AUTOMATED TROLLING AND FAKE NEWS GENERATION IN FUTURE SOCIAL MEDIA: COMPUTATIONAL AND EMPIRICAL INVESTIGATIONS OF THE THREAT AND ITS IMPLICATIONS

MEDIA SOC

Antti Salovaara (Aalto-University)

General description

This project investigates mechanisms by which public discussion – mediated in social media and news websites – may become under attack through recent advances in artificial intelligence technology. In particular, automated approaches to generate seemingly realistic text and photographs have taken tremendous steps in the past ten years. With the help of such technology, it is possible to efficiently conduct targeted, large-scale misinformation campaigns on the internet. This can happen, for instance, through thousands of socially competent conversational AI agents that engage in trolling on social media disguised as humans, or through entire ecosystems of interlinked "news" websites that seem real but ultimately contain mainly disinformation that has been generated by automatic means with very little human effort. When such AI users and synthetic content reach an even higher quality and start to convincingly deceive the majority of human users, they become a threat to the media industry, democracy and ultimately to societal integrity.

The project combines qualitative analysis of trolling strategies and "fake news" articles with modern machine learning methods. Our objective is to devise means and best practices by which individuals and communities can respond to situations in which it is difficult to assess the truthfulness of online content, and the ulterior motives of different discussion participants. This includes technologies that identify trolling and other malicious online behaviour, or computational methods that assess if a text or photograph is AI-generated or real. Reaching these objectives requires a multidisciplinary approach. Hence, the core research team combines expertise in artificial intelligence and machine learning, human-computer interaction, and computational social science.

The project is conducted at the Department of Computer Science, University of Helsinki, with partners from Cornell University (USA), Syracuse University (USA) and City University London (UK).

PROJECT MEMBERS

Antti Salovaara, Lecturer, Aalto University (PI)

Matti Nelimarkka, postdoc, Aalto University

Image: Pond5

CONTACT:

- *Antti Ukkonen*
  antti.ukkonen@helsinki.fi