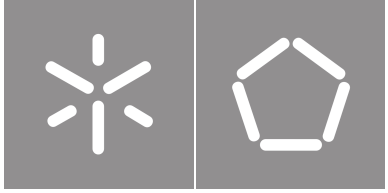


Universidade do Minho
Escola de Engenharia

Maria Laura de Araújo Barbosa

Post generator for social media based on Emotions and Personality

October, 2022



Universidade do Minho
Escola de Engenharia

Maria Laura de Araújo Barbosa

Post generator for social media based on Emotions and Personality

Master Dissertation
Master in Informatics Engineering

Work developed under the supervision of:
Paulo Novais

October, 2022

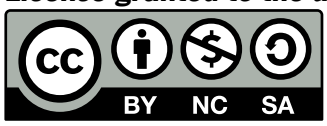
COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositoriUM of Universidade do Minho.

License granted to the users of this work



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

CC BY-NC-SA 4.0

https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en

Acknowledgements

O desenvolvimento desta dissertação não teria sido tão gratificante sem o envolvimento de um conjunto de pessoas, a quem estou profundamente grata. Um agradecimento especial:

Ao meu orientador, Professor Paulo Novais, pelo sentido prático com que sempre me orientou. Obrigada pela confiança, respeito e compressão ao longo do trabalho.

Ao Professor Francisco S. Marcondes, por toda a atenção e apoio, disponibilidade, orientação, colaboração e rigor científico. Todos os seus conselhos e sugestões foram determinantes para a concretização deste trabalho.

A Professora Dalila Durães pela simpatia e colaboração no desenvolvimento científico.

Ao Doutor Adelino Gala pela sua disponibilidade em partilhar conhecimento na elaboração do léxico mental.

A todos os professores que me acompanharam ao longo da licenciatura e Mestrado. Em especial, ao Professor Pedro Rangel Henriques e à Professora Cristiana Araújo, pela simpatia, amabilidade e ensinamentos dados nos vários projetos em que trabalhamos juntos.

Aos meus colegas de curso, Carolina, Constança, Filipa, Filipe, Renata e Vasco, pela paciência, apoio e amizade.

A minha mãe, Conceição Araújo, por ser um exemplo de coragem e me incentivar a ultrapassar os desafios que foram surgindo.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Braga, October 25, 2022

(Place) (Date)

Assinado por: **Maria Laura de Araújo Barbosa**
Num. de Identificação: 15800901
Data: 2022.10.25 19:17:17 +0100

(Maria Laura de Araújo Barbosa)

Abstract

Post generator for social media based on Emotions and Personality

sSocial networks are currently one of the main sources of fake-news dissemination. On the other hand, fact-checking agencies, which emerged with the aim of solving the problem of fake-news, find it difficult to spread their content on social media. The engagement of fake-news is considerably greater than that of fact-checking.

This dissertation contributes to defining a set of heuristics applicable to fact-checking posts in microblog environments, in order to increase their reach and engagement with the reader. The systematic review, inspired by the guidelines defined by Kitchenham, made it possible to identify the main strategies used in the dissemination of fake-news and fact-checkings on social media. The results showed that the dissemination and viralization of posts depends on two aspects: the dissemination strategy and the engagement strategy created with the reader. The first is outside the scope of study in this document. Regarding the second, the systematic review showed that the inclusion of emotions and personality in social media posts is an efficient strategy to improve reader engagement. Furthermore, engagement seems to depend on the type of language and elements that make up the post. Since we are working in the context of fake-news in which ethical limits are frequently tested and extrapolated, it is important to reinforce that the defined approach is ethically valid.

In this sense, the course of the dissertation continued with the design and development of a Post Generator Algorithm based on emotions and personality capable of increasing the engagement of fact-checking posts. The algorithm was tested in an experiment carried out with twenty participants, where patterns were searched between the posts obtained by the algorithm and those created by the fact-checking journal Snopes and the interactions achieved in each of them. This experiment, carried out in a lab environment, proved the hypothesis that emotions and personality in Microblog posts increase user engagement with fact-checking posts.

Keywords: Natural Language Processing, Post Generator, Social Networks, Twitter, Fake-news, Fact-checking, Emotions, Personality, Engagement

Resumo

Gerador de posts para redes sociais com base em Emoções e Personalidade

As redes sociais são, atualmente, uma das principais fontes de difusão de fake-news. Por outro lado, as agências fact-checking, que surgiram com o objetivo de colmatar o problema das fake-news, têm dificuldade em difundir o seu conteúdo nas redes sociais. O alcance das fake-news é consideravelmente superior ao dos fact-checkings.

Esta dissertação contribui para definir um conjunto de heurísticas aplicáveis aos posts de fact-checking em ambientes microblog, com o objetivo de aumentar o seu alcance e envolvimento com o leitor. A revisão sistemática, inspirada nas guidelines definidas por Kitchenham, permitiu identificar as principais estratégias utilizadas na difusão de notícias falsas ou de verificação de conteúdo nas redes sociais. Os resultados mostraram que a difusão e viralização de posts depende de duas vertentes: a estratégia de disseminação e a estratégia de envolvimento criada com o leitor. A primeira não faz parte do âmbito de estudo deste documento. Em relação à segunda, a revisão sistemática mostrou que a inclusão de emoções e personalidade em posts de redes sociais é uma estratégia eficiente para melhorar o envolvimento com o leitor. Além disso, o envolvimento parece depender do tipo de linguagem e elementos que constituem o post. Uma vez que estamos a trabalhar no contexto das fake-news em que limites éticos são testados e extrapolados com frequência é importante reforçar que a abordagem definida seja eticamente válida.

Neste sentido, o curso da dissertação seguiu com a conceção e desenvolvimento de um Algoritmo de Geração de Posts baseado em emoções e personalidade capaz de aumentar o engajamento de posts de fact-checking. O algoritmo foi testado num experimento realizado com vinte participantes, onde foram pesquisados padrões entre os posts obtidos pelo algoritmo desenvolvido e os criados pelo jornal de fact-checking Snopes e as interações alcançadas em cada um deles. Este experimento, realizado num ambiente de laboratório, comprovou a hipótese de que emoções e personalidade em posts de Microblogs aumentam o envolvimento do utilizador com os posts de fact-checking.

Palavras-chave: Processamento de linguagem Natural, Gerador de posts, Redes sociais, Twitter, Fake-news, Fact-checking, Emoções, Personalidade, Envolvimento

Contents

List of Figures xii

List of Tables xiv

1 Introduction 1

1.1 Context and Motivation . . . 1
1.2 Goals . . . 2
1.3 Research Hypothesis . . . 3
1.4 Methodology . . . 3
1.4.1 Ethical statement . . . 4
1.5 Document structure . . . 5

2 A systematic review on news spreading strategies 6

2.1 Search procedure . . . 6
2.1.1 Survey questions . . . 7
2.1.2 Locating studies . . . 7
2.1.3 Study selection: Inclusion and exclusion criteria . . . 8
2.2 Results . . . 8
2.2.1 Study characteristics . . . 10
2.3 Discussion . . . 16
2.3.1 Dissemination fact-checking on social media . . . 16
2.3.2 Engagement and viralisation of Narratives . . . 16
2.3.3 Fake-news strategies to create viral posts . . . 18
2.4 Taxonomy of research work on news spreading behaviour . . . 20
2.5 Summary . . . 21

3 Microblogging Environment Simulator 22

3.1 Challenges of a Research on social media . . . 22
3.2 Requirements . . . 24
3.3 System Architecture . . . 25

3.4 Final Result . . . 29
3.4.1 Home Page . . . 29
3.4.2 New Participation . . . 30
3.5 Conclusion . . . 31

4 MentaLex: a Mental Processes Lexicon 32
4.1 Related work . . . 32
4.2 Procedure . . . 33
4.3 Results and Discussion . . . 34
4.4 Conclusion . . . 36

5 Post Generator 37
5.1 Text Core, Information Retrieval . . . 37
5.2 Summarisation . . . 38
5.2.1 Summarisation approaches Evaluation . . . 42
5.3 Emotional Reinforcement . . . 44
5.4 Personality Reinforcement . . . 46
5.5 Conclusion . . . 47

