Problem Statement:

"Resilient Federated Anomaly Detection in Mission-Critical Networks under Adversarial Conditions"

Mission-critical networks (MCNs)—such as those in defense, healthcare, energy, and transportation—must maintain strict availability, confidentiality, and integrity. These networks are increasingly dependent on AI-based intrusion detection systems (IDS) to defend against evolving threats. However, traditional centralized IDS approaches are unsuitable due to:

- **Data privacy and regulatory constraints**
- **High latency and limited bandwidth**
- **Vulnerability to single-point failures and targeted attacks**

Moreover, **federated learning (FL)** has emerged as a promising alternative, enabling collaborative anomaly detection without raw data sharing. But FL in MCNs faces critical challenges:

- **Data heterogeneity** across distributed MCN nodes (e.g., sensors, gateways, control systems)
- **Adversarial participants (e.g., model poisoning)**
- **Limited resources (e.g., computation, energy)**

There is a lack of robust federated anomaly detection frameworks that can **adapt to adversarial conditions** and **ensure trustworthy collaboration** in the sensitive environment of MCNs.

Proposed Solution:

"TRUST-MCNet: A Trust-Aware Resilient Federated Learning Framework for Securing Mission-Critical Networks"

We propose **TRUST-MCNet**, an end-to-end federated anomaly detection framework designed specifically for mission-critical networks. Key contributions of the solution include:

1. **Trust-Based Client Selection**:
   Integrate a dynamic trust scoring system that monitors local model updates using a combination of:
    - Model deviation analysis (e.g., cosine similarity with global model)

    - Behavioral history (e.g., consistency over rounds)
    - Lightweight local explainability (e.g., SHAP score alignment with known threat patterns)

2. Clients with low trust are down-weighted or excluded from aggregation, mitigating poisoning attacks.

3. **Robust Aggregation Using Trimmed Mean + Gradient Clipping**:
   Enhance global model resilience using:
    - **Trimmed mean** aggregation to suppress outliers
    - **Gradient clipping** to bound update influence and resist malicious spikes

4. **Adaptive Learning Rate for Non-IID Heterogeneity**:
   Introduce per-client adaptive learning rates that account for:
    - Data distribution skew
    - Local performance feedback
    - Communication budget constraints

5. **Explainable Threat Attribution Engine**:
   Post-training, apply **SHAP-based model explainability** on local detectors to provide interpretable anomaly reports, essential for high-stakes environments (e.g., SCADA, aircraft networks).

6. **Experimental Validation on MCN Simulated Testbeds**:
   Evaluate TRUST-MCNet on realistic MCN scenarios using datasets like **TON_IoT, Edge-IIoT, and MedBIoT**, with attacks tailored for:
    - SCADA disruption
    - Industrial control hijacking
    - Data exfiltration