**Final HandsOn 8 Documentation**

1. **How LoRA Improved Model Performance**

Reduced GPU Memory Usage → LoRA fine-tuning updates only a subset of model weights, reducing memory requirements.

Faster Fine-Tuning → Instead of modifying the entire model, LoRA adjusts key layers, speeding up training.

Improved Efficiency → Enables fine-tuning on smaller GPUs, making AI training more accessible.

Better Response Quality → Optimized temperature, top-k, and no-repeat settings to generate more natural responses.

1. **How the Streamlit App Enhances User Interaction**

User-Friendly Interface → Simple text input for flight queries and real-time AI recommendations.

Real-Time Search → Uses FAISS indexing for fast semantic search of flight data.

AI-Powered Travel Insights → Generates context-aware recommendations based on user queries.

Clean & Responsive Design → Utilizes columns and styling for a professional look.

1. **Real-World Impact of the Hands-on**

Faster Travel Planning → Users receive quick AI-generated flight summaries, reducing manual searches.

Personalized Travel Recommendations → AI suggests best flight options based on user needs.

Scalability → Can be extended to hotels, trains, and car rentals for a complete travel assistant.

Business Applications → Airlines & travel agencies can integrate AI-based booking assistants to improve customer experience.

1. **Why This LoRA Configuration?**

Efficient Fine-Tuning → Updates only key attention layers, reducing computation.  
Improved Model Adaptation → lora\_alpha=32 ensures better learning efficiency.  
Balanced Dropout → lora\_dropout=0.05 prevents overfitting while keeping training stable.  
Supports Smaller GPUs → Fine-tunes without modifying the entire model, saving memory.