



CERTIFICATE OF AWARD

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

Gold Award

in The OpenClaw Security Attack and Defense
Challenge

Team: PoliteInjection

Jun Zhang^{1,2*}, Jian Zhao^{1*†}, Rubin He^{3*}, Yuming Liu¹, Huilin Zhou¹,
Tianle Zhang^{1†}, Wenqi Ren^{4†}, Xuelong Li^{1†}

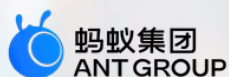
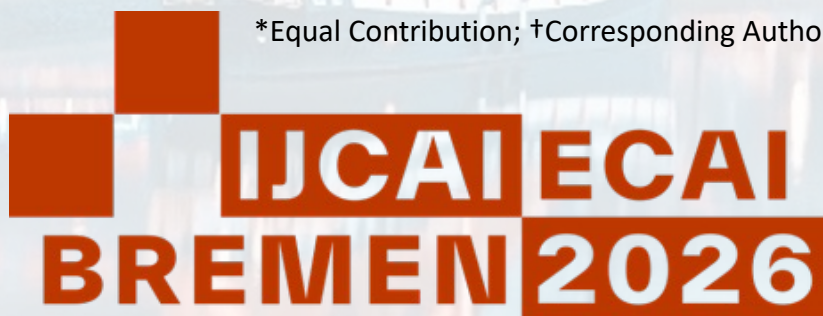
¹ Institute of Artificial Intelligence (TeleAI), China Telecom

² Institute of Information Engineering, Chinese Academy of Sciences

³ Northwestern Polytechnical University

⁴ Sun Yat-sen 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



Centre for
Frontier AI
Research
A*STAR CFAR



Ca' Foscari
University
of Venice



合肥工业大学
HEFEI UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF
OXFORD

MAX PLANCK INSTITUTE
FOR SECURITY AND PRIVACY



ZOLIOZ



蚂蚁天玑实验室
ANT PHOENIX LAB



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