# Momenta Assignment

## Approach Selection:

### 1. Paper Title:

### *Spoofing Attacker Also Benefits from Self- Supervised Pretrained Model*

1. Key-technical innovations:
    - Utilizes **self-supervised speech models** (HuBERT/WavLM) trained on large-scale unlabeled audio.
    - Learns **rich speech representations**, improving **generalization** to unseen attacks.
2. Reported Performance metric:
    - **EER:** ~0.5%-1.5% (best-in-class performance on ASVspoof).
    - **Accuracy:** ~98% (when fine-tuned on ASVspoof datasets).
3. Reason for this approach being promising:
    - Generalizes well to diverse audio sources and unseen spoofing attacks.
    - Captures more complex speech features than traditional MFCC/LFCC-based models.
    - Adapts dynamically with fine-tuning, reducing dataset bias issues.
4. Potential Limitations or Challenges:
    - **Computationally expensive**—requires GPUs for real-time inference.
    - **Large model size**—not ideal for edge devices or embedded systems.
    - Still vulnerable to adversarial attacks targeting its feature space.

### 2. Paper Title:

### *How to Boost Anti-Spoofing with X-Vectors*

1. Key-technical innovations:
    - Uses **TDNN-based x-vector embeddings** for speaker verification.
    - Incorporates **LFCC/MFCC feature extraction** to enhance robustness.
2. Reported Performance metric:
    - **EER:** ~1-2% on ASVspoof 2019 dataset.
    - **Accuracy:** ~95% for detecting known spoofing attacks.
3. Reason for this approach being promising:
    - **Fast inference speed** (suitable for real-time applications).

- **Well-established in ASV (Automatic Speaker Verification) pipelines**.
- **Efficient for limited-resource environments** (low computational cost).

4. Potential Limitations or Challenges:
   - **Struggles with adversarial AI-generated voices** trained to mimic x-vector patterns.
   - **Feature dependency (MFCC/LFCC may not capture advanced spoofing techniques)**.
   - **Overfitting risk if trained on a limited dataset**.

## 3. Paper Title:

## *AASIST: Audio Anti-Spoofing Using Integrated Spectro-Temporal Graph Attention Networks*

1. Key-technical innovations:
   - Uses **Graph Attention Networks (GATs)** to model **spectro-temporal dependencies** in audio.
   - Learns **contextual relationships** between different speech segments for robust anti-spoofing.
2. Reported Performance metric:
   - **EER: 0.24%** (best among recent deep learning approaches).
   - **Accuracy: 99%+** on controlled datasets.
3. Reason for this approach being promising:
   - **Most advanced deep learning method for anti-spoofing**.
   - **Strong against unseen spoofing techniques** due to feature learning via GATs.
   - **Captures both local and global spectral patterns**, making it more adaptable.
4. Potential Limitations or Challenges:
   - **High computational cost**, limiting real-time deployment feasibility.
   - **Overfitting risk if trained on a dataset that lacks spoofing diversity**.
   - **Difficult to interpret (black-box nature of deep learning models)**.