

УДК 004.9

МЕТОДЫ ВЫЯВЛЕНИЯ ДЕСТАБИЛИЗИРУЮЩИХ СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ

Никольская А.Д. (Университет ИТМО)

Научный руководитель – к.т.н., доцент, Зудилова Татьяна Викторовна

(Университет ИТМО)

Аннотация. В работе исследуется проблема большого количества дестабилизирующих сообщений с Интернетом. Рассматриваются существующие современные методы выявления негативных пользовательских сообщений, а также подходов, которые помогут в снижении их количества.

Введение. На сегодняшний день социальные сети – это самый популярный способ общения, и обмена аудио, видео – контентом. Вследствие такого большого количества активных пользователей можно предположить, что информация, передаваемая в данных сетях, не может быть полностью безопасной. Таким образом, вследствие повсеместного использования социальных сетей, тема выявления дестабилизирующих сообщений в них является актуальной. Следует отметить то, что искусственный интеллект все еще не достиг той стадии, что будет способен блокировать запрещенную информацию в Интернете, а это значит, что данный вопрос все еще остается открытым.

Основная часть. В работе проанализированы виды социальных сетей, а также статистика их использования. Далее выявлены наиболее частые дестабилизирующие пользовательские сообщения: призыву к суициду (а также подобные сообщения), рассылки и спам, кибербуллинг, нелегальная торговля. Основные подходы, которые существуют на данный момент, основаны на методах искусственного интеллекта: сервисы «Аура» и «Окулус» (последний находится еще на стадии разработки). Данные сервисы распознают негативную информацию в сообщениях и звонках, а также в видео и фото-материалах.

Выводы. В работе было проведено исследование популярных социальных сетей и статистики их использования, проанализированы возможные дестабилизирующие пользовательские сообщения в сети Интернет, а затем был сделан анализ существующих методов выявления таких сообщений и способов борьбы с ними. В дальнейшем планируется более тщательный анализ существующих современных методов выявления, а также методов, которые находятся на стадии разработки.

Никольская А.Д. (автор)

Подпись

Зудилова Т.В. (научный руководитель)

Подпись