LLM Based B5G/6G Resource Allocator Documentation

1. Introduction

The "LLM Based B5G/6G Resource Allocator" project introduces an innovative AI-driven solution for optimizing resource allocation in Beyond 5G (B5G) and 6G networks. Leveraging Large Language Models (LLMs) via Perplexity AI and Federated Learning, this tool addresses the dynamic, high-demand needs of next-generation networks, ensuring real-time, privacy-preserving management of bandwidth, power, and latency. This report details the project's objectives, architecture, novelty, implementation, and potential applications, positioning it as a pioneering advancement in wireless network optimization.

2. Project Overview

2.1 Objective

Develop a scalable, intelligent system to allocate resources (bandwidth, power, latency) in B5G/6G networks, enhancing efficiency, privacy, and adaptability for advanced applications like holographic communications and autonomous systems.

2.2 Key Features

- Real-time resource allocation using Perplexity AI's LLM.
- Federated Learning for distributed, privacy-preserving training across network nodes.
- SQLite database for tracking and visualizing allocation history.
- Interactive dashboards with Plotly visualizations (gauges, bar charts, 3D plots).
- User-configurable Quality of Service (QoS) levels, refresh rates, and AI optimization.

3. Novelty and Differences from Existing Projects

The "LLM Based B5G/6G Resource Allocator" stands out from existing B5G/6G and resource allocation projects in the following ways:

3.1 Comparison with Existing Projects

Traditional 5G/6G Resource Allocation:

- Most existing projects (e.g., 3GPP standards, Nokia's 6G research) rely on rule-based or heuristic algorithms for resource allocation, lacking AI-driven adaptability. Our project uses LLMs for intelligent, real-time decision-making, offering higher accuracy and flexibility.
- Privacy is often a concern in centralized systems; we implement federated learning to protect sensitive network data, a feature absent in projects like Ericsson's 6G trials or Qualcomm's AI for 5G.

AI-Based Resource Allocation:

- Projects like Google's DeepMind for 5G or Huawei's AI for 6G use machine learning but focus on centralized models or specific use cases (e.g., traffic prediction). Our federated LLM approach ensures distributed learning, scalability, and privacy across B5G/6G networks.
- Unlike Intel's AI-driven 6G prototypes, which emphasize simulation, our tool operates in real-time with Perplexity AI, integrating self-learning for dynamic adaptation.

Federated Learning in Networks:

- While some projects (e.g., Samsung's federated learning for IoT in 5G) apply federated learning, they target device-level data privacy, not network-wide resource allocation. Our project uniquely combines federated learning with LLMs for B5G/6G optimization, addressing both privacy and efficiency at scale.

Database and Visualization:

- Existing systems often lack historical tracking or interactive visualization. Our SQLite database and Plotly dashboards provide a novel, user-friendly interface for analyzing past allocations, a feature not typically found in projects like AT&T's 5G AI or NTT's 6G research.

3.2 Novelty

- The integration of Perplexity AI's LLM for real-time B5G/6G resource allocation is unprecedented, combining natural language processing with network optimization.
- Federated Learning ensures privacy-preserving, distributed training, a first for B5G/6G resource management at this scale.
- Dynamic self-learning adapts to network changes, reducing computational overhead and enhancing efficiency, distinguishing it from static or heuristic-based approaches.
- Historical tracking with SQLite and interactive visualizations enable data-driven decision-making, a gap in current B5G/6G solutions.

4. Technical Architecture

4.1 Model Training

- **Training Process**: The LLM connects to Perplexity AI via API, exchanging network data seamlessly. Self-learning techniques dynamically adjust internal parameters based on real-time inputs (e.g., bandwidth, power, latency).
- **Data Source**: Distributed network data from B5G/6G nodes, processed locally before aggregation via federated learning.

4.2 Federated Learning

- **Implementation**: Trains on distributed data across network nodes without transferring raw data, preserving privacy and security.
- **Benefits**: Reduces computational overhead, enhances scalability, and ensures compliance with data regulations for 6G networks.

4.3 Hyperparameter Tuning

- Tuned LLM parameters (e.g., temperature, max_tokens) via API calls to optimize accuracy, latency, and resource utilization.
- Focused on minimizing prediction errors and maximizing efficiency for B5G/6G scenarios.

4.4 Prompt Tuning

- Addressed potential hallucinations (inaccurate predictions) by refining prompts to ensure contextually accurate outputs for resource allocation.
- Conducted error analysis and adjusted hyperparameters for improved retrieval accuracy.

