



CERTIFICATE OF AWARD

DDL 2.0: AIGC Safety from Dual Perspectives of Model and Content

From Echo to Nexus: How Attacker Personality
Dictates Success in Jailbreaking LLMs

Best Paper Award

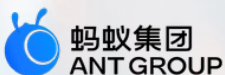
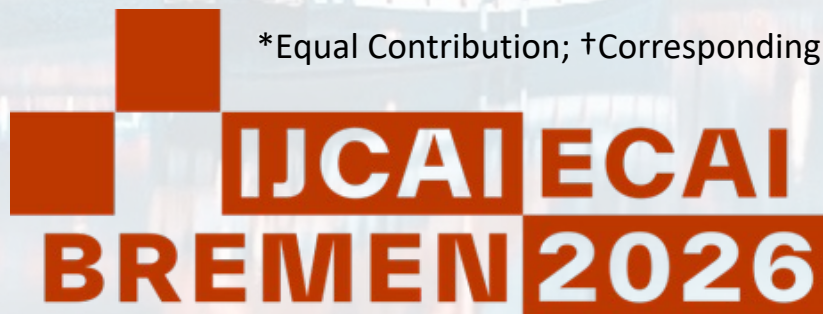
Yuhang Wang^{1,2*}, Jian Zhao^{1*†}, Rubin He^{3*}, Huilin Zhou¹,
Tianle Zhang^{1†}, Rui Feng^{2†}, Xuelong Li^{1†}

¹ Institute of Artificial Intelligence (TeleAI), China Telecom

² Fudan University

³ Northwestern Polytechnical University

*Equal Contribution; †Corresponding Authors.



数字科技



中国科学院自动化研究所
INSTITUTE OF AUTOMATION CHINESE ACADEMY OF SCIENCES



Ca' Foscari
University
of Venice



UNIVERSITY OF
OXFORD

MAX PLANCK INSTITUTE
FOR SECURITY AND PRIVACY



ZOLIOZ



新疆大学
Xinjiang University



香港科技大学
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE



西安交通大学

Jianshu Li, Changtao Miao, Yunfeng Diao, Chenhao Lin, Zhigang Guo, Joey Tianyi Zhou

Chairs