Preprints.org

Article

Not peer-reviewed version

Withdrawn: Optimizing Safety Alignment and Jailbreak Defense for Large Language Models

Jiekai Wu and Shiyin Lin *

Posted Date: 21 November 2025

Withdrawn Date: 22 July 2026

doi: 10.20944/preprints202511.1521.v1

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 November 2025 doi:10.20944/preprints202511.1521.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Withdrawn: Optimizing Safety Alignment and Jailbreak Defense for Large Language Models

Jiekai Wu [1] and Shiyin Lin [2,*]

[1] Juntendo University, Bunkyo City, Tokyo, 113-0033,Japan
[2] University of Florida, Gainesville, FL, 10092, United States
* Correspondence: shiyin.aslin@gmail.com

Withdrawal Statement

This preprint has been withdrawn at the request of the authors because institutional authorization was not obtained.

WITHDRAWAL

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.