6 Results and Discussion 48
6.1 Preparation . . . 48
6.2 Results . . . 49
6.3 Discussion . . . 53
6.4 Conclusion . . . 54

7 Conclusion 55
7.1 Contributions . . . 56
7.2 Publications . . . 56
7.3 Future work . . . 57

Bibliography 58

Appendices

A Detailed posts 66

B Experiment results 71

C Scientific production 74
C.1 Microblogging Environment Simulator: an ethical approach . . . 74

C.2 MentaLex: a Mental Processes Lexicon based on the Essay dataset 76
C.3 A systematic review on news spreading strategies 77

List of Figures

1 Social media post representation . . . 1
2 UNESCO´s definition on Fake-news . . . 2
3 DSR process model . . . 4

4 Publications count . . . 9
5 Search and Selection process . . . 9
6 Example of Disinformation evolution . . . 19
7 Comparison between fact-checking and fake-news engagement strategies. . . . 20
8 General taxonomy of Social media posts spreading strategies . . . 21

9 Example of Shadowban effect . . . 24
10 Twitter feed . . . 25
11 Mockup Twitter feed . . . 25
12 System use case view . . . 26
13 Object diagram: Conduct experiment scenario . . . 26
14 Class diagram of the system . . . 27
15 Partial data model in First Normal Form . . . 29
16 Home page . . . 30
17 User consent. . . . 30
18 Participation Page . . . 31
19 Help Page . . . 31

20 Word-cloud MentaLex words . . . 35
21 The frequency of partial mentaLex words in the Twitter dataset . . . 36

22 Social media post representation . . . 37
23 Extraction example . . . 38
24 Extractive Summarisation example . . . 39
25 Text core obtained selecting the sentence with the highest value of surprise and sadness 40
26 *Text core* created by connecting the fields: allegation and news classification. . . . 40
27 Result of news summarisation by GTP-2 model. . . . 42

28 Result of news summarisation by BART model. 44
29 Emotional reinforcement for Allegation Sentence approach 45
30 Emotional reinforcement for T5 model 46
31 Neuroticism reinforcement 46

32 Voluntaries gender distribution by age. 49
33 Voluntaries distribution by course. 50
34 Global Interactions count. 50
35 Interactions count per approach 51
36 Count Interactions for particular post 51
37 Top 3: Posts with more interactions 53

List of Tables

1 Specification of Software Frameworks, Libraries and Datasets used 5

2 Keywords selection for papers extraction. 7
3 Search query on the different sources. 8
4 Studies characteristics 11
5 Summary of the studies/papers included in the review. 15
6 Sentiment analysis in fact-checking and fake-news tweet 18

7 Tables that compose the database. 28
8 Configuration Table attributes 28

9 The state of the art in personality detection 33
10 Summary of commonly used words Lexicon 35
11 Summary of avoiding prone words Lexicon 36

12 Abstractive summarisation models 41
13 Summary of the automatic evaluation results 43
14 Summary of Human evaluation. 43
15 Example of how emotions, news classification (based on SNOPES), and emojis relate. . 45

16 Engagement score for each approach. 52

17 Generated posts presented to users. 70

18 Number of interactions with the posts 72
19 Content related to the five replies made by the users. 72
20 Posts that users made throughout the research. 73

Chapter 1

Introduction

1.1 Context and Motivation

Social media platforms are very popular nowadays. Millions of users are using *Posts* for sharing their thoughts, opinions, news and personal information [1]. A Post is a message that a person publishes online on a social network that may include text or images [2].

In social media studies, the content of posts is frequently employed as a research foundation since it provides insight into public opinion and what people are talking about [3]. With Natural Language Processing (NLP) techniques, the content of Posts is transformed into data that can be interpreted by a computer. Sentiment analysis and text generation are two NLP techniques relevant to this dissertation: Sentiment analysis examines the given text and determines the predominant emotion by classifying it as positive, negative, or neutral based on the degree of polarity; and text generation combines computational linguistics and artificial intelligence to generate new texts [4].

The working object of this document is a Post Generator that produces social media post with the elements shown in Figure 1. The serious and relevant problem of fake-news in social media was chosen as a case study for this project.

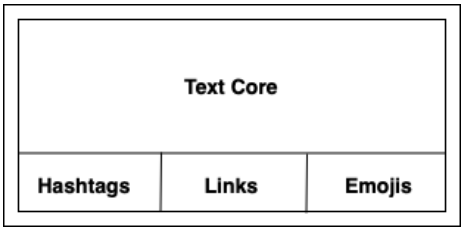


Figure 1: Social media post representation

The definition of fake-news followed in this document results from the union of information disorders

present by UNESCO [5], which aggregate three categories, as indicated in Figure 2, misinformation, disinformation, and mal-information. The first relates to false information that is distributed by someone who believes it is correct. Disinformation, on the other hand, is the act of spreading false information while being aware of it. Finally, mal-information is information based on facts that is intended to hurt or discredit a person, organisation, or country. It's vital to remember that, despite their differences, the consequences of each domain are, for the most part, similar.

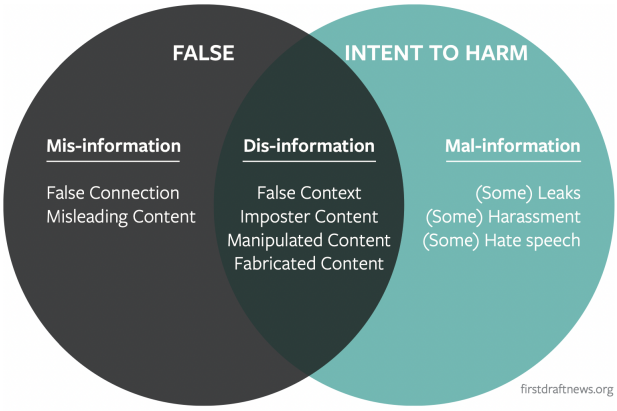


Figure 2: UNESCO´s definition on Fake-news [5].

Although fact-checking organisations have made efforts to reduce the negative impact of fake-news, the results obtained are not entirely satisfactory [6]. The main motivation for this dissertation is the difficulties that fact-checking agencies face in sharing their work and effectively using the information systems they have created.

1.2 Goals

This dissertation focuses on the development of an emotional and personality-based Post generator that can engage the reader. The resulting generator will be applied in the fact-checking world to create informative and engaging posts for social media users. To achieve this goal, it is necessary to understand how fact-checks currently act in social media, what strategies are used to generate engagement and viral posts, and what limitations exist in current approaches, following a line of research guided by the results of the systematic review for the Post Generator.

In order to verify the results of the Generator, an experiment is carried out using an interface developed for this purpose, with which the following objectives can be achieved:

- Investigate what actions users take when exposed to fact-checking posts;
- Evaluate the engagement of the Post Generator Algorithm.
- Compare the change (or not) in users actions from the original posts to the posts generated with personality and emotions.

This document focuses on the elements and characteristics of microblogging, with special attention to Twitter. The main reason for this decision was the vast number of studies related to this platform. Nonetheless, social networks are in a constant process of adaptation. While new social networks emerge and spread, others fall into oblivion. The younger generations play an important role in giving a stage to new platforms, which often take their place in the world of social networks. One example is Telegram, which have channels that allow users to interact with posts and create a more flexible environment, as there is no maximum words size and it is possible to limit public access to the content shared. Therefore, it is important to point that even with other social networks emerging and occupying an important position among the older ones, the work developed in the context of microblogging platforms, with the appropriate changes, can also be applied to another social network.

1.3 Research Hypothesis

This work tests the hypothesis:

> Microblogging posts, built with emotions and personality, are able to improve users engagement on fact-checking posts.

1.4 Methodology

The research hypothesis refer to using NLP approaches, to create posts (based on the elements from Figure 1) with special attention to the emotions and personality. At the end, the result should be a post related with some context, respecting microblogging rules, with a emotional and personality appeal allowing fact-checking agencies to grow their audience and improve results in the battle against fake-news.

In order to answer the research hypothesis, this study followed the Design Science Research (DSR) methodology. The fundamental principle of DSR is that for understanding a problem and its solution, an prototype should be built and used in the context of that problem [7]. Six steps can be followed to create the work in accordance with this process [8], as show Figure 3.

