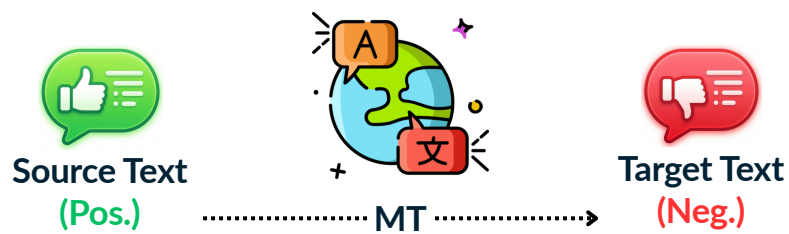


TL;DR

Lost in translation: *your opinion!*

We prove that current MT systems systematically alter subjective meaning and political stance, even when translations seem perfectly fluent.

1 THE PROBLEM: Blind Spots in MT Evaluation

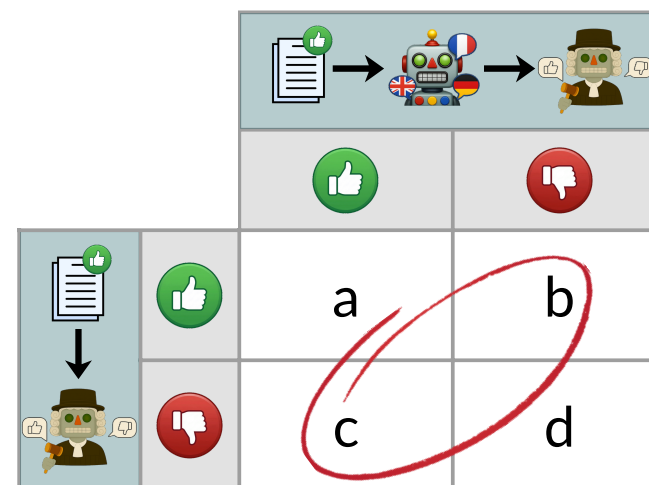


- Widespread MT deployment implicitly assumes **literal meaning and subjective framing** (e.g., stance/opinion) are preserved.
- Standard metrics (BLEU, COMET) prioritize surface similarity & adequacy over the preservation of subjective meaning.
- Subtle translation artifacts or lexical changes can completely **distort a speaker's intended position**.

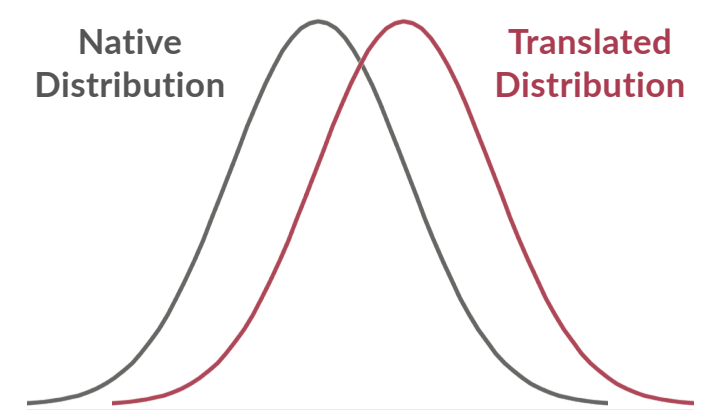
2 THE SOLUTION: Testing for Stance Invariance

- We cast **stance preservation** as an **invariance problem**: translation should not alter the stance conveyed by the source text.
- We adapt two classical **statistical tests** to diagnose systematic opinion shifts introduced by MT systems.

