



PROPAGANDOS APTIKIMAS IR KLASIFIKAVIMAS SOCIALINĖS MEDIJOS PRANEŠIMUOSE NAUDOJANT HIBRIDINIO GILIOJO MOKYMOSI METODUS

Doktorantė: Ieva Rizgelienė

Vadovas: prof. dr. Darius Plikynas

Mokslo kryptis: N009 Informatika

Studijų laikotarpis: 2022 – 2026 m.

Studijų metai: 2023/2024



Tyrimo objektas, tikslas

Objektas:

- socialinėje medijoje publikuojami pranešimai informacinio karo kontekste ir su jais susijusi informacija;
- hibridiniai giliojo mokymosi algoritmai propagandos požymių aptikimui.

Tikslas: sukurti hibridinį giliojo mokymosi metodą, skirtą propagandos požymių aptikimui socialinės žiniasklaidos pranešimuose lietuvių kalba.

Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2022/2023)	2	1	1	1			
II (2023/2024)	2		1				
III (2024/2025)			1		1		
IV (2025/2026)			1		1		
Iš viso:	4	1	4	1	2		



Ataskaitinių metų darbo planas ir jo įvykdymas

Egzaminai

Planas	Įvykdyta	Būklė
Mašininis mokymas	Egzamino pavadinimas: Mašininis mokymas Data: 2023-03-02	Išlaikytas



Ataskaitinių metų darbo planas ir jo įvykdymas

Dalyvavimas konferencijose

Planas	Įvykdyta	Konferencijos tipas
13th Data analysis Methods for Software Systems Conference, 2022 12 02. Lietuva, Druskininkai	Pristatymo pavadinimas: Automated Propaganda Detection Using Deep Learning Methods: a Pilot Systematic Review Research Autoriai: Rizgelienė, I., Plikynas, D. Konferencijos pavadinimas: 13th Data analysis Methods for Software Systems Conference Data: 2022 metų gruodžio 1-3 dienomis Vieta: Lietuva, Druskininkai	Nacionalinė
Skaitmeninės humanitarikos seminaras, VU, 2022 11 23. Lietuva, Vilnius	Pristatymo pavadinimas: Propogandos automatinio atpažinimo mašininio mokymo metodais tyrimas ir socialinio poveikio modeliavimas. Autoriai: Ieva Rizgelienė, Darius Plikynas. Seminaro pavadinimas: Skaitmeninės humanitarikos seminaras. Data: 2022 metų lapkričio 23 diena. Vieta: Lietuva Vilnius	Nacionalinė (seminaras)

Visų mokslinių tyrimų ir disertacijos rengimo etapai (1 pusmetis)

Darbo pavadinimas	Atlikimo terminai	Pastabos
<p>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</p> <ol style="list-style-type: none"> 1. Disertacijos tyrimo objekto detalizavimas. 2. Gilių neuroninių tinklų mokymu grįstų metodų apžvalga kompiuterinės lingvistikos ir natūralios kalbos apdorojimo kontekste, siekiant aptikti propagandos požymius. 3. Metodų apžvalgos palyginimas, apibendrinimas ir pateikimas disertacijos analitinės dalies aprašyme. 4. Tyrimo tikslo suformavimas. 	<p>2022 m. lapkričio mėn. - 2023 m. vasario mėn.</p>	<p>1.1. Įvykdyta.</p> <p>1.2., 1.3. Įvykdyta. Kartu su vyresn. m. darbuotojais</p> <p>Dariumi Plikynu ir Gražina Korvel ruošiamas sisteminės apžvalgos straipsnis.</p> <p>1.4. Įvykdyta.</p>