Following the DSR methodology, problem, motivations and objectives were presented in the introductory chapter of this document. For the design of the solution, a systematic review is conducted to identify the best approach for the Post Generator. Also, the development of the functionality and architecture of the artefact will be based on the results of the systematic review. The software developed during the creation of the artefact use the libraries, frameworks and datasets listed in Table 1. The related source-code files are available in GitHub at `https://github.com/maria85290/Generator`.

Next, following the DSR methodology, a case study will be created in the demonstration phase to gain concrete and contextual knowledge about social media engagement. It allows the collection and analysis of qualitative and quantitative data. The outcome will be evaluated by a group of volunteers and the results will be analysed to draw conclusions about the effectiveness of the Post Generator. Regarding the

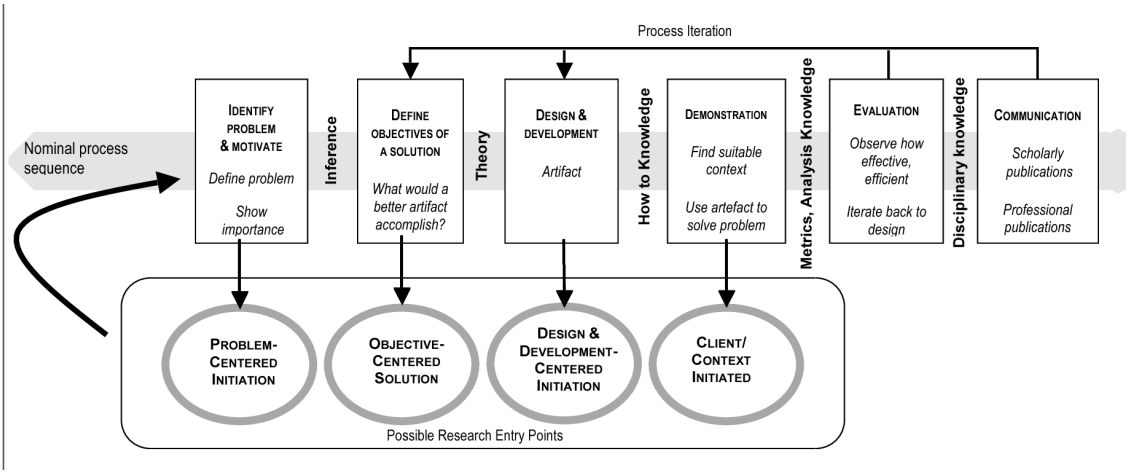


Figure 3: DSR process model retrieved from [8]

communication phase, it will be done through publications in conferences and the final defence of this document.

The chosen design and approach allow analyse the engagement (is proportional to the interactions each post receives) of the generated post compared to engagement values of the original post (obtained in Twitter). Results will be evaluated in-lab through voluntary participation. Posts have a visual component that usually has a large influence on whether the content is read or not, so design validation is also necessary. Based on the results presented in the study [11], to identify simple design problems, the participation of 20 volunteers in the study is sufficient. Furthermore, if we consider the 1.4 million Portuguese people who use the social network Twitter in 2022 [12], we conclude that using a sample of 20 volunteer Twitter users guarantees a confidence level of 80% and a margin of error of 14%. During the experiment, the twenty participants are invited to interact with the platform by liking, sharing, replying, following, blocking, and even creating their own posts. All the interactions are stored in a database for later analysis.

1.4.1 Ethical statement

During research, it is necessary to keep in mind some ethical considerations that are important for human rights and human dignity and to help maintain a strong collaboration between science and society. This ensures that participation in studies is voluntary, informed and safe for the subjects.

The social study considered in this dissertation uses an in-lab-environment that simulates a social media interface (a more detailed description is present in Chapter 3). It is important to note that this interface ensures that users know they are participating in a study and, therefore, provide consent. Furthermore, no personal identifiable information is stored. In addiction, all participants take part in the research voluntarily, with full information about it.

Phase	Frameworks and Libraries	Dataset
Programming language	• python 3.9.13	N/A
Sentiment Analysis	• EmotionsMeter.com	N/A
Microblogging Environment	• Django 4.0.6 • MySQL server 8.0.21	N/A
MentaLex	• pandas 1.4.2 • nltk 3.7 • spacy 3.3.1 • sklearn 1.0.2	• Twitter Personality [9] • Essays dataset [10]
Post generator	• huggingface's transformers • nltk 3.7 • sumy 0.10.0 • keybert 0.5.1 • spacy 3.3.1 • pandas 1.4.2 • gensim 4.2	N/A

Table 1: Specification of Software Frameworks, Libraries and Datasets used

1.5 Document structure

This document, after the Introduction, is composed of others six chapters. First, Chapter 2 conducts a systematic review of the literature to identify relevant topics and studies related to social media news-spreading behaviour. Chapter 3 details the web application Microblogging Environment simulator. Next, Chapter 4 presents the *mentalLex*, a Mental Processes Lexicon based on the Essay dataset. In Chapter 5 is described the proposed approach for the Post Generator. As for Chapter 6, addresses the results and discussion about the experiment conducted to test the research hypothesis by analysing the interaction between social media users and posts created by the generator on the microblogging simulator. Lastly, the document is concluded in Chapter 7, presenting a synthesis of the contribution with mentions of future work.

Chapter 2

A systematic review on news spreading strategies

The term *fake-news* became popular during the 2016 American election [13]. Since then, social networks have suffered greatly from the effects of its spread. To combat this problem, fact-checkers emerge and try to restore the truth [14]. However, their work does not have the same reach as fake-news. Fact-check agencies have only been concerned with the task of disproving or confirming news, and have not been able to find and define a strategy for disseminating the content they publish online.

Both fake-news and fact-checking as we know it are two recent concepts whose exploration by the scientific community is moving towards their debunking. The study of spreading behaviour has been neglected so far. This systematic review will address this direction, as it is not yet very well researched.

2.1 Search procedure

A systematic review is a secondary study that uses a well-defined methodology to identify studies, select and evaluate papers, analyse and select data to identify relevant information on a particular question or topic [15]. It also enables researchers to identify gaps and justify why research is needed in a particular field [16]. This study is inspired in the guidelines proposed by Kitchenham [15]. Adapting from [15], the steps considered in the study were:

- *Planing the review* by identifying the research questions of the study;
- *Conducting the review*: Defining the search strategy and string; defining the inclusion and exclusion criteria; Reviewing the quality of the studies; and Analysing the selected studies;
- *Reporting the review*: Presenting and discussing the results;

Because the included studies do not have statistical or numerical measurements that would allow comparison of results, this work is conducted using a qualitative systematic review. Which can be defined as a means of comparing results from qualitative experiments.

2.1.1 Survey questions

This work serves three functions. The first one is to provide an overview of the strategies followed by fact-checking to disseminate the content on social media, and thus define the question: RQ1) How are fact-checking posts disseminated on social media? The second is to identify the benefits of using emotions and personality in a post to get viral narratives. This leads to the second guiding question: RQ2) How can post generators create engagement narratives that go viral? The last goal is to analyse the tactics of fake-news to reach as many people as possible, which leads us to the final research question: RQ3) What was the fake-news strategies to create viral posts?

2.1.2 Locating studies

The study [17] examines and compares the systematic search qualities of the 28 databases that support academic research and concludes that 14 of them are suitable for systematic reviews considering parameters such as accuracy, retrieval and reproducibility. Considering the findings of this study, the *ACM digital library* [1] and *SCOPUS* [2] were selected as sources for searching relevant articles

To answer the above research questions, four main areas and associated keywords can be identified. The Table 2 shows a summary of the keywords and a brief description of each subject. The most popular synonyms have been included to the list of keywords in consideration of the fact that same concept can be expressed in several ways.

Description / Justification	**Keywords**
The field of action of this work are Social media, especially *Twitter*.	Social network; Social media; *Twitter*
It is intended to use *fact-checking* for the fight against fake-news.	Fact-check; Fact-checking
Look forward to a posts generator for social media posts that improves reach and engagement.	Emotion; Feeling; Sentiment; Personality; Text generator; Post generator
Identify how fake-news create engage posts.	Fake-News; False-News; Spread; Dissemination;

Table 2: Keywords selection for papers extraction.

