



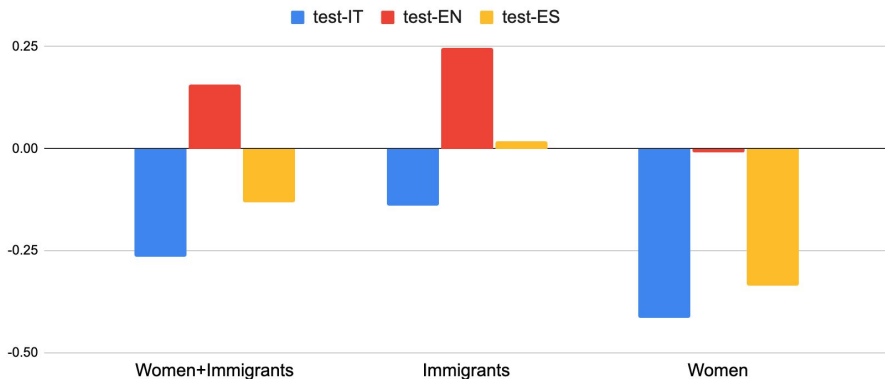
# Exposing the limits of Zero-shot Cross-lingual Hate Speech Detection

Debora Nozza

The lack of hate speech detection models and labeled corpora for non-English languages seems a perfect application for zero-shot, cross-lingual learning, **but is it?**

**Performance does not transfer to different hate speech target types.**

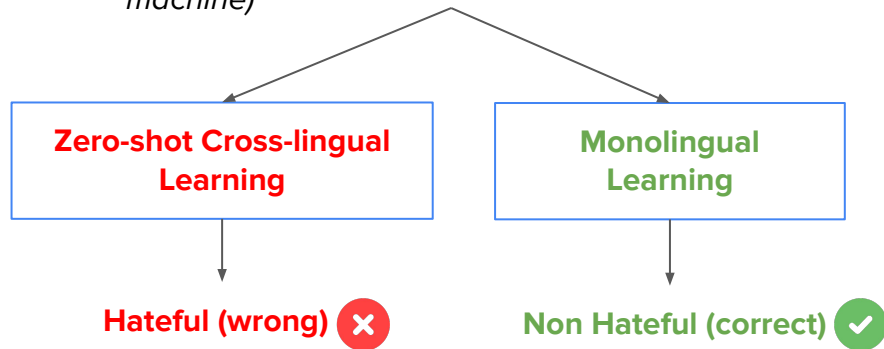
*Difference between Zero-shot Cross-lingual and Monolingual performance*



**They are not able to capture taboo language-specific expressions that contain literal insults but are not hateful per se**

*ma poi come si fa a rompere la lavatrice  
porca puttana (lit. slut bitch)*

*(en. how the hell can you break the washing machine)*



Zero-shot, cross-lingual transfer learning **is not a feasible solution** for hate speech detection, which is **language- and target-specific.**