Propagandos kaip tyrimo objekto sąvokos apibrėžiamas

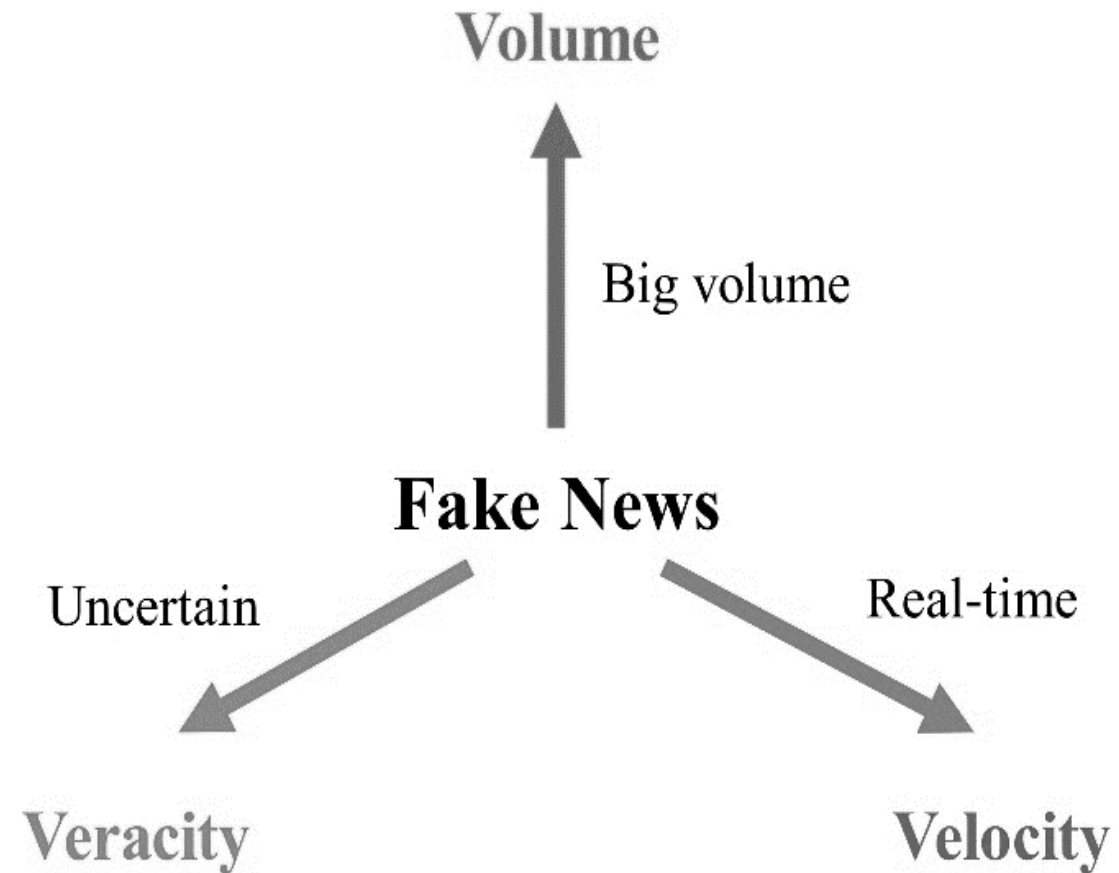
Apibrėžta propagandos socialinės medijos pranešimuose sąvoka, kurios metu nutarta, kad šiame projekte bus tirama **propaganda ir dezinformacija socialinės medijos pranešimuose, informacinio karo kontekste.**

- NATO apibrėžia informacinį karą kaip operaciją, vykdomą siekiant įgyti informacinį pranašumą prieš priešininką. (org. information warfare is an operation conducted in order to gain an information advantage over the opponent.)
- „Oxford“ žodyne informacinis karas apibrėžiamas kaip konflikto forma, kai siekiama užgrobti, susilpninti arba sunaikinti priešo priemones, skirtas duomenims rinkti, analizuoti ir trikdyti. (org. a form of conflict in which the objective is to capture, degrade or destroy the enemy’s means of gathering, analysing, and disturbing data.)
- „Cambridge“ žodyne propaganda apibrėžiama kaip informacija, idėjos, nuomonės ar vaizdai, dažnai pateikiantys tik vieną argumento dalį, kurie transliuojami, publikuojami ar kitaip skleidžiami siekiant paveikti žmonių nuomonę (org. *information, ideas, opinions, or images, often only giving one part of an argument, that are broadcast, published, or in some other way spread to influence people’s opinions*).

Informacinio karo sąvoka yra platesnė nei propagandos, propaganda tai vienas iš informacinio karo įrankių.

