

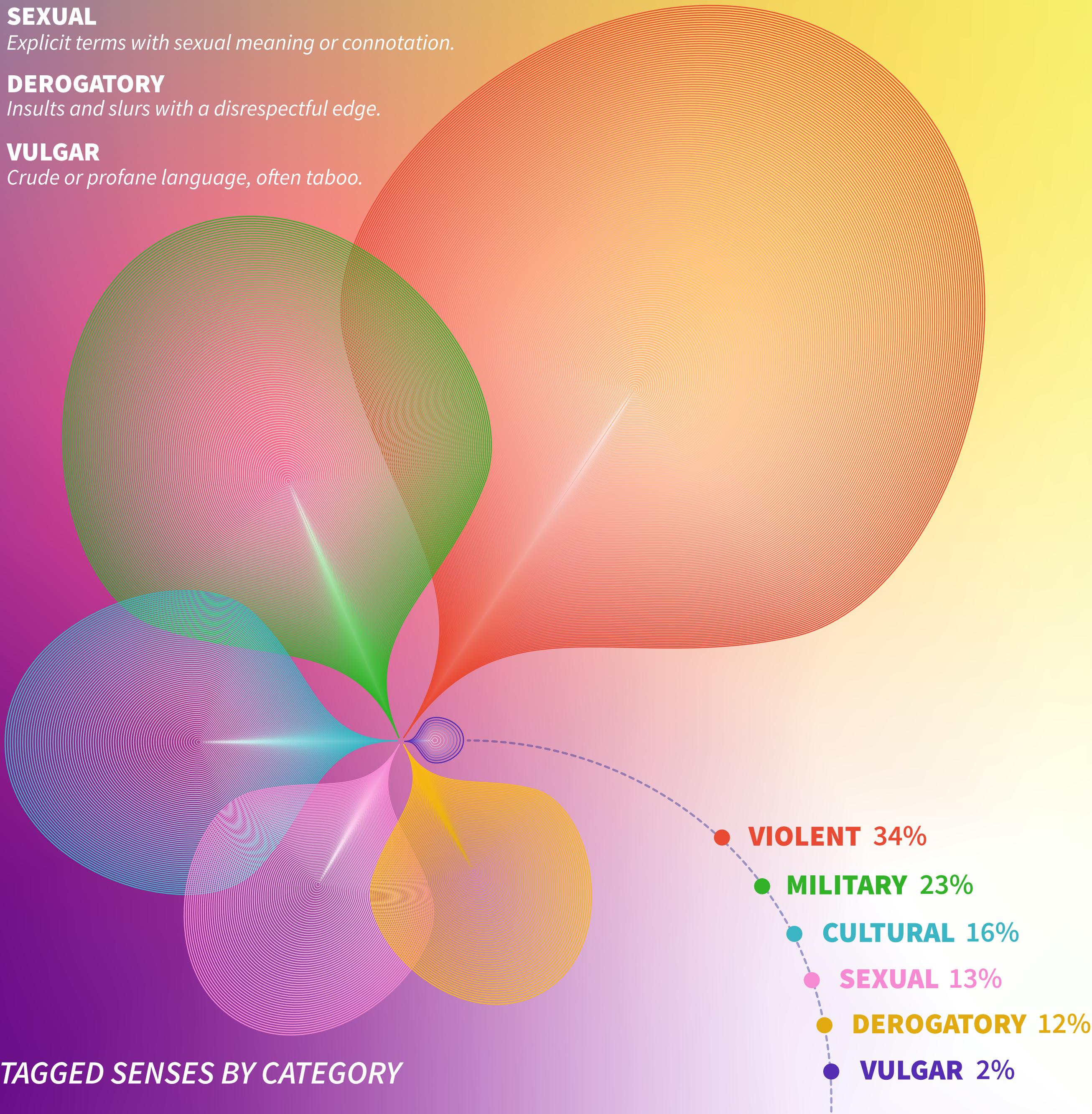
28 languages, 1 goal: Safeguarding AI across cultures

An expanded dataset to map, analyze,
and mitigate linguistic risks in global contexts.

