



Автоматично
Разпознаване на
написана от човек
или генерирана чрез
дипфейк методи
дезинформация в
Социалните медии в
съобщения,
написани на език с
малко ресурси

Проект и цели



- Откриване на целенасочено разпространявана дезинформация по намерението да се разпространява (освен определянето дали става дума за невярно твърдение).
- Прилагане на нови методи за машинна обработка на естествения език за откриване на следи на умишлено разпространявана дезинформация, в текстове, написани от човек или генерирани автоматично за езици с малко езикови ресурси
- В конкретния случай – за български език (нуждата е критична)
- Употреба на интердисциплинарни методи: Изкуствен интелект, Machine Learning, журналистика, лингвистика, психолингвистика, машинна обработка на естествен език
- В съответствие с множество европейски регулации (GDPR, AI Act, Европейския кодекс за поведение във връзка с дезинформацията, Европейската харта за правата на човека, Европейската конвенция за правата на човека).
- Повишаване на чувствителността към темата в обществото чрез образователно събитие (зимно училище)



Очаквани резултати

- Създаване на модели за машинно обучение (ML) за български
- Нови дейтасети за български на здравна и политическа тематика - коронавирус и избори
- Първи **инструмент с Technology Readiness Level 4 за откриване на дезинформация** на български, който да подпомага журналисти и фактчекъри



- **Ръководство за създаване на такива технологии за други езици с малко ресурси**
- Откритие дали съществува дийпфейк дезинформация на български в социалните медии
- Нови психолингвистични и езикови ресурси за българския език
- Повишаване на уменията на обществото в откриването на дезинформация (създадени уебсайт, видео, Twitter, зимно училище, статии в медии, научни статии)



Екипът

★ **Институция:** Институтът GATE – големи данни в помощ на интелигентното общество към СУ „Св. Климент Охридски“).



★ **Департамент интелигентно управление**
★ **Екип:** интердисциплинарен, допълващи се области



Dr. Irina Temnikova (Principal Investigator)
NLP, media bias, Psycholinguistics



Prof. Milena Dobрева
Media and information literacy, Big data

Ruslana Margova

Journalist/news editor

Ph.D. student on the language of Bulgarian fake news



Prof. Ivan Koychev

Deep learning methods for fake content detection,
Machine Learning, Artificial Intelligence



Hristiana Nikolaeva

Bulgarian linguistics, disinformation



Ivo Dzhumerov

Linguistics, Disinformation detection in social media



Silvia Gargova

Deepfake detection, NLP



TRACES

SOFIA UNIVERSITY
ST. KLIMENT OHRIDSKI



AI4media



GATE
big data for smart society