

Context-Aware Harassment Detection on Social Media

Challenge:

- i. Little past research on **automatic detection of harassment on social media**.
- ii. Existing work on harassment detection predominantly uses machine learning, **relying on message content while ignoring message context**.
- iii. Social networking sites, Facebook, Twitter, and YouTube, have **not** yet developed **effective techniques** against harassment.
- iv. Massive **social and emotional cost**, esp. on young

Solution:

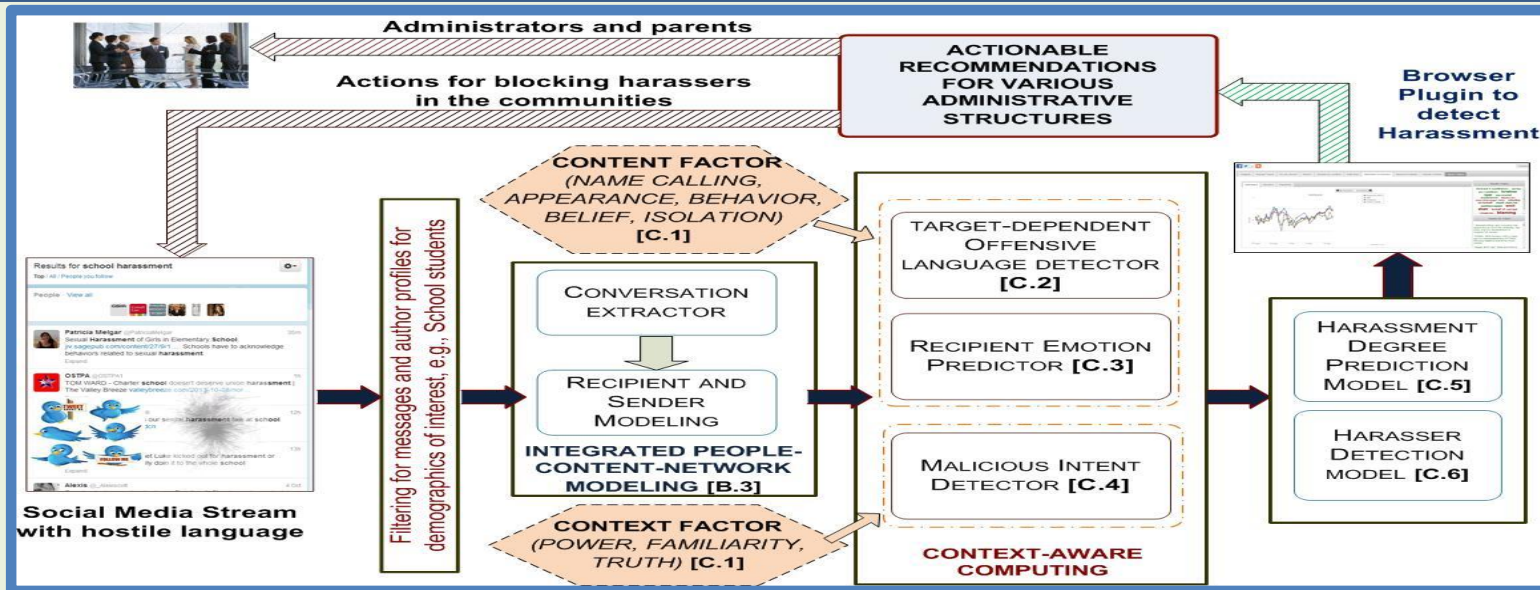
- i. Develop comprehensive and reliable **context-aware techniques** to glean information about the people involved and their interconnected network of relationships, and to determine and evaluate potential harassment and harassers
 - Analyze not just a message, but conversation w/ context; broad set of features
 - Use more diverse techniques: *machine learning, text mining, NLP, and network analysis*

Scientific Impact:

- i. Identification of generic and **target-dependent** language of insult.
- ii. Prediction of harassment-specific **emotion** evoked in a recipient.
- iii. Recognition of **sender's malicious intent**.
- iv. **Harm assessment** of harassing messages.
- v. Detection of **harassing users** from their aggregated behaviors.
- vi. Open source tool for harassment detection

Broader Impact:

- i. Involvement of **social scientists, urban and public affairs professionals**, along with **college and high schools**, to enable broader appreciation of technology's impact on **safe social interactions**.
- ii. The research problems, solutions, and findings inform our **educational activities** including senior student classes, and as in the past, **course projects**.



NSF Award#: CNS 1513721

PI: Prof. Amit P. Sheth

Kno.e.sis, Wright State University