

Borse di studio GARR
Orio Carlini

TRADISAN – conTRAstare la Disinformazione in ambito SANitario tramite fake news detection sui social media

Luca Giordano

Tutor: Maria Pia di Buono

UNIOR NLP Research Group
Università degli Studi di Napoli
“L’Orientale”



- Notizie medico-sanitarie online (social media)
- Scarsa qualità e disinformazione → rischi
- Prevenzione della disinformazione in Italia

Risultati primo anno [1]

1 – Risorsa linguistica italiana annotata e multimodale, specifica per il dominio medico-sanitario

ID	Fonte	Data	Titolo	Testo	Immagini	URL
3762	la Repubblica	18/04/2019	Fagioli e spinaci tengono lontano il tumore della vescica	...	Immagine_1.jpg Immagine_2.jpg ...	...
3763	la Repubblica	17/04/2019	Buona Pasqua a tavola, anche per il vostro cuore	...	Immagine_1.jpg Immagine_2.jpg ...	...
3764	la Repubblica	10/04/2019	Legumi, soia e vegetali: il giusto apporto proteico per proteggere il cuore	...	Immagine_1.jpg Immagine_2.jpg ...	...

Risultati primo anno [2]

2 - Modello testuale basato su set di feature

Categoria	# (Totale: 31)	Esempi
Stilometriche	11	uppercase_word_w consecutive_excla_count ellipses_count typo_count_w
Lessicali	15	superl_count_w rev_hurtlex_count_w buzzwords_count_w
Sentiment	5	opos - oneg nrc_anger_w nrc_trust_w nrc_joy_w

Risultati primo anno [3]

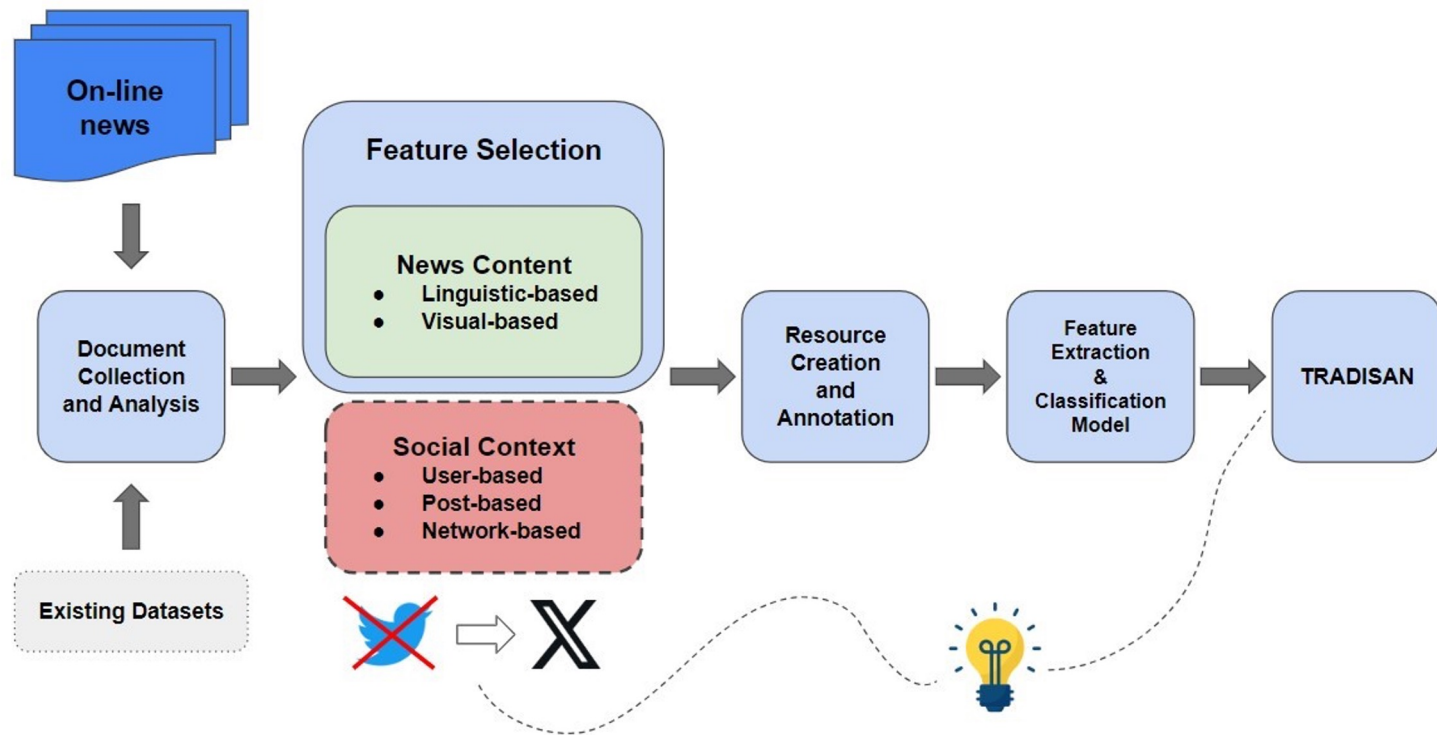
		38 1999 - 2023				31 Feature			Annotazioni aggiuntive				
		id	source	date	url	headline	char_count	...	onег	lemmas	pos	ioб	ner
NON ATTENDIBILI	ATTENDIBILI	0	...	...	...	...	...	...	...	...	...	...	...
		...	...	...	...	...	...	...	...	...	...	...	...
		9972	...	...	...	...	...	...	...	...	...	...	...
	ATTENDIBILI	9973	...	...	...	...	...	...	...	...	...	...	...
		...	...	...	...	...	...	...	...	...	...	...	...
		32100	...	...	...	...	...	...	...	...	...	...	...

