

## **Recommendations on Spanish-Language Disinformation**

We demand social media companies to act immediately on our Spanish-language disinformation action plan:

1. Hire and publicly identify a C-Suite position to oversee U.S. Spanish-language content moderation policy & enforcement.
2. Publicly explain the translation process of the content moderation algorithms, including how the algorithms are trained and audited to ensure proactive detection, reporting, and processing of Spanish-language content.
3. Clarify whether Content Moderators evaluate content in-language, translated, or both. Explain precisely how many US-based Spanish-language content moderators the company employs, as well as the overall number of US-based content moderators. In each case, provide the breakdown of between direct employees and contractors involved in moderation.
4. Publicly share the materials that are used to train content moderators of U.S.-based Spanish-language content.

### **Additional Issue Areas on Spanish-Language Disinformation**

The efforts that social media companies implement to reduce hateful activities and disinformation must be replicated in Spanish and other languages in which they operate. The following issue areas and action steps for social media companies align with our recommendations on matters of enforcement, training, transparency, staff, and governance.

#### Enforcement

- Social media companies should effectively enforce their content moderation policies for all bilingual and Spanish-language content at a commensurate rate of that of English content.

#### Evaluation and Training

- Spanish-language content moderators should be culturally competent in the social and political issues affecting the Spanish speaking community, as well as receive anti-racist training.
- Social media companies should share all content moderation training materials, this includes the Known Questions and Implementation Standards documents.
- Because Spanish-language translation issues are prominent and can be delicate depending on cultural and dialect context, social media companies should regularly audit and test its machine learning algorithm for effectiveness, accuracy, and bias.

#### Transparency

- Social media companies must clarify how it is training its hate speech detection AI to proactively search for content in Spanish.

- Social media companies must clarify how it identifies content in Spanish for content moderation policies or Terms of Service violations, and whether it uses its translation software for these purposes.
- Incorporate all Spanish-language content moderation data into its transparency reports.
- Include Spanish-language content moderation as an integral part to any future independent audit assessments.

#### Governance and Authority

- Publicly identify a C-Suite position, to overlook efforts to moderate hateful activities, misinformation, and disinformation in Spanish in the United States.
- Social media companies should prioritize culturally competent candidates who are fluent in U.S. Spanish dialects for this position.