



PAPER DIGEST

CERTIFICATE OF RECOGNITION

MOST INFLUENTIAL NAACL 2024 PAPER

This is to certify that the paper
**XSTest: A Test Suite for Identifying Exaggerated
Safety Behaviours in Large Language Models**

Paul Rottger, Hannah Kirk, Bertie Vidgen, Giuseppe Attanasio, Federico Bianchi, and Dirk Hovy

presented at the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2024
is recognized among the Most Influential NAACL 2024 Papers



Paper Digest
New York, New York



Edition 2026-03
Issued 2026-03 - Impact Factor 5

[Verify this certificate at paperdigest.org](https://paperdigest.org)

Certificate ID 6330401b92c1afb1 - Ranking constructed from citations in research papers and granted patents.

Scan to verify

