Proposed Standard Outline for new PWI on

"Privacy Preservation of training data for ML"

- ❑ Scope
- ❑ Normative References
- ❑ Terms & Definitions
- ❑ Symbols & Abbreviated terms
- ❑ Overview
- ❑ Impact of privacy preservation on effectiveness of AI Models
- ❑ Approaches to privacy preservation*:
  - ❑ Methods that use data de-identification before training
  - ❑ Alternate methods for output control (in lieu of de-identification)
- ❑ Methods of de-identification for various data formats:
  - ❑ Structured
  - ❑ Unstructured – documents, weblogs
  - ❑ Unstructured - Voice, Photo, Video
- ❑ Integrating privacy preservation into AI life cycle
- ❑ Guidance on selection of right privacy preservation method

- ❑ Bibliography

**Illustrative list based on exploratory study is provided in next page*

Privacy Preservation Methods

- ❑ *During data preparation stage:*
  - *Attribute Suppression (k-anonymity, l-diversity, t-closeness)*
  - *Perturbation Techniques (e.g., differential privacy)*
  - *Surrogate dataset (e.g., synthetic data)*
  - *PAC (Probably Approximately Correct) – under development at MIT*
  - *Irreversible Video Redaction*

- ❑ *During Model development stage:*
  - *Differential Private Training*
  - *Encrypted Machine learning*
  - *Federated Learning*
  - *Lottery ticket Hypothesis – under development at CSAIL*
  - *PATE framework*

- ❑ *During Model Serving stage*

- *Encrypted inference*
- *Oblivious Transformation*
- *Probability Randomization (or Confidence masking of model outputs)*