CONVERSATIONAL AGENTS
TO ADDRESS
ABUSIVE ONLINE BEHAVIORS

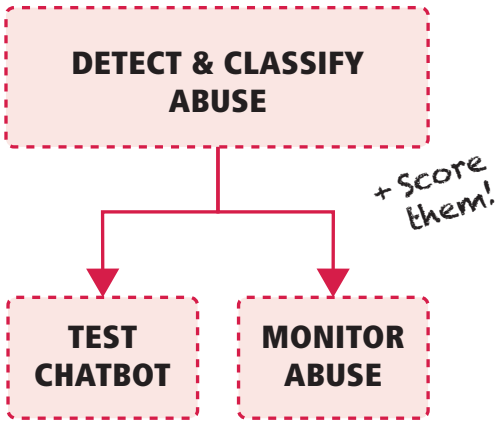
Chatbots for Cyberbullies (in brief)

PROBLEM DESCRIPTION

Emma Beauxis-Aussalet
Amsterdam University of Applied Science
Digital Society School

What to do with cyberbullying ?

Abusive online behaviors occur at a large scale, on all social media, and have dire consequences for their victims. Most interventions focus on victim support. We must do more to address abusers. **Technical solutions remain limited to detecting and hiding abusive comments**, blocking abusers, or replying autimatically with a single message. **The potential of chatbots is unexplored.**



What can we tell abusers ?

Multidisciplinary research is needed to design and evaluate effective dialogues. It requires psychologists and social scientists, not only computer scientists. **We want to develop a framework that supports such research**. We propose 3 initial approaches:

Educate - The dialogues can encourage healthier social skills, introspection of personal issues or insecurities, empathy for the victim, or higher social intelligence.

Deter - The dialogues can explain the legal consequences of harassment and hate speech.

Keep busy - While abusers talk to chatbots, they are not abusing others.

What can go wrong ?

-> Chatbots may add to abusers' motivation.

-> Chatbots may worsen abuses (more frequent and intense).

-> Abusers may build their own chatbots.

-> Abusers may totally ignore chatbots.

-> Chatbots may become abusive (target innocents, or learn & repeat abusive language when personalized).

How to understand abusers ?

Social learning theory - By witnessing abusive peers, abusers acquire mental models where aggression is rewarding.

Coercion theory - If escalating pressure on peers is successful, abuses are deemed rewarding.

Cognitive behavioral theory - Assumptions and interpretations associated with life events are the primary determinants of abusers' behaviors.

Attribution theory - Abusers' perception of hostility may be exaggerated, and lead abusers to respond aggressively.

Is it ok to intervene without asking the victim's consent ? e.g., on Twitter

Is it ok to pretend that the chatbot is a real person ?

Is it ok to impersonate a stereotype of a victim ?

Is it ok to provoke abusers ? e.g. challenge their lack of social intelligence

Emma Beauxis-Aussalet
e.m.a.l.beauxis@hva.nl
@emma_beauxis