Propagandos socialinėje medijoje charakteristikos

- **Kiekis (Volume).** Socialinėje medijoje bet kas gali skleisti bet kokia informaciją, todėl propagandos kiekis socialinėje medijoje yra didelis.
- **Tikrumas (Veracity).** Socialinėje medijoje informacija skleidžiama įvairiausiomis temomis, todėl dažnu atveju yra sunku patikrinti skleidžiamos informacijos patikimumą.
- **Greitis (Velocity).** Propagandai naudojamos konkrečiu momentu visuomenei aktualios temos, todėl propaganda sklinda greitai, o jos kūrėjai dažnu atveju yra trumpalaikiai socialinės medijos vartotojai.



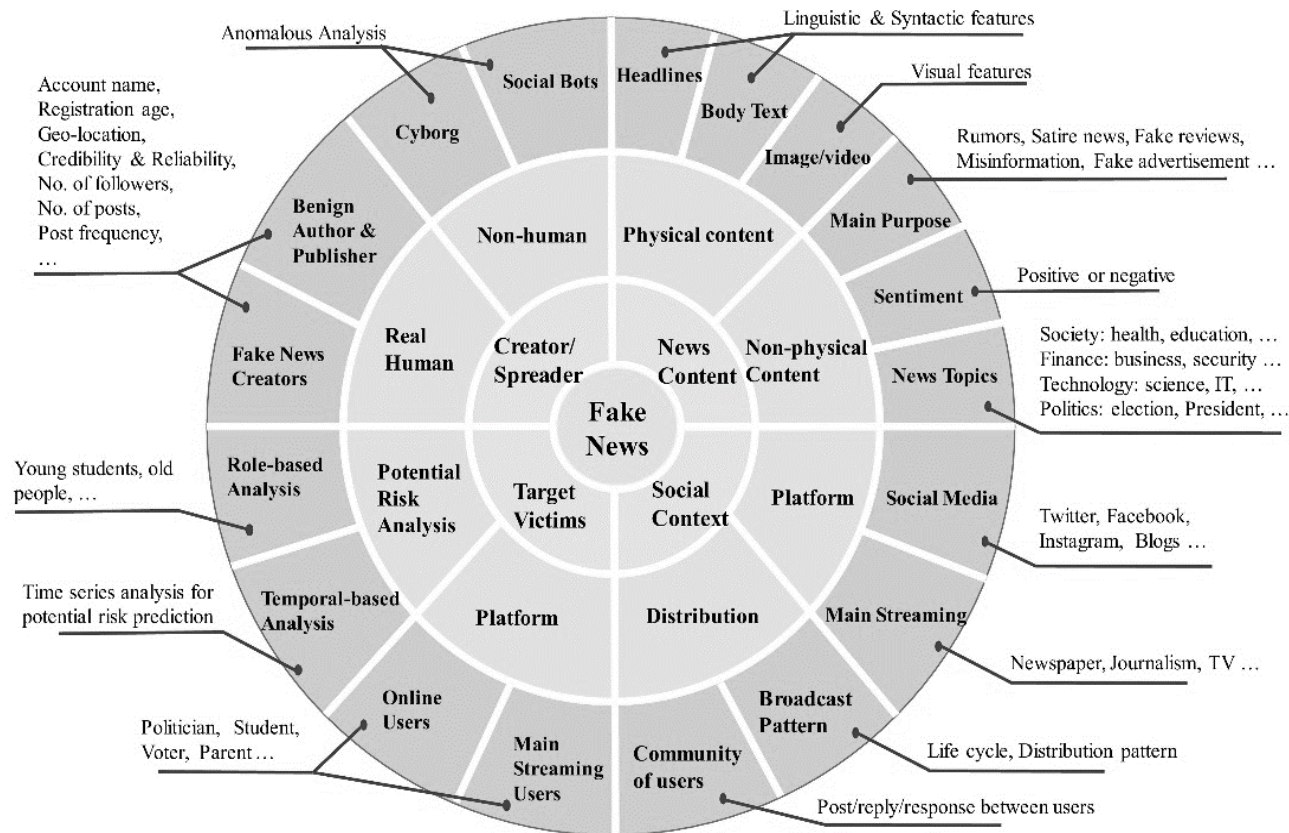
1 pav. Propagandos socialinėje medijoje kiekis, tikrumas ir greitis [ZhG20]