McNemar's Test



Two Proportion Z-Test



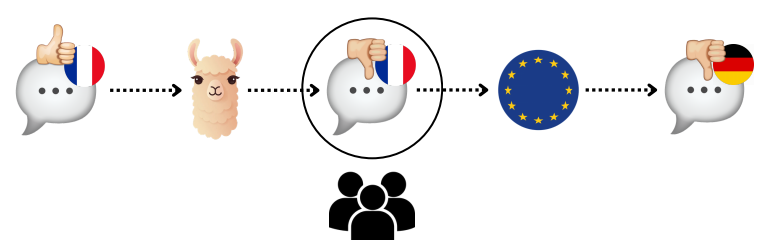
3 CALIBRATION: Validating the Framework

----- Graded, Synthetically-Biased Translation Sets -----

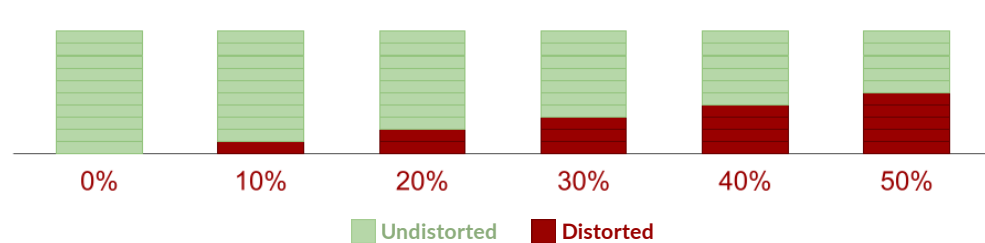
Undistorted ($n=200$)



Distorted ($n=100$)



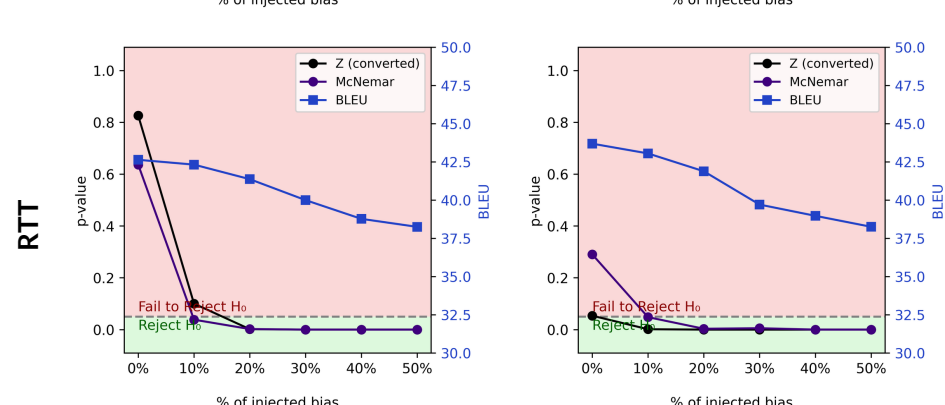
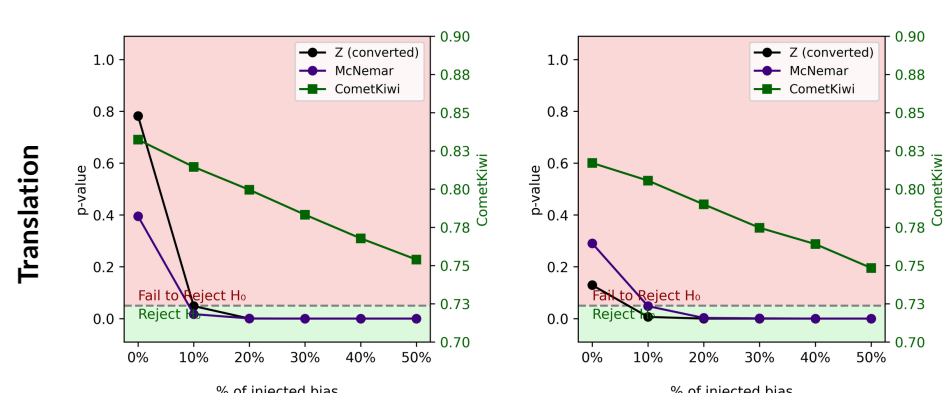
Combination



Experiment

FAVOR

AGAINST



- Both tests are **highly sensitive**, picking up stance shifts from as little as ~10% manipulation.
- BLEU & CometKiwi react on a much **narrower scale** and offer **no indication of the onset** of distortion.