[1] https://dl.acm.org
[2] https://www.scopus.com

As indicated in Table 3, the search protocol was developed with dijunctions and conjunctions carefully positioned to maximise search.

Data Base	Query
SCOPUS	TITLE-ABS-KEY (((“fact-check” OR “fact-checking”) AND (“social network” OR “social media” OR “twitter”)) OR ((“text generator” OR “post generator”) AND (“Social Media” OR “Social network” OR “twitter”) AND (“Emotion” OR “feeling” OR “Sentiment” OR “Personality”)) OR (“fake-news” OR “false news” AND “spread” AND “dissemination”))
ACM Digital Library	((“fact-check” OR “fact-checking”) AND (“social network” OR “social media” OR “twitter”)) OR ((“text generator” OR ”post generator”) AND (“Social Media” OR “Social network” OR “twitter”) AND (“Emotion” OR “feeling” OR “Sentiment” OR “Personality”) OR (“fake-news” OR “false news” AND “spread” AND “dissemination”))

Table 3: Search query on the different sources.

The bibliography was compiled over the first two days of November 2021. The previous research was then expanded by conducting a second bibliographic survey on June 1, 2022.

2.1.3 Study selection: Inclusion and exclusion criteria

Only scientific articles that were peer-reviewed and whose content was freely available were considered. Then, studies that were not in English or Portuguese, that were duplicated and whose content did not fit the context of this dissertation were excluded (*e.g.* studies dealing with the techniques for the detection of fake-news were removed).

The prevalence of fake-news has increased sharply since the American election in November 2016 [18]. As a result, most studies analysing the impact of fact-checking appeared in early 2017, as show in Figure 4. For this reason, the search was limited to the last five years (since 2017).

2.2 Results

Using the established protocol, 866 articles relevant to the research questions were identified from the two databases. First, duplicate papers (n=42), studies before 2017 (n=77) and records without peer-review (n=28) were removed. In this way, 147 records were excluded. The remained papers were screened by reading the titles. To automate this process, the tool Rayyan [3] was used, which identified 190 studies as eligible. The next step was to read the titles and abstracts of the studies. Of these, 34 were considered suitable for full reading. The reasons for excluding the others are related to the discussed topic does not fit the context to explore here.

[3] https://www.rayyan.ai/

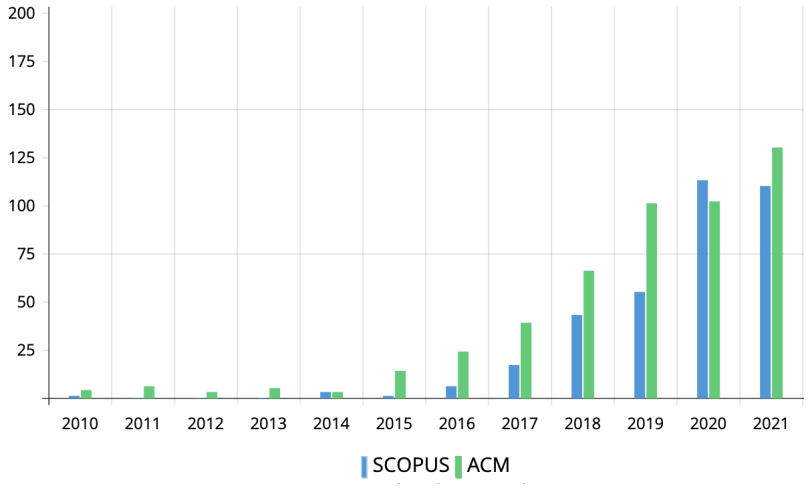


Figure 4: Count of publications obtained by queries (presented in Table 3) since 2010. The increased trend from 2016 coincides to the fake-news and fact-checks recognised growth.

The list of included articles also incorporate studies that had already been identified as relevant to the review but were not included by the systematic review. Figure 5 presents a flowchart that provides more details of the exclusion process. Finally, a total of 14 articles were included for the review.

While there are many studies that focus on the development of fact-checking tools and strategies to detect fake-news, no studies were found that present a systematic review in the peer-reviewed literature in the news spreading strategies

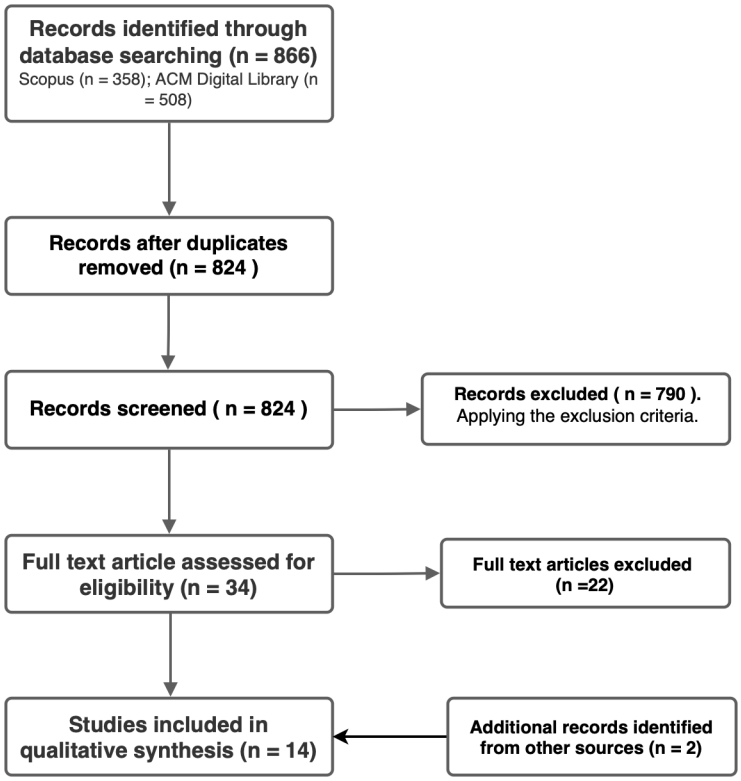


Figure 5: Search and Selection process

2.2.1 Study characteristics

For each work included in this review, the publication source, ranking, publication year, country of publication, and total number of citations were listed in the Table 4.

Related to the citations, Vosoughi's *et al.* [19] work has 1530 citations and explores the spread strategies of fake-news. The two works by N. Vo *et al.* [20], [21], with 50 and 14 citations respectively, focus on identifying a list of Twitter users who combat misinformation and tweet fact-checking information. They also analyse the characteristics of their fact-checking posts and use the results to create a generator for automatic replies with fact-checking intent. With 11 citations, I. Singh *et al.* proposes a text generation model that incorporates emotions, based on NLP and GTP-2. It is important to note that this values are taken from the official page of the document's source of publication on June 7, 2022.

The articles included in this review were published in a variety of relevant journals and conferences with a high ranking classification. The examination of demographic distribution identify five countries. The United States (USA) has the most articles, with nine, followed by China (CN) with two, and Portugal (PT), India (IN), and the United Kingdom (UK) with one.

Study	Country	Publication Type	Rank	PY	TC
Vosoughi *et al.* [19]	USA	Journal (Science)	Q1	2018	1530
Adali *et al.* [22]	USA	Conference (ICWSM)	–	2017	442
S. Sorok *et al.*[23]	USA	Journal (Organizational Behavior and Human Decision Processes)	Q1	2015	313
C. Heath [24]	USA	Journal (Organizational Behavior and Human Decision Processes)	Q1	1996	199
Britt *et al.* [25]	USA & FR	Journal (SAGE Publications)	Q1	2019	58
N. Vo *et al.* [20]	USA	Conference (SIGIR)	A* (CORE 2018)	2018	50
N. Vo *et al.* [21]	USA	Conference (SIGIR)	A* (CORE 2019)	2019	14
I. Singh *et al.* [26]	IN	arXiv (open-access archive)	–	2020	11
C. Shao *et al.* [27]	CN	Journal (IEEE Access)	Q1	2018	6
K. S. Autry *et al.* [28]	USA	Journal (Applied Cognitive Psychology)	Q2	2021	3

J. Lee *et al.* [29]	USA	Journal (Online Information Review)	Q2	2021	3
Z. Wang *et al.* [30]	CN	Conference (PRICAI)	B (CORE 2019)	2019	3
F. S. Marcondes *et al.*[31]	PT	Conference (worldCIST)	C (CORE 2020)	2020	2
L. Piccolo *et al.* [32]	UK	Conference (CUI)	–	2021	0

Table 4: Studies characteristics: Country, publication source, ranking, total citation (TC) and publication year (PY).