Risultati primo anno [4]

3 - Valutazione granulare e interpretabile dell'attendibilità dei titoli delle notizie (+ 1 pubblicazione a riguardo), scalabile per corpo del testo

Model	Classifier	P _{MacroAVG}	R _{MacroAVG}	F1
<i>All Features</i>	RandomForest	0.70	0.70	0.70
<i>Top13 Features</i>	RandomForest	0.68	0.68	0.68
<i>fastText</i>	LinearSVC	0.76	0.76	0.76
<i>fastText</i> + <i>All Features</i>	LinearSVC	0.78	0.78	0.78
<i>fastText</i> + <i>Top13 Features</i>	LinearSVC	0.79	0.79	0.79
<i>GloVe</i>	LogisticRegression	0.78	0.78	0.78
<i>GloVe</i> + <i>All Features</i>	LogisticRegression	0.79	0.79	0.79
<i>GloVe</i> + <i>Top13 Features</i>	LogisticRegression	0.79	0.78	0.79
<i>BERT_{BASE}</i>	Multi-Layer Perceptron	0.855	0.855	0.855
<i>BERT_{BASE}</i> + <i>All Features</i>	Multi-Layer Perceptron	0.871	0.871	0.871
<i>BERT_{BASE}</i> + <i>Top13 Features</i>	Multi-Layer Perceptron	0.887	0.884	0.884

Modifiche pipeline originale [1]



Modifiche pipeline originale [2]



