



Association for Computational Linguistics

Copyright Transfer and Assignment Agreement

FinHarmBench: Financial Jailbreak Benchmark and Unsupervised Safety Fine-Tuning via Refusal Steering Distillation

Yubin Choi, Yujin Yang, Subin Kim, Seokil Ham, Seungju Cho, Jungmin Son, Youngjun Kwak, Changick Kim

This Copyright Transfer and Assignment Agreement (“Agreement”) is entered into between the Association for Computational Linguistics (“ACL”) and the Author(s) listed above (“Author”) in connection with Author’s submission of the above referenced Work to ACL.

In exchange of adequate consideration, ACL and the Author(s) agree as follows:

Transfer/Assignment of Copyright. In the event the Work is accepted for publication by ACL, including for presentation at a meeting/event sponsored by the ACL and for publication in the proceedings of that meeting/event, Author transfers and assigns to ACL without compensation the Author’s rights, title, and interests pertaining to the Work, including all copyrights, copyright licenses and copyright interests.

Copyright License. The Parties acknowledge and agree that the Work is subject to the [Creative Commons Attribution 4.0 International Public License](#), and that ACL will be identified as the Licensor of the Work with the copyright notice:

Copyright © 2020 Association for Computational Linguists (ACL). All Rights Reserved.

Author is granted the same Licensed Rights and subject to the same terms and conditions provided under the Creative Commons Attribution 4.0 International Public License.

Warranty. Author represents and warrants that the Work is Author’s original work and does not infringe on the proprietary rights of others. Author further warrants that he or she has obtained all necessary permissions from any persons or organizations whose materials are included in the Work, and that the Work includes appropriate citations that give credit to the original sources.

Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea

By signing below, I confirm that (please select one):

- (All authors signing) I/we are the Author(s) and have full rights to sign this Agreement.
- (Company representative signing) The Work was prepared within the scope of the Author’s employment and deemed a Work for Hire, and I am authorized to sign this Agreement on behalf of the Author’s employer.
- (One author signing) The Work was prepared jointly by all authors identified above, all authors of the Work have agreed to the above terms, and I am authorized to sign this Agreement on their behalf.

If you have questions about this agreement, please contact your venue chairs.

/ YuJin Yang / graduate student (doctoral program)

Date: April 20, 2026