Propagandos tipai

#	Technique	Snippet
1	Loaded language	Outrage as Donald Trump suggests injecting disinfectant to kill virus.
2	Name calling, labeling	WHO: Coronavirus emergency is ' Public Enemy Number 1 '
3	Repetition	I still have a dream . It is a dream deeply rooted in the American dream . I have a dream that one day ...
4	Exaggeration, minimization	Coronavirus ' risk to the American people remains very low ', Trump said.
5	Doubt	Can the same be said for the Obama Administration?
6	Appeal to fear/prejudice	A dark, impenetrable and "irreversible" winter of persecution of the faithful by their own shepherds will fall.
7	Flag-waving	Mueller attempts to stop the will of We the People!!! It's time to jail Mueller.
8	Causal oversimplification	If France had not have declared war on Germany then World War II would have never happened.
9	Slogans	"BUILD THE WALL!" Trump tweeted.
10	Appeal to authority	Monsignor Jean-Francois Lantheaume, who served as first Counsellor of the Nunciature in Washington, confirmed that "Vigan said the truth. That's all."
11	Black-and-white fallacy	Francis said these words: "Everyone is guilty for the good he could have done and did not do ... If we do not oppose evil, we tacitly feed it."
12	Thought-terminating cliché	I do not really see any problems there. Marx is the President.
13	Whataboutism	President Trump — who himself avoided national military service in the 1960's— keeps beating the war drums over North Korea.
	Straw man	"Take it seriously, but with a large grain of salt." Which is just Allen's more nuanced way of saying: "Don't believe it."
	Red herring	"You may claim that the death penalty is an ineffective deterrent against crime – but what about the victims of crime? How do you think surviving family members feel when they see the man who murdered their son kept in prison at their expense? Is it right that they should pay for their son's murderer to be fed and housed?"
14	Bandwagon	He tweeted, "EU no longer considers # Hamas a terrorist group. Time for US to do same."
	Reductio ad hitlerum	"Vichy journalism," a term which now fits so much of the mainstream media. It collaborates in the same way that the Vichy government in France collaborated with the Nazis.

2 pav. Propagandos tipai [MBW20]

Propagandos socialinėje medijoje charakteristikos



3 pav Propagandos socialinėje medijoje charakteristikos [ZhG20]

Kūrėjai/Skleidėjai. (Creator/Spreader)

Pagrindiniai propagandos kūrėjai realūs **žmonės (human)**, sąmoningai skleidžiantis dezinformaciją. Ne žmonės (**non – human**), tai kompiuteriniai algoritmai, sukurti taip, kad pasižymėtų į žmogų panašiu elgesiu ir skleistų dezinformaciją socialinėje medijoje.

Tikslinės aukos. (Target victims). Atsižvelgiant į propagandos tikslus, nustatomi konkretūs taikiniai, pvz. konkrečios socialinės grupės.