stance + social engagement analysis

Raccolta
e pulizia
dati



Annotazione
automatica



Analisi

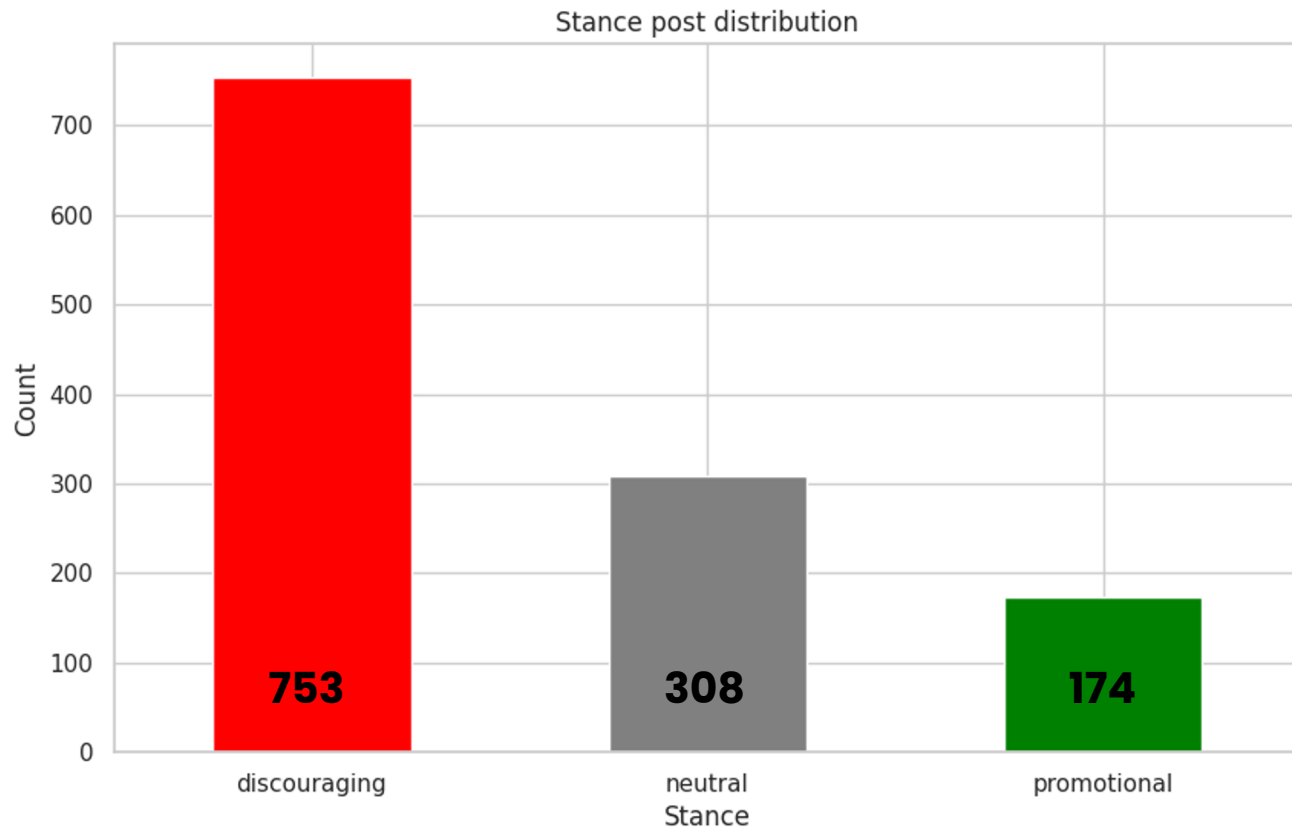
gab

1235 post
Argomento : misure
di contenimento
della pandemia
COVID-19

stance
=

pro
neutrale
contro

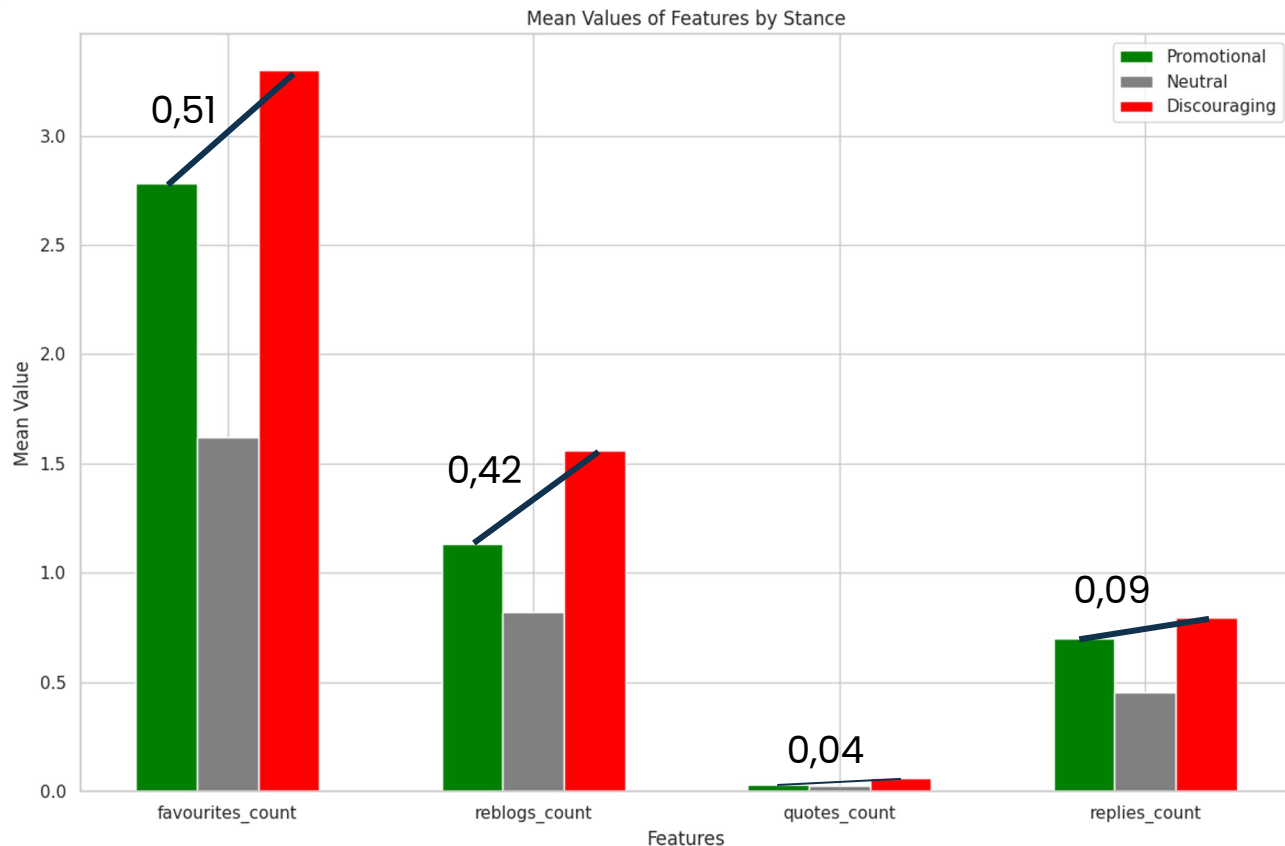
Modifiche pipeline originale [3]



