



PAPER DIGEST

CERTIFICATE OF RECOGNITION

MOST INFLUENTIAL ICLR 2024 PAPER

This is to certify that the paper
**Fine-tuning Aligned Language Models Compromises
Safety, Even When Users Do Not Intend To!**

Xiangyu Qi, Yi Zeng, Tinghao Xie, Pin-Yu Chen, Ruoxi Jia, Prateek Mittal, and Peter Henderson
presented at the International Conference on Learning Representations (ICLR), 2024
is recognized among the Most Influential ICLR 2024 Papers



Paper Digest
New York, New York



Edition 2026-03
Issued 2026-03 - Impact Factor 7

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

Certificate ID 5438f61af5af9d6b - Ranking constructed from citations in research papers and granted patents.

Scan to verify