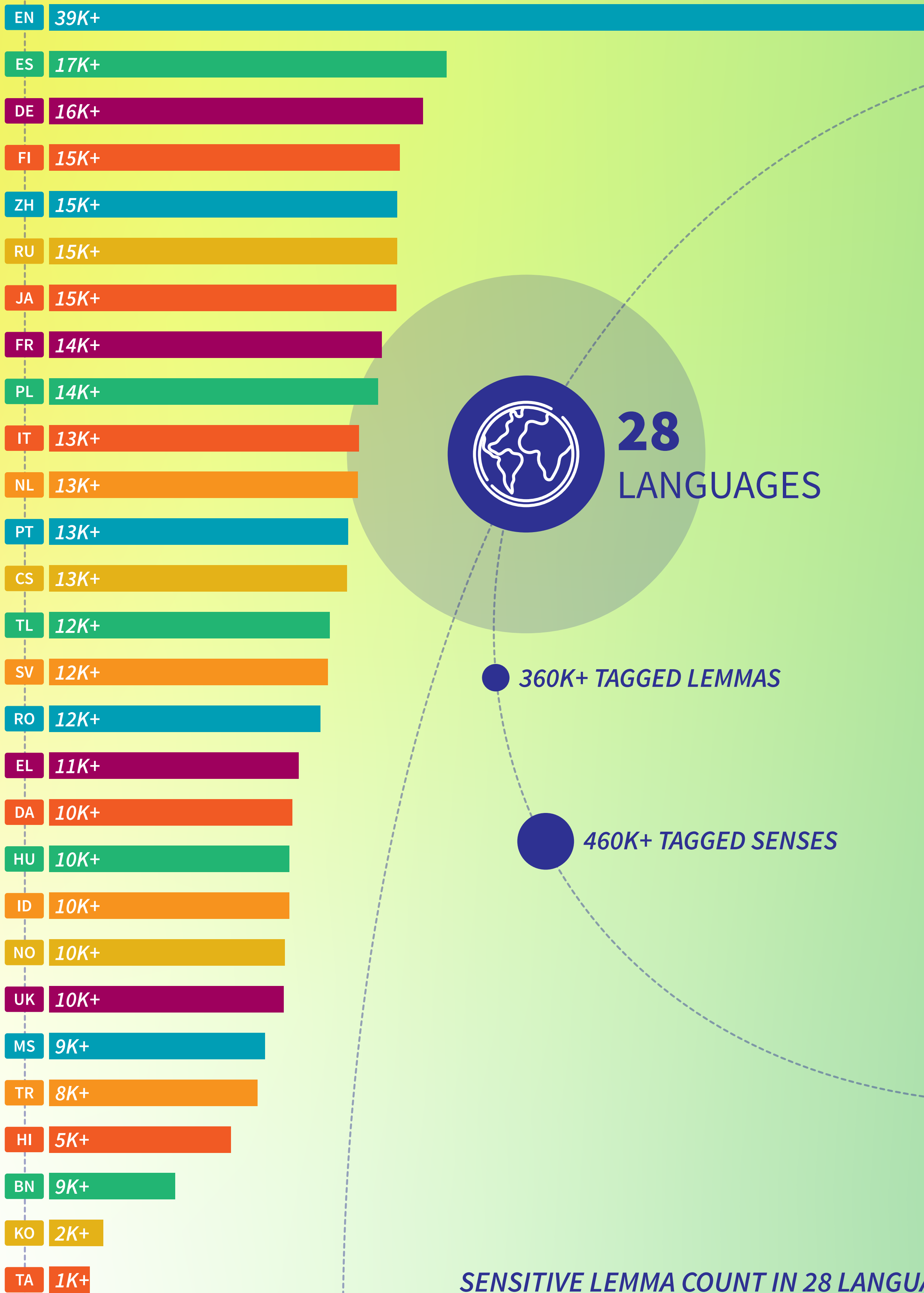


Categories of sensitive concepts

- VIOLENT**
Terms linked to aggression, harm, or historical violence.
- MILITARY**
Words related to defense, armies, warfare, and armed forces.
- CULTURAL**
Context-sensitive terms tied to religion, history, or identity.
- SEXUAL**
Explicit terms with sexual meaning or connotation.
- DEROGATORY**
Insults and slurs with a disrespectful edge.
- VULGAR**
Crude or profane language, often taboo.



TAGGED SENSES BY CATEGORY



28
LANGUAGES

● 360K+ TAGGED LEMMAS

● 460K+ TAGGED SENSES

SENSITIVE LEMMA COUNT IN 28 LANGUAGES