Next, Table 5 presents a summary of the results found in the review to answer the research questions.

Study	Findings
N. Vo *et al.* [20]	The authors have identified and collected a list of Twitter users they call guardians who combat misinformation by tweeting fact-checking URLs. Guardians need to be careful about what they share with their followers to maintain engagement. In addition, to keep the engagement, they typically took less them a day to respond to claims in online conservation's and another day to share verified information with its followers.
F. S. Marcondes *et al.*[31]	The authors proposed a fact-checks dissemination profile and a behavioural design. For that, they explore the hypothesis that *fact-checking spread-bot will have more impact on their followers by simulating human-like behaviour, compared to the simple 'link only' tweet*. The design of the behaviour assumes that for each fact to be communicated, *A-Guardian* takes a record on a particular topic and generates text fragments, which it inserts with the link and decides whether to mention someone on Twitter (or not). In the end, *A-Guardian* sleeps for a certain time until the next share event. The authors goal is that people will recognise the bot profile behaviour, like Human, and thus connect with the post content.

N. Vo *et al.* [21]	The study concluded that fact-checking tweets (*FC-tweets*): 1) Use impersonal pronouns more often compared to normal responses on Twitter; 2) Responses tend to refute the original content of the tweets by using the adjectives fake, wrong, dump, false and untrue; 3) *FC - Tweets* have more formal language and use few swear words; and 4) *FC-Tweets* show a clear intent to fact-check. Based on these results, the authors purposed a framework, in order to generate responses with fact-checking intention, used to support fact-checkers and increase their engagement. The model created successfully generates responses with the intent of fact-checking.
C. Shao *et al.* [27]	Fact-checking is competing with misinformation in two dimensions: fact-checkers and fact-checking consumers. The firsts, are working hard to debunk a claim, but their efforts struggle to keep up with the huge volume of misinformation. The fact-checking consumers are sharing tweets of debunking the misinformation. In addiction, the authors study the strategies used by social bots to increase the spread of fake-news, and presented three case studies: producing a large number of original tweets, use trending hashtags and injecting fake-news content into the conversations.
J. Lee *et al.* [29]	They identify angry in fact-checking tweets.
Z. Wang *et al.* [30]	The authors proposes a model capable of writing short Chinese texts in human-like form by introducing BFP. For this purpose, they use a CNN-based text classifier and a conditional LSTM-based text generator.

L. Piccolo *et al.* [32]	The authors show a multi-agent approach to engage Twitter users with fact-checkers. The first agent encourages users to share verified content. To this end, it actively searches for URLs on Twitter that have been classified as misinformation by the fact-checker. Based on this information, the bot replies to the tweets spreading misinformation in an attempt to enlighten the author of the post. The second agent uses natural language to explain the credibility rating of certain posts. To develop the best design for the bot and to understand the preferences for credibility indicators in explaining the evaluation of misinformation, the authors conducted two online studies.The results suggest that bots should be consistent, friendly, engaging, and credible in their presentations to connect the users with the posts.
I. Singh *et al.* [26]	The authors have presented a model based on the GPT-2 and the plug-and-play language model [33] that is able to incorporate one of the eight emotion categories into the output of the generated text while ensuring grammatical correctness.
Vosoughi *et al.* [19]	The authors examined a dataset of Twitter rumours to better understand how fake-news spreads. They discovered that fake-news engages more individuals than truth. True stories induced feelings of anticipation, sadness, joy, and trust, whereas fake ones prompted responses of fear, disgust, and surprise. Another conclusion is that bots spread either true and fake-news at the same speed, meaning that fake-news spreads faster than true news because humans, rather than algorithms, are much more likely to spread it.

Adali *et al.* [22]	The authors analysed three datasets and attributes that capture the style and language of articles to show that fake-news is more comparable to satire than real news in most situations. They demonstrate that overall title structure and the usage of proper nouns in titles are particularly important in distinguishing false from real. This brings to the conclusion that fake-news are focused at developing strategic links between entities and statements for audiences who are hesitant to go beyond titles.
Britt *et al.* [25]	The authors have study how memory works and our biases in selecting and interpreting information in the fake-news environment. Some of the important findings are: 1) Simply repeating information can increase confidence in its perceived truth; 2) initial incorrect information remains available and can continue to have an effect despite learning the corrected information; 3) and we are more likely to accept information that is consistent with our beliefs.
K. S. Autry *et al.* [28]	Negated fake-news can be increase belief in misinformation when people do not been exposed to the misinformation.
C. Heath [24]	People like to share news that matches their emotional valence when presented with extreme information. Contrary to some theoretical assumptions, people in emotionally negative environments are willing to communicate terrible news, even when it is exaggeratedly unpleasant. However, in emotionally happy environments, people frequently search excessively happy news.
S. Sorok *et al.* [23]	People frequently give greater weight to negative information than to positive information.

Table 5: Summary of the studies/papers included in the review.

2.3 Discussion

2.3.1 Dissemination fact-checking on social media

According to the previous results, two different types of actors can be involved in the dissemination of fact-checked news: human and non-human. An example of the first type are the Guardians who develop and use social media profiles to disseminate the URLs of fact-checked articles [20]. But even good Guardians cannot fight against highly professional and technological propaganda structures [34]. This enables the use of non-human applications, such as social media bots, to better perform artificial actions, in the context of spreading fact-checked posts on Twitter. These bots are the outcome of algorithms that mimic human behavior and are capable of creating content and interacting with people on social media. Like they currently have a significant presence on Twitter, between 9% and 15% of Twitter accounts [35], they are well suited to accelerate the spread of fact-checking through actions such as sharing or replying. One strategy is simulate human behaviour to engage people with the content of the post [31]. This can be done by giving the bot instructions with actions that social media users often take to achieve engagement for their posts. For example, identifying influence users, which encourages their engagement in the [31] outreach process, or even use trending hashtags.

Beyond that, the bots should also be consistent in their presentation, friendly and credible [32]. Another approach to bots that have been tested with good results is to use the ability to actively search for posts whose content is fake and respond to show users the truth [32]. Although this strategy has been frequently used by fact-checking agencies, it is necessary to be careful using words that negate the content, as this can have an opposite effect on the reader, especially If is the first time that he is confronted with such information.

2.3.2 Engagement and viralisation of Narratives

To understand how a story goes viral, it is important to first define what spread means. In the context of social networks, it is common to define the spread of a topic as social contagion. Each individual in a social network may be thought as a node that is linked to other nodes through social ties. When a person or a node is confronted with a piece of information, can easily expose anyone who is socially connected [36]. This spread continues in a cascade and becomes more intense the more interactions it reaches, as shown in [37]. It is important to remember that the social media algorithm does not allow every post to appear in the feed at the same time, and that interactions are the main influence of the algorithm that decides how many people will see it. But then, for the first user to come across some post, parameters such as the number of interactions cannot be used to start the cycle that can lead to viralization. So an alternative is to pay attention to the content and the way the narrative is built. In this sense, the study made by Nguyen Vo and Kyumin Lee [21] concluded that, compared to normal responses on Twitter, fact-check posts more often resort to the use of impersonal pronouns, use more formal language, show a clear intention to confirm facts, and frequently resort to the use of the adjectives: fake, wrong, dump,

false and untrue. But it is necessary to bear in mind that these identifications are the opposite of what is expected by social networks users. For example, the use of formal language may be contributing to increasing the distance between the user and the posts, as it is not the type of content that engage social media user.

Another problem that can arise is a mechanisation of the text, which can lead to a distance between the post content and the reader [30]. When thus distance is perceived the content can appear artificial and fall into the uncanny valley [38]. Two approaches have been presented to circumvent this problem: Introducing personality in text for microblogs using the model big5[4] [30] and identify the emotions present in the tweets and how they relate to the possibility of going viral. Both options could be a way to capture the user's attention and bring the text closer to human speech. Although this has not yet been tested in the context of fact-checking, some studies have examined the relationship between the presence of certain emotions or personality and user engagement. In terms of emotions, tweets showing melancholy and surprise are more likely to be retweeted and liked by users, but tweets displaying other emotions (for example, anger) are less likely to receive engagements. On the other hand, tweets associated with corrections (*e.g.*, tweets posted by fact-checking agencies) are known to contain anger [29]. This could be a possible explanation for why fact-checking posts do not receive many interactions.