Modifiche pipeline originale [4]

ANOVA

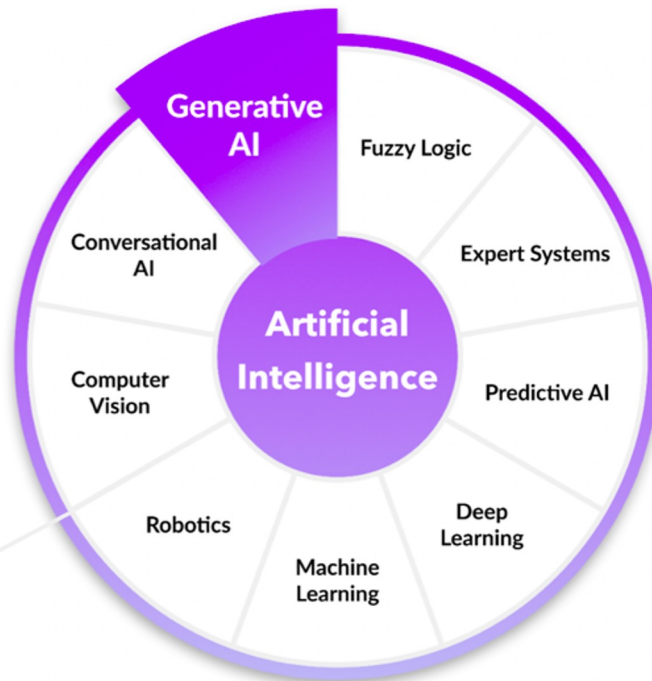
Feature	F-Score
# Mi Piace	14,5
# Condivisioni	7,6
# Citazioni	2,5
# Commenti	1,9



“Enhancing Journalism through Large Language Models”

What is Generative AI in context of AI Pie

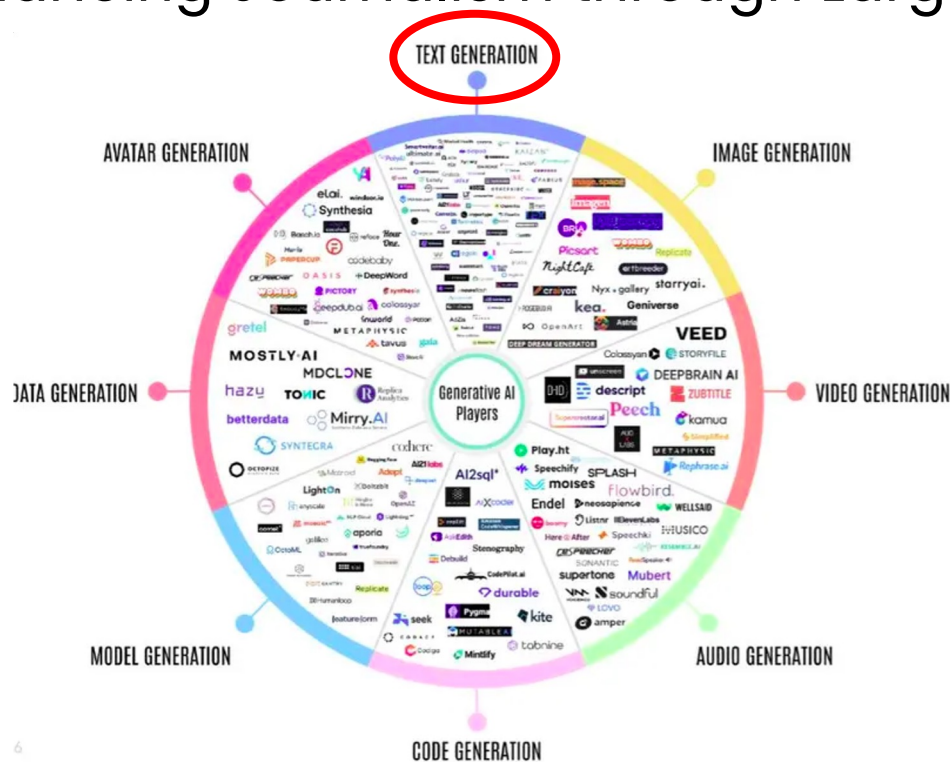
AI definition:
(N) Technology that can perform
tasks traditionally require
human intelligence



