

Towards Synthetic Social Media Data

Justina Mandravickaitė, Milita Songailaitė, Veronika Gvozdovaitė,
Danguolė Kalinauskaitė, Tomas Krilavičius

CARD

CENTRE
FOR APPLIED
RESEARCH
AND
DEVELOPMENT



VYTAUTO
DIDŽIOJO
UNIVERSITETAS
M C M X X I I



UNIVERSITY OF
OXFORD

SYSTEM DESCRIPTION

1 Text generation

2 Network structure

3 Social media site



fakesocial

Fake social network using generated content.



Antonia Keron | a month ago

Not there's though, so tough shit.

□ 41 □ 7



Neth Odgley | a month ago

Lol Kamala wanted to jail students for missing class.

□ 27 □ 9



Pomela Posiorek | a month ago

Che Guevara or the US flag? 10 seconds to answer...

□ 22 □ 7

EVALUATION

Classes of text generation models	BLEU	ROUGE	Perplexity
Republican	0.02	0.87	11.37
Qanon	0.01	0.65	11.95
Democrat	0.03	0.48	11.20
Conservative	0.02	0.32	12.01
Mean	0.02	0.58	11.63

- Our models reached **0,58 mean ROUGE** value and **11,63 mean perplexity score**, indicating good quality models
- We got a **low mean BLEU** score which indicates that generated texts differ in their structure, though they are grammatically correct and meaningful.

Towards Synthetic Social Media Data

Justina Mandravickaitė, Milita Songailaitė, Veronika Gvozdovaitė,
Danguolė Kalinauskaitė, Tomas Krilavičius

CARD

CENTRE
FOR APPLIED
RESEARCH
AND
DEVELOPMENT



VYTAUTO
DIDŽIOJO
UNIVERSITETAS
M C M X X I I



UNIVERSITY OF
OXFORD