5. Implementation

5.1 Dependencies

- streamlit: For the interactive web interface.
- requests: For API calls to Perplexity AI.
- pandas: For data manipulation and storage.
- plotly: For visualizations (gauges, charts, 3D plots).
- sqlite3: For database management (included in Python).

5.2 Setup

1. Clone the repository

```
git clone <repository-url>

cd LLM-Based-B5G-6G-Resource-Allocator
```

2. Install dependencies:

   bash

   WrapCopy

```
pip install streamlit requests pandas plotly
```

3. Ensure SQLite3 is installed (default in Python).

5.3 Usage

1. Run the application:

   bash
   ```
   python -m streamlit run app.py
   ```

2. Configure network settings in the sidebar (users, bandwidth, power, etc.).
3. Click "Start Allocation" to begin real-time resource allocation.
4. Use "Continuous Mode" for ongoing updates or stop manually.
5. View historical allocations via the sidebar "Enable History Viewer" checkbox.
6. Export allocation data as JSON for analysis.

6. Project Components

6.1 Available Bandwidth

- **Definition**: Total MHz available for allocation (default 100 MHz, user-configurable via sidebar).
- **Purpose**: Sets the upper limit for bandwidth distribution across users.

6.2 Available Power

- **Definition**: Total Watts available (default 50 W, user-configurable).
- **Purpose**: Defines the power capacity for network operations.

6.3 Refresh Rate

- **Definition**: Seconds between allocation updates (default 5 s, user-configurable).
- **Purpose**: Controls how often the system recalculates and updates allocations.

6.4 AI Optimization

- **Definition**: Toggle in the sidebar to enable LLM-driven refinement of allocations (user-enabled).
- **Purpose**: Enhances accuracy and efficiency using Perplexity AI's intelligence.

6.5 Quality of Service (QoS) Levels

- **Levels**:
  - **High**: Low latency (1–20 ms), high reliability (95–100%).
  - **Medium**: Moderate latency (20–50 ms), reliability (80–95%).
  - **Low**: Higher latency (50–100 ms), reliability (50–80%).
- **Configuration**: Set via sidebar per user, influencing LLM prompts for allocation.
- **Purpose**: Tailors resource allocation to user needs for performance and reliability.

6.6 Max Latency and Reliability

- **Max Latency**: User-defined per user (1–100 ms) via sliders, driving allocation requests.
- **Reliability**: User-defined per user (50–100%), ensuring stable connections.
- **Purpose**: Customizes allocation to meet specific performance requirements.

6.7 Bandwidth, Power, and Health Score

- **Bandwidth and Power**: Calculated by summing allocated values from the LLM response (e.g., User_X: Y MHz, Z Watts, W ms latency).
- **Health Score**: 0–100, derived from LLM's energy efficiency and network health scores.

- **Calculations**:
  - Bandwidth/Power: Sum of values in the LLM response.
  - Latency: Average across users.
  - Health Score: Parsed directly from LLM output.

6.8 Resource Gauges

- **Definition**: Plotly indicators showing real-time bandwidth and power usage.
- **Usage**: Displayed in dashboards during and after allocation for monitoring.
- **Applications**: Helps network administrators track resource utilization, optimize allocations, and prevent over-allocation.

7. Applications and Use Cases

- **Smart Cities**: Optimize traffic lights, IoT sensors, and public infrastructure in B5G/6G networks.
- **Holographic Communications**: Ensure low latency and high bandwidth for 6G holographic applications.
- **Autonomous Systems**: Provide real-time resource allocation for drones, vehicles, and robotics in 6G.
- **Healthcare**: Enable reliable, low-latency connections for telemedicine and remote surgeries in 6G.
- **Industry 4.0**: Support AI-driven factory automation with privacy-preserving, efficient allocation.

8. Future Work

- Integrate a real AI Assistant (e.g., xAI Grok, free Hugging Face models) for interactive guidance.
- Scale federated learning to more network nodes for broader coverage and robustness.

- Implement predictive analytics for proactive resource allocation based on historical trends.
- Develop energy-harvesting optimizations for 6G sustainability.
- Create a mobile app or web dashboard for broader accessibility and user engagement.

9. Conclusion

The "LLM Based B5G/6G Resource Allocator" revolutionizes network resource management with its AI-driven, privacy-preserving approach. By integrating LLMs, federated learning, and a robust database, it addresses B5G/6G challenges, enabling next-generation applications while reducing overhead. This solution sets a foundation for 6G adoption, offering scalability, efficiency, and innovation for global networks.

10. Technical Details

- **Dependencies**: streamlit, requests, pandas, plotly, sqlite3.
- **Setup Instructions**: See Section 5.2.