This information can be easily verified by selecting two random tweets as examples, one referring to fake-news and the other to fact-checking. For each of these tweets, a sentiment analysis was conducted using the *EmotionsMeter.com* tool. The results, in Table 6, show that the emotions in the fake-news tweet are more melancholic, as expected, which may be the reason for higher engagement compared to the fact-check post.

Type	**Tweet**	**Emotions**	**Polarities**
Fake-news	"R..P. to the 8 year-old boy who died in Boston's explosions,while running for the Sandy Hook kids. prayforboston"	• Disgust: 20%; • Fear: 40%; • Sadness: 20%; • Surprise: 20%;	• Negative: 100%

[4]The Big Five personality theory is a hierarchical organisation of personality traits based on five basic dimensions: Extroversion, Amiability, Consciousness, Neuroticism, and openness to experience [39]

Fact-check	"No, a missing airliner (Santiago Flight 513) did not mysteriously land 35 years later with a cockpit and passenger cabin full of skeletons."	• Anticipation: 100%	• Positive: 100%

Table 6: Sentiment analysis in fact-checking and fake-news tweet

As a final note, the presence of emotion has an impact on how the user interacts with a post and are consistent with the idea that it is important to analyse the different emotions and the impact each of them can have on users when inserted into a post.

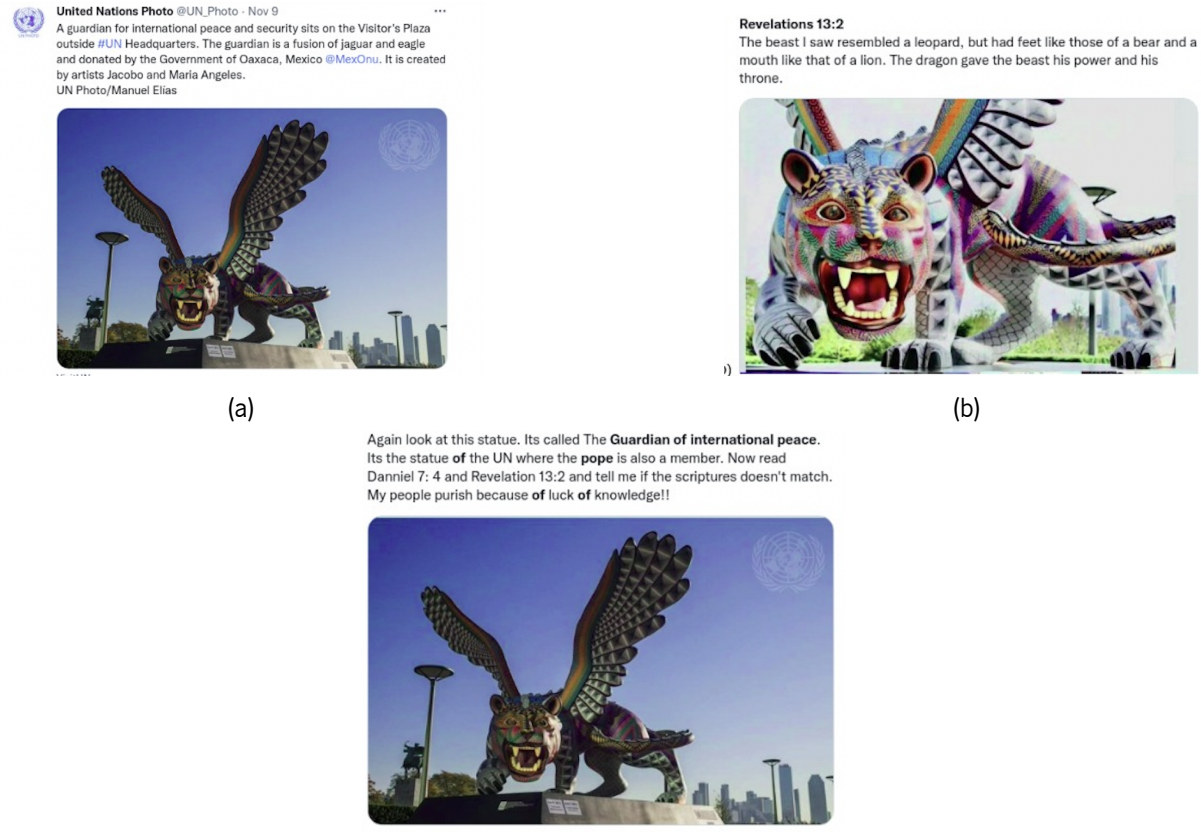
2.3.3 Fake-news strategies to create viral posts

There is no secret recipe for the viral spread of fake-news, but a number of reasons and factors, which in this analysis have been categorised into two areas that potentially influence the success of fake-news spread: 1) the semiotic structure and content of fake-new posts; 2) the cognitive bias; For the first area, the graphics and images, the form of headlines and the language used in the content, can be used to explain why fake-news spread so quickly on social media. The second is related to the systematic errors in judgment that human beings consistently make.

In 1996 the study [24] conducted by Heath showed that people have a preference by exaggeration, especially when the news is truly bad, therefore fake-news featuring accidents, disasters, or murders may elicit higher emotional involvement. Sarcoka and Mandamus in 2015, observe the same phenomenon, that is, really unfavourable news attracts more public attention [23]. Vosoughi *et al* in the 2018 study, discovered that political content was more widely disseminated and reached a bigger population than news related with the everyday life [19]. Also found that fake material spreads faster than daily life content because it evokes various feelings of disgust, fear and surprise in readers [19]. The results from Table 6 confirm this statement.

These emotions are often reinforced by the spread of altered images that are spectacular, unusual, highly visual [40] and the psychological phenomenon Social validation. For example, the Figure 6, which suggests that the new statue at UN headquarters is an anti-Christian representation associated with the disinformation campaign against the Pope. Figure 6a, was a official statement and the Figures 6b, 6c transform it into disinformation. These transformations follows the traditional disinformation strategy of using uncontextualised information with visual content about matters that generate controversy on society.

Another strategy relates to the visual structure of the post/article like the clickbait [22], often through long titles, sensationalism, titles and texts that arouse curiosity. Fake-news articles are usually captivating



(a) (b)

(c)

Figure 6: Desinformation evolution instance: a) Original published on 09/11/2021; b) published on 06/12/2021 linking with the Bible: c) published on 19/12/2021 linking with the Pope.

narratives that can be read by anyone, as they require a lower level of education to interpret [22], and more informal and simple language, without technological production:

> "Real news articles are significantly longer than fake-news articles and fake-news articles use fewer technical words, smaller words, less punctuation, fewer quotes, and more lexical redundancy", Adali *et al.* [22]

The second dimension to be aware, in the viral spreading of fake-news, is the readers perception of the (fake) news. The reader is more likely to act on headlines and tags without first reading the text [25], [41]. Second, social media transmits signals that impact our sense of information popularity, leading to greater acceptance [41]. This is due to the idea that mass-medium communications have a bigger impact than others. Another factor is the inclination to attach higher accuracy to an authoritative figure's opinion (unrelated to its content) and be more affected by it [41].

The artificial intelligence is usually responsible for the dissemination process of the fake-news, usually made by social bots that simulate users behaviour. While bots transmit a lot of fake-news, real people are responsible for the vast majority of fake-news re-transmissions [19]. Vosoughi et al. [19] found that real human users are 70% more likely to retweet fake-news than true stories.

Figure 7 provides an review of the strategies used by both fake-news and fact-checking agencies. The engagement strategies are presented in the center, and the methods implemented by fact-checking and fake-news on both sides of the image.

