Part 1/3:

Attention-Based Interpretability

1. Attention Visualization
2. Layer-wise Relevance Propagation
3. Gradient-Based Attribution
4. Integrated Gradients
5. SHAP for Medical Applications
6. LIME for Clinical Text
7. Concept Activation Vectors