October 16, 2025

Editors of *SoftwareX*

Dear Editors,

I am pleased to submit our manuscript entitled **"Sleuth: Detecting Circular Bias in AI Model Evaluation via Statistical Consistency Metrics"** for consideration as a Software Article in *SoftwareX*.

The rapid adoption of machine learning in scientific research has exposed a critical but under-addressed threat to reproducibility: **circular bias**—the practice of adaptively tuning evaluation protocols (e.g., datasets, compute budgets, hyperparameters) in response to observed model performance. This creates self-reinforcing feedback loops that inflate results and undermine scientific claims, as recently highlighted in *Patterns* (Kapoor & Narayanan, 2023) and the broader reproducibility literature.

While tools exist for tracking experiments or auditing model fairness, none provide **automated, statistically rigorous diagnostics** for circularity in the evaluation process itself. *Sleuth* fills this gap by introducing three novel indicators—PSI, CCS, and ρ_{PC}—combined into a composite Circular Bias Score (CBS) with bootstrap-based uncertainty quantification. The tool operates entirely client-side via a no-code web interface, ensuring privacy while enabling researchers, reviewers, and auditors to detect and mitigate evaluation bias.

Key strengths of Sleuth include:

- First open-source framework specifically targeting circular bias in AI evaluation
- Statistical rigor: 94% detection accuracy on synthetic benchmarks, validated on real-world ImageNet logs
- Privacy-preserving: zero data transmission, fully client-side execution
- Immediate applicability: integrates with existing evaluation workflows via CSV input
- Open science compliance: Creative Commons Attribution 4.0 International license, Zenodo DOI (10.5281/zenodo.17201032), and public GitHub repository

This work directly supports *SoftwareX*'s mission to publish high-impact, reusable research software that advances methodological rigor. Sleuth addresses a timely and urgent need in the ML and AI communities, where evaluation integrity is increasingly scrutinized.

The manuscript is original, unpublished, and not under consideration elsewhere. All authors have approved this submission.

Thank you for your consideration. I look forward to your feedback.

Sincerely,

Hongping Zhang[†]
Independent Researcher
ORCID: 0009-0000-2529-4613
Email: zhanghongping@gmail.com
[†]Corresponding author