



PAPER DIGEST

CERTIFICATE OF RECOGNITION

MOST INFLUENTIAL ACL 2025 PAPER

This is to certify that the paper
**PKU-SafeRLHF: Towards Multi-Level Safety
Alignment for LLMs with Human Preference**

*Jiaming Ji, Donghai Hong, Borong Zhang, Boyuan Chen, Josef Dai, Boren Zheng, Tianyi Alex Qiu, Jiayi Zhou, Kaile Wang,
Boxun Li, Sirui Han, Yike Guo, Yaodong Yang*

presented at the Annual Meeting of the Association for Computational Linguistics (ACL), 2025
is recognized among the Most Influential ACL 2025 Papers



Paper Digest
New York, New York

Edition 2026-03
Issued 2026-03 - Impact Factor 4

[Verify this certificate at paperdigest.org](https://paperdigest.org)

Certificate ID 3ae2c6cce526fb8e - Ranking constructed from citations in research papers and granted patents.

Scan to verify