12

Secondo anno

“Enhancing Journalism through Large Language Models”



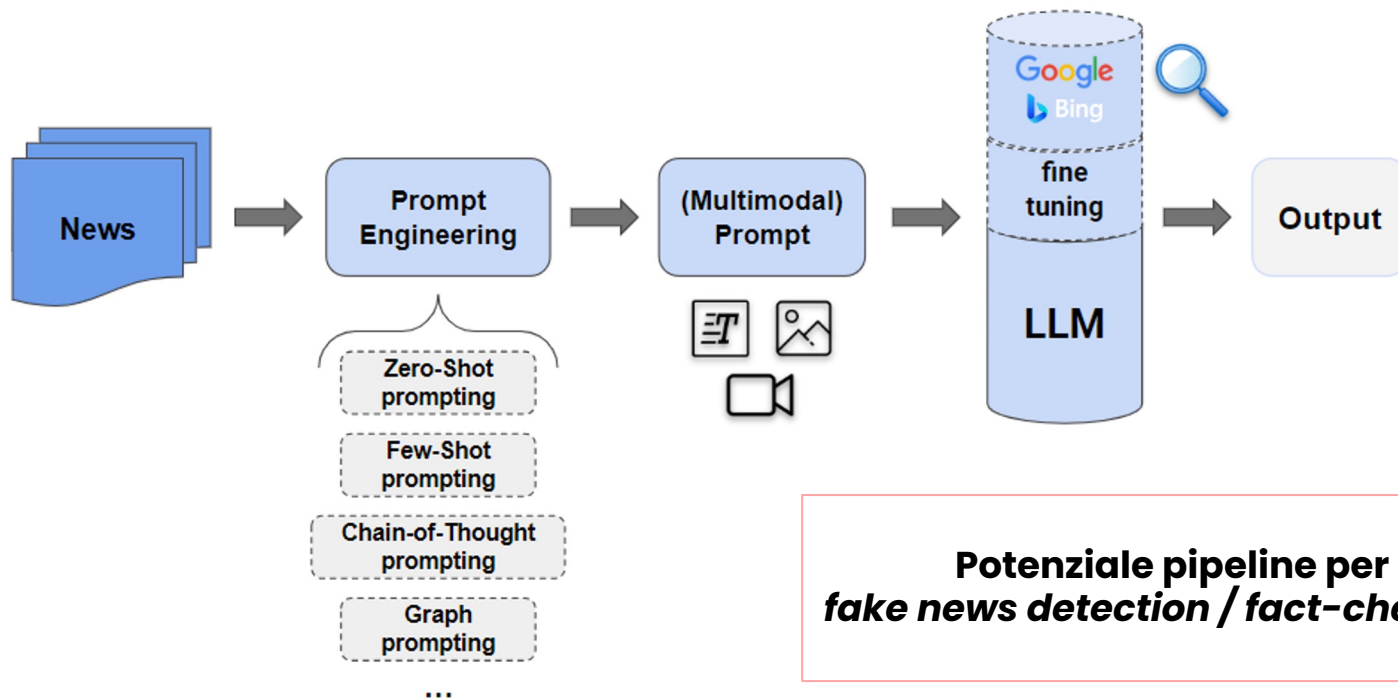
Secondo anno

“Enhancing Journalism through Large Language Models”



Secondo anno

“Enhancing Journalism through Large Language Models”



Borse di studio GARR
Orio Carlini

Grazie per l'attenzione!

Luca Giordano

Tutor: Maria Pia di Buono

UNIOR NLP Research Group
Università degli Studi di
Napoli "L'Orientale"

