10) NL - NegL Research Note（NL - NegL 研究備忘）

Date: 2025-12-24

Purpose

Track the current status of the NL/NegL exploration and how it could help the FER model training pipeline.

Where the work lives

- Research folder:
  - `research/nl_negl/`

Motivation (project context)

- The project combines multiple datasets with label noise and domain shift.
- Distillation (KD/DKD) helped calibration after temperature scaling, but macro-F1 gains were not guaranteed in the first student run.
- NL/NegL methods are being explored as a way to reduce the impact of noisy labels and improve robustness.

Current status

- This area is **scaffolded** as research notes/rules.
- No final integrated training result is claimed in this mini-report.

Success criteria (for future experiments)

- Macro-F1 improves vs baseline (CE, and/or KD/DKD) on the chosen benchmark(s).

- Calibration improves or stays stable (lower ECE/NLL after temperature scaling).

- Confusion on hard pairs reduces (e.g., Fear↔Surprise, Sad↔Neutral) as measured by confusion matrix / per-class F1.

- Training remains reproducible (same manifests, seeds, and evaluation protocol).

Integration idea (future)

Potential integration points (to be validated experimentally):

- Apply NegL-style reweighting or sample filtering during CE/KD/DKD training.

- Track impact on:

  - macro-F1

  - calibration (ECE/NLL)

  - confusion between similar emotions (e.g., Fear vs Surprise, Sad vs Neutral)

Next steps

- Define a minimal experiment matrix:

  - baseline CE vs CE+NegL

  - KD vs KD+NegL

- Ensure identical manifests and evaluation protocol so results are comparable.