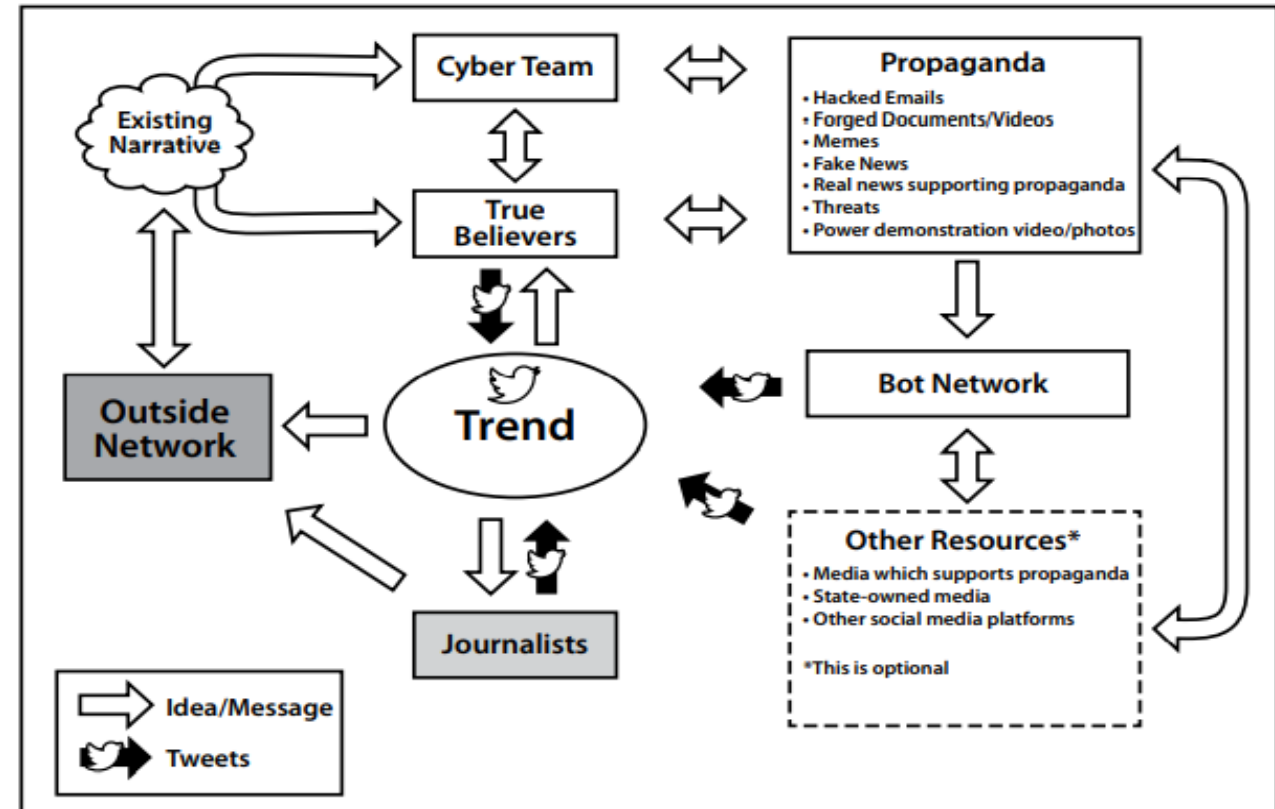
Turinys. Antraštė, pagrindinis tekstas, vaizdinė informacija.

Socialinis kontekstas. Socialinio tinklo analizė (social network analysis). Kaip sklinda propaganda ir kaip vartotojai tarpusavyje sąveikauja.

Organizuota propaganda socialinėje medijoje

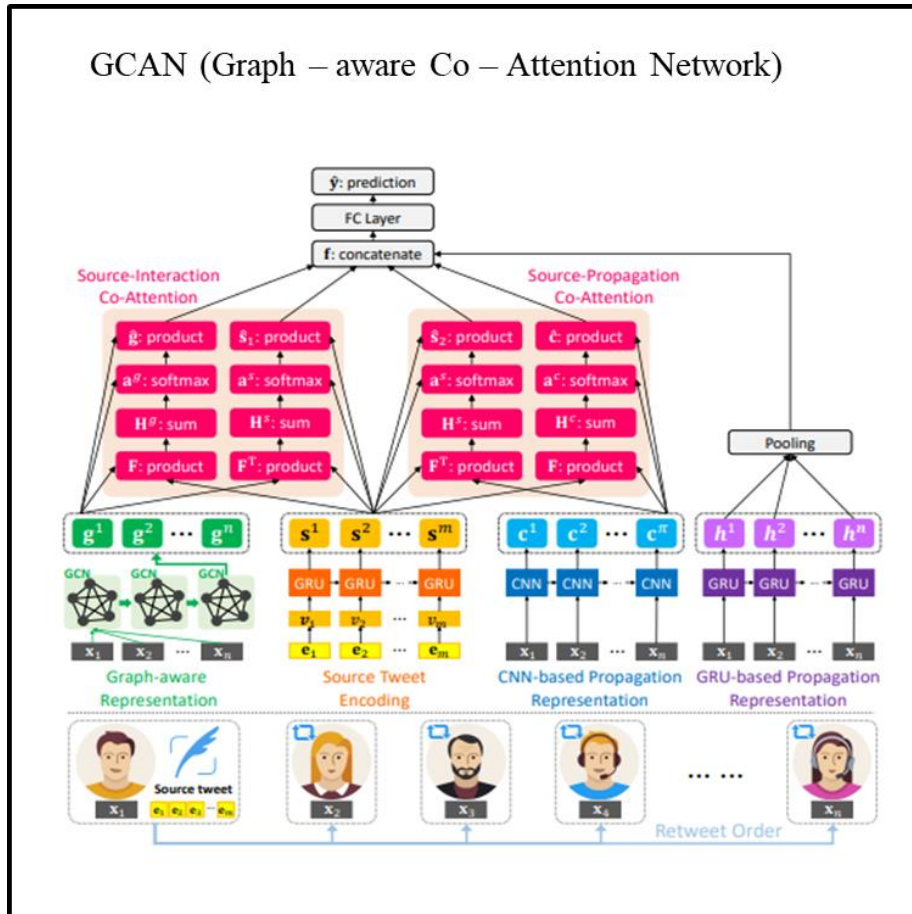
Pagrindiniai propagandos socialinėje medijoje sklaidimo keliai:

- žinutė, atitinkanti realiu laiku esantį naratyvą;
- grupė žmonių, linkusių patikėti, norima skleisti informacija;
- kibernetinė komanda (angl. *cyber team*), kuri yra atsakinga už propagandos sklaidą (paruošia informacija, nustato taikinius ir t.t.);
- netikrų paskyrų tinklas (ang. *bot network*).

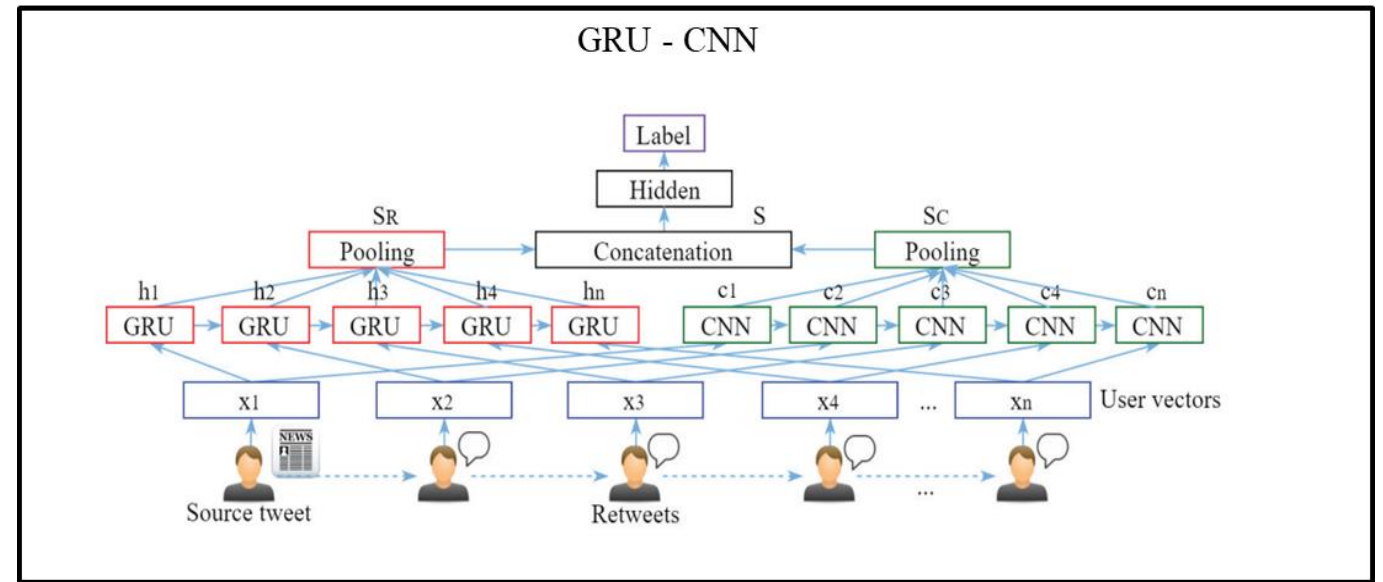


4 pav. Propagandos sklaidimo procesas [Pri17]

Hibridinių giliojo mokymo metodų naudojimas propagandos aptikimui, socialinėje medijoje



5 pav. GCAN modelis [LuL20]



6 pav. GRU - CNN modelis [LiW18]



