource constrained learning framework incorporated allocation consistency checks, structured reasoning based allocation evaluation, and multi scenario training trajectories collected from heterogeneous network slicing environments.

The resulting dataset therefore exposed the transformer model to a wider range of traffic conditions including high demand eMBB slices, delay sensitive mMTC traffic, and inactive slice conditions.

Metric	Value
MAE	3.86 PRBs
Slice-wise MAE	3.72 – 4.02 PRBs
Correlation	0.8947
Mean PRB Allocated (KBRL)	106.70
Mean PRB Allocated (LLM)	105.93
Invalid Output Rate	0%
Over-limit Rate	0.80%

Table 7.3: Performance evaluation of the resource constrained LLM allocation framework

The obtained results indicate noticeable improvement in allocation prediction quality compared with the earlier supervised policy imitation framework.

The overall allocation error decreases to approximately 3.86 PRBs while the correlation between transformer generated allocations and expert KBRL allocation behaviour increases to 0.8947.

The reduction in Mean Absolute Error demonstrates improved alignment between generated PRB allocation outputs and expert allocation behaviour under heterogeneous traffic conditions.

Similarly, the higher correlation coefficient indicates that the transformer model successfully captures the allocation dynamics and resource utilization behaviour associated with the baseline KBRL controller.

An additional improvement is observed in PRB utilization efficiency. The generated transformer allocations maintain average PRB occupation closely matching the expert KBRL policy while simultaneously reducing excessive allocation behaviour.

Furthermore, the observed reduction in over limit allocation rate indicates improved compliance with PRB resource constraints during inference generation.

The obtained experimental behaviour therefore demonstrates that multi scenario instruction tuning combined with resource constrained allocation learning significantly improves transformer based allocation consistency, resource

utilization efficiency, and generalization capability across heterogeneous RAN slicing conditions.

7.6 Generalization Across Multiple Scenarios

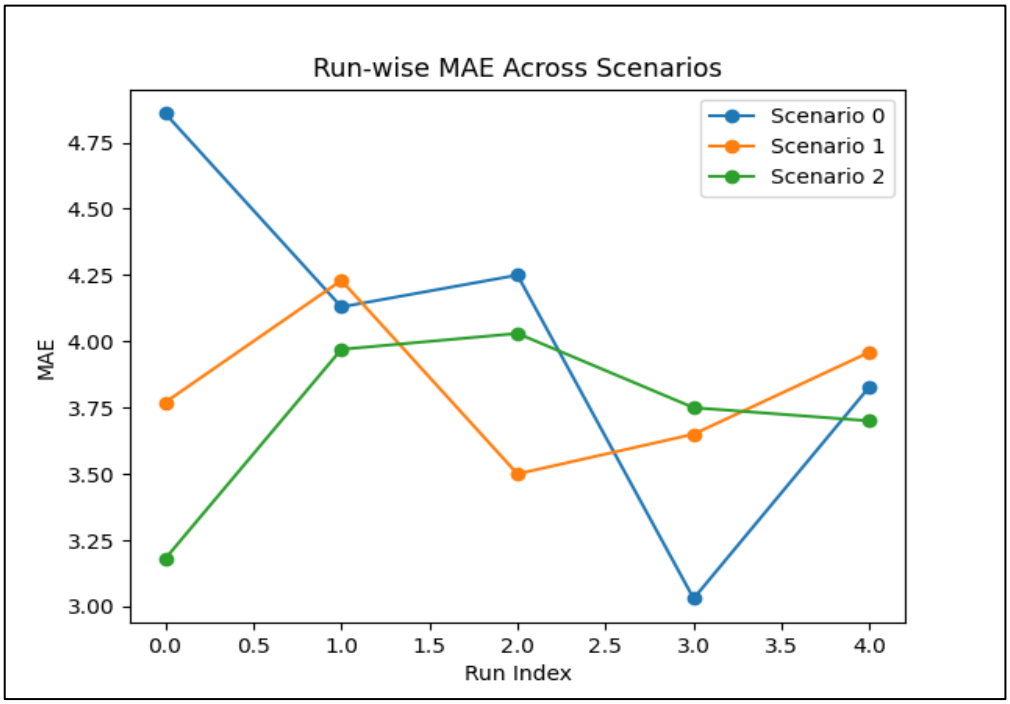


Figure 7.6: Generalization behaviour of the proposed LLM allocation framework across Scenario 0, Scenario 1, and Scenario 2

This section evaluates the generalization capability of the proposed transformer based allocation framework under heterogeneous Radio Access Network slicing conditions.

Unlike the initial single scenario experiments, the final instruction tuning dataset incorporated allocation trajectories collected from multiple traffic environments in order to improve robustness and allocation consistency during inference generation.

