<u>**Lecture 3 :**</u> **Synthesis of Research Paper: "Automated Source Code Generation and Auto-Completion Using Deep Learning: Comparing and Discussing Current Language Model-Related Approaches"**

1. Introduction:

- The paper focuses on using deep learning in language models, namely programming languages.

- Fills a gap in the empirical literature by comparing different deep learning architectures for source code creation and auto-completion.

2. methodology:

- The experiment included three neural network architectures: AWD-LSTM, AWD-QRNN, and Transformer (GPT-2, BERT, RoBERTa).

- Tokenization methods included word, unigram, character, and Byte-Pair Encoding (BPE).

- Used a Python dataset from the "GitHub CodeSearchNet Challenge dataset" for training and testing.

- Applied transfer learning, pre-training models on English texts, and fine-tuning to the Python dataset.

3. Results:

- AWD-LSTM with char tokenization achieved the highest accuracy (0.779633) in source code creation.

- Transformer models (BERT and RoBERTa) obtained near-perfect auto-completion accuracy (0.999238 and 0.999468, respectively).

- Pre-trained models consistently beat non-pre-trained ones.

- The results contrasted favorably to current literature, outperforming numerous cutting-edge models in source code generation.

4. discussion:

- Recognized potential underfitting in AWD-LSTM char, AWD-QRNN char, and Transformer models, recommending additional investigation via training.

- emphasized the importance of pre-training on overall model performance.

- Results were compared to current literature to illustrate the competitive accuracy attained in source code generation.

5. Conclusion:

-This article examines and compares various deep learning architectures and tokenization strategies for source code generation and auto-completion.

- AWD-LSTM with char tokenization and Transformer models (BERT and RoBERTa) showed promising results.

- Offers useful insights to scholars and practitioners in the field of deep learning for programming languages, promoting additional research.