



PAPER DIGEST

CERTIFICATE OF RECOGNITION

MOST INFLUENTIAL ICML 2024 PAPER

This is to certify that the paper
**HarmBench: A Standardized Evaluation Framework
for Automated Red Teaming and Robust Refusal**

*Mantas Mazeika, Long Phan, Xuwang Yin, Andy Zou, Zifan Wang, Norman Mu, Elham Sakhaee, Nathaniel Li, Steven Basart,
Bo Li, David Forsyth, Dan Hendrycks*

presented at the International Conference on Machine Learning (ICML), 2024
is recognized among the Most Influential ICML 2024 Papers



Paper Digest
New York, New York



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

Verify this certificate at paperdigest.org

Certificate ID 6ee7b0115913bed4 - Ranking constructed from citations in research papers and granted patents.

Scan to verify

