Summary and Next Steps

🎯 Key Concepts

🔍 **Self-Attention** Mechanism

🎭 **Multi-Head** Attention

📍 **Positional Encoding**

✅ Advantages

⚡ **Parallelization** - Fast training

🔗 **Long-range dependencies** - Global context

🎨 **Flexibility** - Multiple domains

⚠️ Challenges

📐 Quadratic complexity $O(n^2)$

💾 Large memory requirements

🚀 Next Steps

1 **Next Lecture**
Advanced topics: Sparse attention, Linear attention

2 **Efficient Transformers**
Explore Reformer, Linformer, Performer

3 **Practice**
Implement a mini-Transformer from scratch

4 **Explore Libraries**
Hugging Face Transformers library

📚 Recommended Reading

"Attention Is All You Need" paper

Vaswani et al., 2017