Figure 7.6 illustrates the allocation performance trend of the fine tuned transformer model across Scenario 0, Scenario 1, and Scenario 2. The obtained results demonstrate that the proposed LLM based allocation framework maintains stable allocation behaviour and consistent prediction performance across multiple network slicing conditions.

The introduction of multi scenario training trajectories significantly improves the capability of the transformer model to generalize beyond individual traffic environments. The resulting model therefore learns broader allocation patterns associated with varying eMBB traffic demand, mMTC activity levels, queue dynamics, and channel quality conditions.

Although minor allocation variations are observed across different scenarios due to heterogeneous traffic distributions and slice level KPI dynamics, the overall allocation behaviour remains closely aligned with expert KBRL allocation trends across all evaluated environments.

The obtained experimental analysis therefore demonstrates that multi scenario instruction tuning substantially improves transformer robustness and allocation generalization capability under dynamic RAN slicing conditions.

7.7 Structured Reasoning and Explainability

In addition to generating PRB allocation predictions, the proposed transformer based allocation framework was extended with structured reasoning capabilities to improve interpretability and transparency during Radio Access Network slicing decisions.

Unlike conventional reinforcement learning based allocation, the proposed framework additionally produces contextual reasoning associated with the generated allocation behaviour.

The reasoning framework reformulates RAN slicing as a reasoning aware allocation problem in which allocation decisions are jointly analysed using traffic demand, queue backlog conditions, device activity behaviour, and channel quality observations.

Component	**Example**
Slice State	Slice1_State: [0.021, 1.571, 9.291, 0.275]
Generated Allocation	Slice1_PRB: 55
Generated Reasoning	High queue backlog with stable SNR indicates increasing traffic demand requiring higher PRB allocation

Table 7.4: Structured reasoning output generated by the proposed LLM allocation framework

The generated reasoning outputs further classify allocation behaviour into multiple severity categories namely OK, Mild, and Severe depending on the alignment between expected traffic demand and generated PRB allocation behaviour.

This enables detection of over allocation, under allocation, and inefficient spectrum utilization during inference generation.

Classification	**Interpretation**
OK	Allocation behaviour aligned with expected traffic demand
Mild	Minor deviation between expected and generated allocation
Severe	Significant over allocation or under allocation behaviour

Table 7.5: Structured reasoning severity classification used during allocation evaluation

An additional advantage of the proposed reasoning framework is the capability to identify inactive slice resource wastage conditions during inference generation.

Inactive slices represented using empty or zero initialized state vectors are expected to receive minimal PRB allocation.

Consequently, excessive PRB allocation assigned to inactive slices is automatically classified as severe resource wastage behaviour.

Similarly, the reasoning framework can identify traffic aware allocation inconsistencies associated with high demand eMBB slices, delay sensitive mMTC traffic, and unstable queue backlog conditions.

This enables qualitative evaluation of allocation behaviour beyond conventional numerical allocation accuracy metrics.

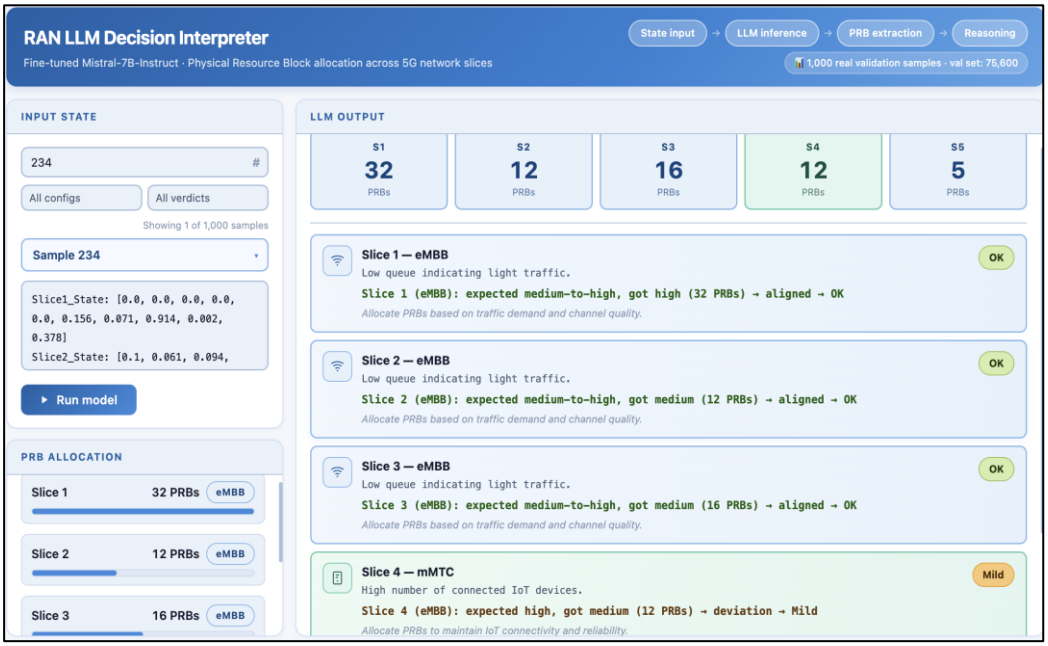


Figure 7.7: Interactive transformer based PRB allocation and reasoning interpretation interface for RAN slicing evaluation

An interactive reasoning visualization interface was additionally developed for qualitative analysis of transformer generated PRB allocation behaviour during RAN slicing evaluation.

