Session 1 — Benchmarking and Evaluation

Day 2 · Friday, 26 June 2026

Chair: Lishan Yang Rapporteur: Pietro Liguori

Patrick Grother (NIST) — Face Recognition: A Solved Problem?

Xiaoming Liu (UNC Chapel Hill) — Biometric Recognition: Accuracy, Privacy, and Trustworthiness

How Face Recognition Works

- 1:N identification (duplicate IDs; licenses have no chip)
- Image → feature vector → score vs. threshold
- The template is invertible → privacy, anti-spoofing
- John Lennon: missed in 2014, matched in 2025

When Is It “Solved”?

- NIST: the independent, transparent benchmark
- ≈100% accuracy but only when everything lines up:
  - recent algorithm
  - a mate in the database
  - recent, unmanipulated image
  - limited pose/quality

Where It Still Fails

- Hard cases: twins, look-alikes, low quality, bad angle
- Error rates uneven across groups: ~1/16,000 → ~1/72
- Bias and Issues:
  - China algorithms ↔ Chinese faces;
  - wrongful arrest;
  - humans worse than algorithms

Accuracy: Recognition at a Distance

- BRIAR (IARPA): recognition far away; **FarSight** (MSU), a finalist team
- 5 modalities: face · body · gait · fusion · unified
- FusionAgent (CVPR’26) · SapiensID: one model, face + body (CVPR’25)
- Best in NIST BRIAR evaluation

Privacy: Training Without Real Faces

- Real faces → consent / privacy problem
- GenAI: synthesize fake identities + styles
- Good synthetic set: ID control + style diversity + many subjects
- Synthetic ≈ real

Trustworthiness: Defending Against Attacks

- Anti-spoofing: masks, replay, makeup (long IARPA effort)
- Deepfake detection + image attribution
- Manipulation localization (which pixels were edited)
- Model parsing: reverse-engineer the generator

Takeaways

- Recognition ≈ solved in good conditions → frontier: hard cases, fairness, privacy, trust
- GenAI cuts both ways: tool (synthetic data) · threat (deepfakes)
- Accuracy benchmarking is mature (NIST); new problems need new benchmarks