WristPrint Module Researcher Checklist: Mitigating Re-identification in Wearable Data

1. Data Collection & Processing for Sharing

- **[ ] Minimize Sampling Frequency:** High-frequency data (e.g., numerous samples per second) captures more micro-movements that can be used for identification. Reduce sampling rate and duration of sampling to the minimum required for your specific activity recognition goals.
- **[ ] Activity-Specific Partitioning:** WristPrint findings show that **high-intensity exercise** carries the highest re-identification risk. Consider aggressive down-sampling or adding noise specifically during high-intensity intervals when sharing raw sensor data.
- **[ ] Segment Randomization:** Avoid sharing long, continuous sequences of raw data. Break data into shorter, non-contiguous epochs to disrupt the temporal patterns (RNN-detectable "fingerprints") used for re-identification.

2. Privacy-Preserving Transformations

- **[ ] Implement Differential Privacy:** Add calibrated Laplacian or Gaussian noise to raw acceleration signals. Note the trade-off: increased work to protect identity may reduce the accuracy of fine-grained activity classification. Consider the specific research questions for which the data are needed, and combine with several mitigation strategies for the best, balanced, effect.
- **[ ] Feature-Level Sharing:** Whenever possible, release derived features (e.g., measures of central tendency, derived variables such as Step Counts) rather than raw time-series data.
- **[ ] Signal Distillation/Abstraction:** Use technical methods to generate synthetic data that preserves the statistical properties of the movement for research utility while discarding the unique nuances of the original participant data.

3. Administrative & Governance Controls

- **[ ] Tiered Access Models:** Instead of public repositories, consider "Data Enclaves" where researchers run code on the data within a secure environment without downloading the raw files.
- **[ ] Linkage Attack Hardening:** Conduct a "Linkage Audit." Consider whether your participants are also in public datasets (like the UK Biobank). If a participant's data exists elsewhere, their risk of re-identification via cross-referencing is significantly higher. Informing participants of this risk is important.
- **[ ] Updated Informed Consent:** Explicitly inform participants that motion data is potentially re-identifiable, similar to DNA or fingerprints, moving beyond the "de-identified data" label.