4 APPLICATION: Can Current Models Be Trusted?

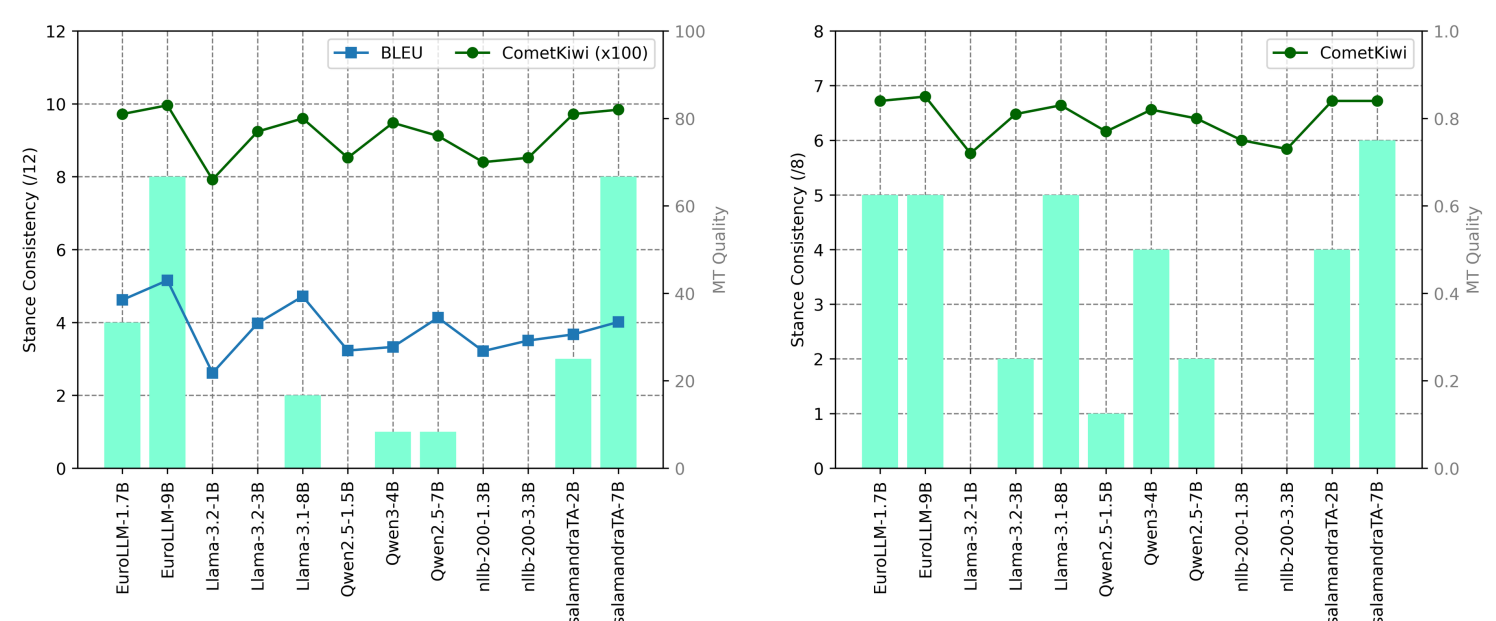
- We evaluate 12 multilingual models across 5 families.

Setup:

Test	Comparison	Source langs	Target/pivot langs	Polarity	Total*
McNemar	Original vs. RTT				12
Two-proportion Z	Original vs. Direct T				8

*Total Tests = $(|L_{src}| \times |L_{tgt/pvt}| - |L_{src} \cap L_{tgt/pvt}|) \times |P|$

Results:



- None of the twelve evaluated systems reliably preserves stance across all tested language directions or polarities.
- Smaller instruction-tuned models are particularly unstable, displaying **significant stance shifts** even when standard MT metrics report high scores.

5

TAKE AWAY

Current multilingual MT systems cannot yet be trusted to faithfully preserve stance.

CONTACT
Nazanin Shafiabadi
Sorbonne University, CNRS, ISIR
nazanin.shafiabadi@isir.upmc.fr

Funded by
bpifrance



PAPER