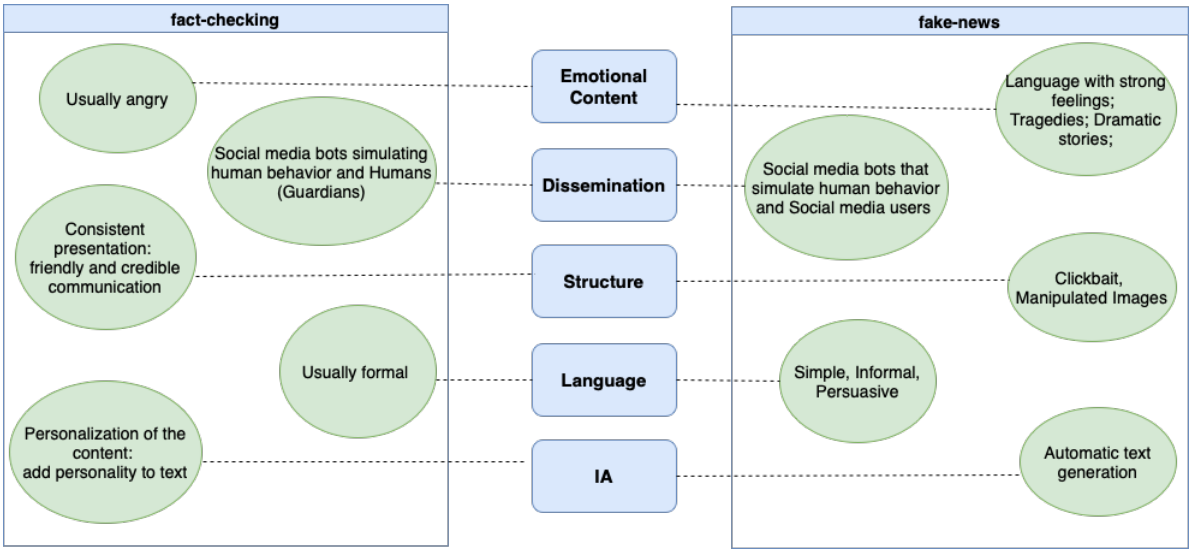


Figure 7: Comparison between fact-checking and fake-news engagement strategies.

2.4 Taxonomy of research work on news spreading behaviour

As part of this systematic review, fourteen studies were selected to be examined to identify the most important issues and contributions in this area of research. Figure 8 summarises and suggests a taxonomy on news dissemination behaviour in social media. In the taxonomy, the first topic highlights the content of the post, which may be a true fact (as the content produced by fact-checking) or fake-news. Next, this content needs to be disseminated, which can be done by humans and/or social media bots. Finally, the last facet relates to engagement, which is associated with user interaction with posts. In this facet, the linguistic aspect, the emotional appeal or the structure of the post seem to have an impact on the dissemination behaviour.

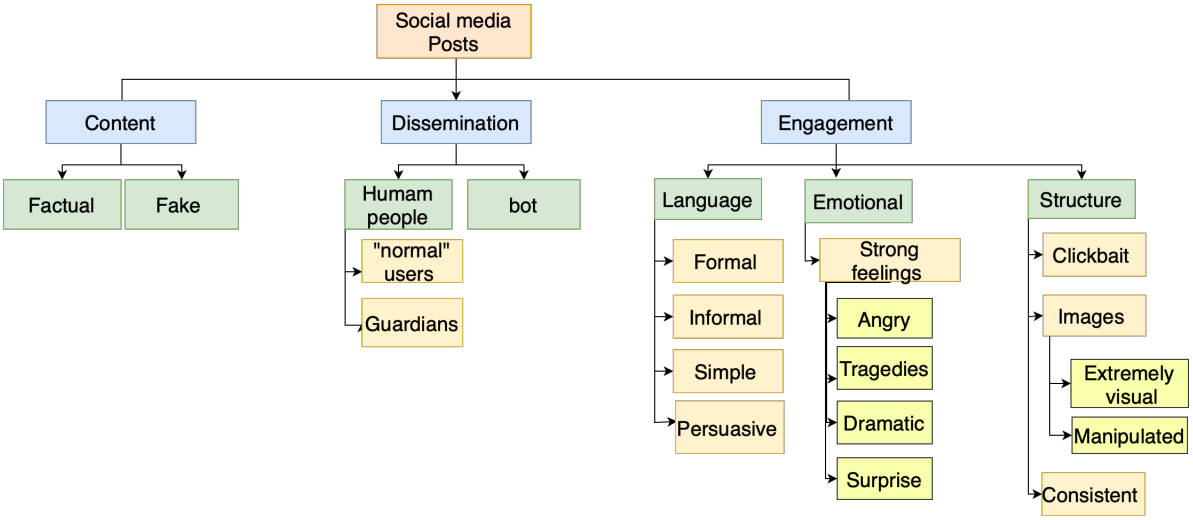


Figure 8: General taxonomy of Social media posts spreading strategies

2.5 Summary

Throughout the literature review, fact-checking was mentioned as a important strategy for combating the broad dissemination of fake-news. It was possible to investigate the various linguistic patterns used by fact-checkers on social media in the studies examined. An important finding, was: **Include emotions and personality in social media posts demonstrated to be an efficient method of improving the reader reach and engagement**. But, it's necessary to pay attention to the ethical statements. Researchers can't follow the error of fighting fake-news with the un-ethic methods. Although some studies have been done in the field of Natural Language Generation (detecting patterns in language used in fact-checking posts), there is a gap in the use of Post Generators to create engagement posts for social media. It is important to note that this systematic review includes only two databases, so other relevant studies may not have been evaluated. In addition, the choice of keywords also influences the results, and although they were carefully selected, the authors' beliefs and opinions may have influenced the choice.

Based on these results, it was propose a new taxonomy related to the news spreading behaviour. The taxonomy includes a variety of different facets, such as content, dissemination and engagement.

Chapter 3

Microblogging Environment Simulator

Numerous studies developed in social media give the perception of ethical problems that can arise if the research doesn't pay attention to some ethical dilemmas.

In order to fulfil the goal of this project, a safe and controlled environment is needed to test the work. In this chapter, the Microblogging Environment Simulator is presented. This platform will make possible draw conclusions about the functionality and acceptance of the crated Post Generator algorithm. It also seeks to ensure that the created experiment has been adequately tested in a laboratory environment before being placed on the social network, so that a bug that can be detected in a lab environment does not jeopardise the experiment's success. This Chapter begins with a description of the interest and requirements to the platform, followed by a presentation of the general structure and functionalities, as well as an indication of the technologies used and the main decisions made during its development.

3.1 Challenges of a Research on social media

For this study, it could be essential to extract data related to real interactions on social media to measure the impact of the developed algorithm. Working directly in social media, on the other hand, necessitates an understanding of the complexities of ethical considerations. Several ethical issues emerge, which can be mapped into four main dilemmas [42].

- **Data:** What is public data? How would we deal with public data posted by a user?
- **Consent:** When and how should we seek informed consent?
- **Traceability:** How can we de-identify participation in a database?

- **Research with children and young people:** How do you ethically work with young people to ensure they understand the implications of participating? [1]

Moreover, the inherent characteristics of social media can make its use extremely harmful to the mental health of its users, making them feel a physical and emotional need for it. It becomes clear that the integrity of the human person can be called into question, which takes to another dilemma:

- **Integrity of the Human Person**: How to ensure respect for the physical and mental integrity of the people involved in a study?

The most common way social media can damage mental health is by rewiring the brain with *instant gratification*. This happens through the production of dopamine, a chemical in the brain associated with pleasure and reward systems. When a user first receives notifications on social media, their brain releases a shot of dopamine that makes them feel very good. With time of use, these shots become frequent. This trains the brain to expect immediate gratification from notifications [43]. The dangerous thing about this compulsive use is that when gratification fails to materialise, users can internalise the belief that this is because they are "unpopular"or "unfunny". Another psychological problem that can occur is Fear of Missing Out, which is characterised by the desire to be constantly connected to what others are doing [44]. These feelings can give people the false sense of being involved.

If a user does not understand social media and why their brain reacts the way it does, it can easily lead to emotional issues. Figure 9 illustrates a real example of a problem that can arise from the use of social media. In a *social forum*, a Twitter user explains that the *shadowban* [2] phenomenon caused him to lose his only source of social communication, leading to self-harm. Although the example refers to a specific situation, conducting a study on social media over a period of time and then stopping can trigger similar situations.