The developed interface enables real time visualization of slice level state observations, generated PRB allocation outputs, traffic aware allocation reasoning, and severity-based allocation interpretation.

The interface further integrates structured reasoning analysis for active eMBB and mMTC slices by jointly evaluating traffic demand, queue conditions, device

activity behaviour, and channel quality observations during inference generation.

Allocation outputs are subsequently classified into OK, Mild, and Severe categories depending on the consistency between expected traffic demand and generated PRB allocation behaviour.

The developed reasoning interface therefore improves transparency and interpretability of transformer-based resource allocation decisions while enabling interactive evaluation of allocation consistency across heterogeneous network slicing conditions.

The obtained reasoning-based evaluation demonstrates that the proposed transformer framework can generate explainable and context aware allocation decisions in addition to numerical PRB allocation prediction.

This significantly improves interpretability and operational transparency compared with conventional reinforcement learning based resource allocation approaches.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 Conclusion

This thesis presented a transformer based resource allocation framework for Radio Access Network slicing using Large Language Models and supervised policy learning from Knowledge Based Reinforcement Learning allocation trajectories. The proposed framework integrates kernel based reinforcement learning, structured dataset generation, LoRA based transformer fine tuning, and reasoning aware allocation interpretation for dynamic PRB allocation under heterogeneous network slicing conditions.

Initially, the baseline KBRL framework was analysed for SLA aware PRB allocation under multiple traffic scenarios consisting of eMBB and mMTC slices. The obtained baseline results demonstrated that the KBRL controller achieves stable allocation behaviour with reduced SLA violations and controlled PRB occupation compared with conventional model free reinforcement learning approaches.

Subsequently, structured allocation trajectories generated using the KBRL controller were converted into instruction based prompt response datasets suitable for supervised transformer fine tuning. The Mistral-7B-Instruct model was fine tuned using LoRA based parameter efficient adaptation to learn state to allocation mappings from heterogeneous network slicing conditions.

Experimental evaluation demonstrated that the proposed transformer based framework successfully approximates expert KBRL allocation behaviour using supervised instruction tuning. The obtained results showed strong allocation correlation, reduced allocation prediction error, stable PRB utilization behaviour, and low invalid output generation across multiple network slicing scenarios.

The final implementation additionally introduced resource constrained policy learning and structured reasoning-based allocation interpretation to improve allocation consistency and operational transparency during inference generation. An interactive reasoning visualization interface was further developed for qualitative evaluation of transformer generated PRB allocation behaviour and severity-based allocation interpretation.

Overall, the obtained results demonstrate the feasibility of utilizing Large Language Models as intelligent and explainable allocation frameworks for next generation RAN slicing environments.

8.2 Limitations

Although the proposed transformer based allocation framework demonstrated promising allocation prediction and reasoning capabilities, several limitations remain associated with sequential deployment and real time RAN control environments.

The current implementation primarily focuses on supervised policy imitation of expert KBRL allocation trajectories rather than direct online reinforcement learning based optimization. Consequently, allocation performance may degrade under long horizon closed loop deployment scenarios involving rapidly changing traffic distributions and previously unseen network conditions.

Additionally, the generated transformer allocations are dependent on the quality and diversity of the expert KBRL allocation trajectories utilized during instruction tuning. Limited trajectory diversity or insufficient representation of rare traffic conditions may affect generalization capability during inference generation.

The proposed reasoning framework currently performs structured qualitative interpretation based on allocation consistency and traffic demand analysis. However, deeper causal reasoning and long horizon allocation planning remain open research challenges for transformer based RAN control systems.

Furthermore, large transformer models introduce computational overhead and inference latency compared with lightweight heuristic allocation approaches, thereby requiring efficient optimization strategies for real time deployment in practical wireless communication environments.

8.3 Future Work

Several research directions can be explored in the future to further improve transformer-based RAN slicing and resource allocation frameworks.

One important direction involves integration of online reinforcement learning with transformer-based policy learning to enable adaptive closed loop allocation optimization under dynamic network conditions. Hybrid RL plus LLM architectures may significantly improve long horizon allocation stability and real time policy adaptation capabilities.

Future work may additionally explore retrieval augmented networking frameworks in which external network telemetry databases, historical traffic traces, and KPI repositories are incorporated during inference generation to improve allocation reasoning and contextual decision making.