Modelių palyginimas

Model	Dataset	User features	Short text	Content	Visual	Social context	Models used	Accuracy
GCAN	Twitter15	yes	yes			yes	CNN, RNN, GCNN, MLP	0.8767
	Twitter16	yes	yes			yes		0.9084
TABML	Twitter 15	yes	yes			yes	GCAN, BERT, Baeyesian optimiazion	0.8565
	Twitter 16	yes	yes			yes		0.7565
MCAN	Twitter		yes		yes		VGG-19 (textual info from images), CNN (image), BERT	0.809
	Weibo			yes	yes			0.899
HCSN	Twitter 15		yes	yes			GRU, gumbel - max trick method,	0.912
	Twitter 16		yes	yes				0.897
Pre-CoFact	Factify		yes	yes	yes		DeiT, DeBERTa, MLP	0.74585
GAME-ON	Twitter			yes	yes		GNN, RCNN, BERT,	0.958
	Weibo			yes	yes			0.899
MML	CCMR		yes		yes		BERT, CNN, MLP	0.9225
DEAP-FAKED	KFN-UB		yes				biLSTM, KG embedding	0.9
	CoAID-UB		yes					0.78
GCN- GAT	FakeNewsNet	yes	yes			yes	social graph, GloVe, GraphSAGE, GAT	0.913
		yes	yes			yes		
BDCoNN	Weibo	yes	yes			yes	BERT, CNN, TextCNN, BiGRU	0.957
	CED	yes	yes			yes		0.946
ZoKa	GossipCop	yes	yes	yes		yes	GAT, EGAT, ELECTRA,	0.802
	PolitiFact	yes	yes	yes				0.874
AMFB	Weibo			yes	yes		ABS-BiLSTM, ABM-CNN-RNN, MLP	0.832
	Twitter			yes	yes			0.883
PPC_RNN+CNN	Weibo	yes						0.921
	Twitter15	yes						0.842
	Twitter16	yes						0.863
Bi -GCN	Twitter15	yes					GCN	0.886
	Twitter16	yes						0.88
UFD (Unsupervised)	LIAR	yes	yes				Bayesian network, Gibbs sampling approach	0.759
	BuzzFeed	yes	yes					0.679
MT-ES	LIU+ (rumor detection)		yes			yes	RNN	

1.lentelė. Modelių palyginimas



Rezultatai

- Tam tikros vartotojų charakteristikos gali parodyti, kokie kokie vartotojai yra labiau linkę pasikliauti ir skleisti propagandą socialinėje medijoje.
- naudojant viešai prieinamus duomenų rinkinius (pvz. Twitter, Weibo, GossipCop, Factify) ir skirtingo pobūdžio duomenis (pvz. vartotojų charakteristikos ir tekstinis turinys, vaizdinė informacija ir tekstinis turinys, socialinis kontekstas ir tekstinis turinys ir t.t.), bei hibridinius neuroninių tinklų modelius, t.y. dviejų ar daugiau neuroninių tinklų modelių junginius, kurie yra naudojami apdoroti skirtingų tipų informaciją (pvz. CNN (apdoroti vaizdinę informaciją) + RNN (apdoroti tekstinę informaciją) +GCN (apdoroti socialinį kontekstą) +MLP (apjungti gautų modelių rezultatus ir klasifikuoja), galima aptikti ar tiriamas tekstas ir tiriamos sąsajos su tekstu, turi propagandos požymių; taip pat pastebėta, kad būtent tokių modelių tikslumas yra didžiausias lyginant su modeliais naudojančiais tik vieno tipo duomenis (pvz. tik tekstinius arba tik vaizdinius).

Kito pusmečio darbo planas

Mokslinio tyrimo vykdymas:	
2.1. Tyrimo metodikos sudarymas:	
2.1.1. Uždavinių, skirtų tyrimo tikslui pasiekti, suformulavimas.	2023 m. kovo mėn. –
2.1.2. Tyrimo metodikos išsikeltiems uždaviniams spręsti parinkimas.	2023 m. rugsėjo mėn.
2.1.3. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.	

Planuojama veikla		Data
1.	Dalyvavimas konferencijoje Lietuvoje.	2023 m. rugsėjo mėn.
1.	Apžvalginių tyrimų rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.	2024 m. rugsėjo mėn.
1.	Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.	2025 m. birželio mėn.
1.	Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.	2026 m. birželio mėn.

Kartu su vyresn. m. darbuotojais: Dariumi Plikynu ir Gražina Korvel paruošti sisteminės apžvalgos straipsnį.

Ačiū už dėmesį.