Another promising research direction involves development of agentic networking architectures capable of autonomous traffic analysis, anomaly

detection, SLA monitoring, and multi objective allocation optimization using collaborative Large Language Model agents.

The proposed reasoning framework can also be extended toward multimodal network intelligence by incorporating channel traces, visual telemetry dashboards, and temporal traffic dynamics into transformer-based allocation reasoning pipelines.

Finally, deployment-oriented optimization including model quantization, lightweight transformer adaptation, and edge accelerated inference can be explored for practical real time implementation of transformer based intelligent RAN slicing systems in future 6G wireless communication environments.

REFERENCES

[1] J. J. Alcaraz, F. Losilla, A. Zanella, and M. Zorzi, “Model-Based Reinforcement Learning With Kernels for Resource Allocation in RAN Slices,” IEEE Transactions on Wireless Communications, vol. 22, no. 1, pp. 486–501, Jan. 2023.

[2] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions,” IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 2429–2453, 2018.

[3] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. C. Liang, and D. I. Kim, “Applications of Deep Reinforcement Learning in Communications and Networking: A Survey,” IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3133–3174, 2019.

[4] R. Li, Z. Zhao, X. Zhou, J. Palicot, and H. Zhang, “Deep Reinforcement Learning for Resource Management in Network Slicing,” IEEE Access, vol. 6, pp. 74429–74441, 2018.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[6] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “LoRA: Low-Rank Adaptation of Large Language Models,” in International Conference on Learning Representations (ICLR), 2022.

[7] T. Brown, B. Mann, N. Ryder, et al., “Language Models are Few-Shot Learners,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.

[8] H. Touvron, T. Lavril, G. Izacard, et al., “LLaMA: Open and Efficient Foundation Language Models,” arXiv preprint arXiv:2302.13971, 2023.

[9] A. Jiang, A. Sablayrolles, A. Mensch, et al., “Mistral 7B,” arXiv preprint arXiv:2310.06825, 2023.

[10] H. Wang, Y. Wu, G. Min, J. Xu, and P. Tang, “Data-Driven Dynamic Resource Scheduling for Network Slicing: A Deep Reinforcement Learning Approach,” Information Sciences, vol. 498, pp. 106–116, 2019.

[11] Y. Sun, M. Peng, S. Mao, and S. Yan, “Hierarchical Reinforcement Learning for Network Slicing in Fog Radio Access Networks,” IEEE Transactions on Vehicular Technology, vol. 68, no. 12, pp. 12061–12076, 2019.

[12] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, “Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial,” IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3039–3071, 2019.

[13] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network Slicing in 5G: Survey and Challenges,” IEEE Communications Magazine, vol. 55, no. 5, pp. 94–100, 2017.

[14] M. Bennis, M. Debbah, and H. V. Poor, “Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale,” Proceedings of the IEEE, vol. 106, no. 10, pp. 1834–1853, 2018.

[15] S. Zhang, S. Xu, and G. Y. Li, “Deep Learning in Physical Layer Communications,” IEEE Wireless Communications, vol. 26, no. 2, pp. 93–99, 2019.

[16] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, “Wireless Network Intelligence at the Edge,” Proceedings of the IEEE, vol. 107, no. 11, pp. 2204–2239, 2019.

[17] O. Simeone, “A Very Brief Introduction to Machine Learning With Applications to Communication Systems,” IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 4, pp. 648–664, 2018.

[18] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, pp. 436–444, 2015.

[19] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[20] C. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[21] D. Silver, A. Huang, C. J. Maddison, et al., “Mastering the Game of Go with Deep Neural Networks and Tree Search,” Nature, vol. 529, pp. 484–489, 2016.

[22] R. Sutton and A. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[23] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2020.

[24] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” in Advances in Neural Information Processing Systems (NeurIPS), 2013.

[25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[26] Hugging Face, “Transformers Documentation,” 2025. [Online]. Available: https://huggingface.co/docs/transformers

[27] Hugging Face, “PEFT Documentation,” 2025. [Online]. Available: https://huggingface.co/docs/peft

[28] Hugging Face, “Datasets Documentation,” 2025. [Online]. Available: https://huggingface.co/docs/datasets

[29] Hugging Face, “Accelerate Documentation,” 2025. [Online]. Available: https://huggingface.co/docs/accelerate

[30] PyTorch, “PyTorch Documentation,” 2025. [Online]. Available: https://pytorch.org/docs/stable/index.html

[31] OpenAI Gym, “Gym Documentation,” 2025. [Online]. Available: https://www.gymlibrary.dev/

[32] S. Haykin, Neural Networks and Learning Machines, 3rd ed. Pearson, 2009.

[33] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[34] A. Goldsmith, Wireless Communications. Cambridge University Press, 2005.

[35] T. S. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Pearson, 2002.