To overcome the previous five dilemmas, it was decided to develop a prototyping environment for social media. This platform, acting as a laboratory environment, will allow data to be collected to understand human opinion about the created posts, without compromising ethical values.

[1] Most social media (including Twitter) require a minimum age of thirteen to create a profile. But young people can easily overcome this hurdle by simply changing their date of birth, because this platforms do not perform verification.

[2] A shadowban consists of blocking a user or part of their online content without them realising it. This is a common practice in social media, including Twitter.

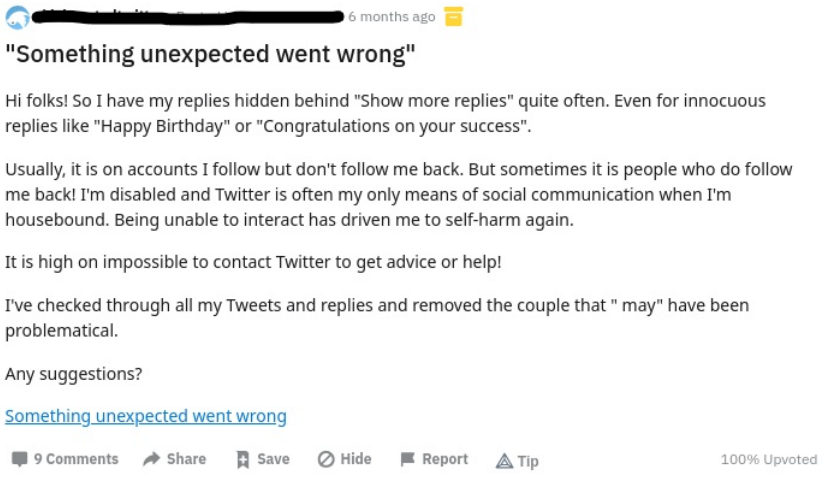


Figure 9: Example of Shadowban effect on Twitter [45]

3.2 Requirements

Before start building the application, it is important to perform a requirements analysis to ensure the success of the developed system. The requirements have been divided into:

- Functional, capturing the intended behaviour of the system.
  - Allow the configuration of the experiment to be changed so that the researcher can select the variables to include in the study, giving greater elasticity and adaptability to studies with slightly different scopes.
  - Save the interactions that are performed on the platform.
  - Link data to the corresponding configuration.
  - Present detailed information about the project where the application is inserted.
- Not-functional, restricting or limiting the design.
  - **Re-usability**: Considering the numerous social media platforms out there, it might be useful to undertake future studies in the context of others platforms. As such, the developed system must be able to integrate another interface with few changes.
  - **Usability**: Considering the purpose of the platform, it is necessary to design and implement an environment that provides an immersive experience to the user. It is also necessary to ensure that the platform is interactive so that real people can respond to the generated content, which is crucial for the improvement and validation of the work.

Since there are a variety of social networks with different functions, purposes, and audiences, the prototyping environment has modelled microblogging platforms and is inspired in *Twitter*, which was studied in this work. Examples of other social media that fall within this context include Weibo, GETTR or GAB.

Choosing a specific platform allows to take into account the particular characteristics of a microblogging platform, to customise the developed environment , and to bring it closer to a browsing experience on Twitter. It is important to note that although Twitter served as the inspiration for the developed interface, efforts were made so it would be similar but not exactly equal to Twitter. That ensures it is clear the post was not published on Twitter even if the posted content was widely distributed.

The images 10, 11 and 18 show, respectively, an example of a Twitter feed, a model of the prototyping environment to be developed (mockup) and the result obtained.

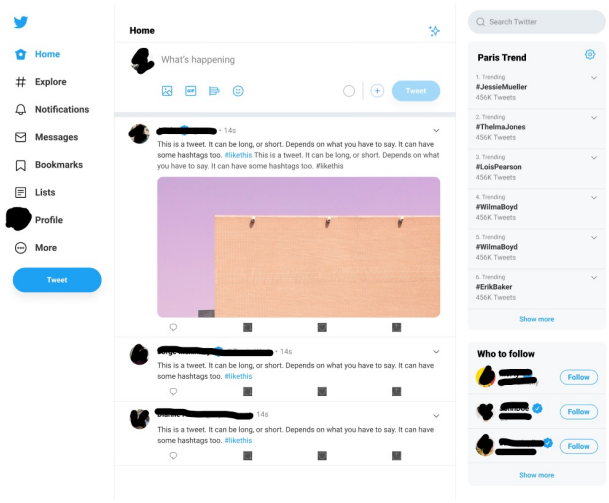


Figure 10: Twitter feed

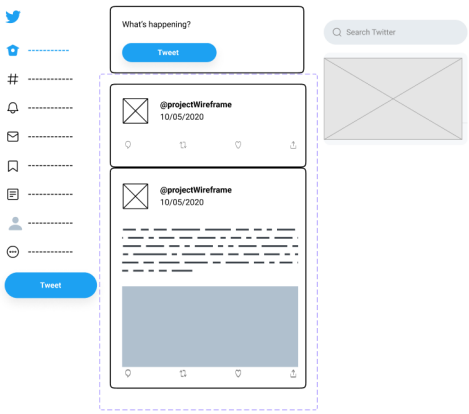


Figure 11: Mockup Twitter feed

3.3 System Architecture

This section describes the functionalities and architecture of the platform. Figure 12 shows the diagram of use cases and the three main blue scenarios with some sub-scenarios for each of them. The Researcher performs experiments and the *SuD* (System Under Development) has the task of creating the environment that allows the Subject to access and participate in the study. Thus, the Subject is a secondary actor who only provides a service to the SuD [46].

The object diagram in the Figure 13 represents the second blue scenario and shows the Subject interaction with the experiment. When the SuD reacts to the instructions of the researcher (first blue scenario) creating the experiment and the Subject gives consent to participate, the environment is ready for the study. The application has a :Listener role that waits for interactions from the participant and send them to the :Database Handler, which is responsible for correctly record the data.

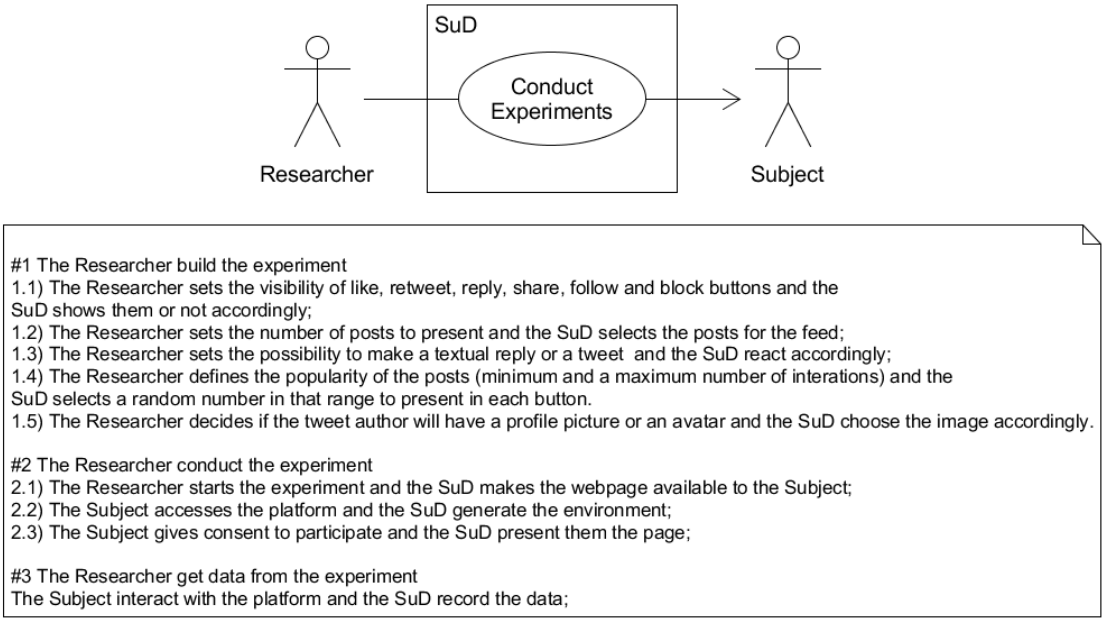


Figure 12: System use case view

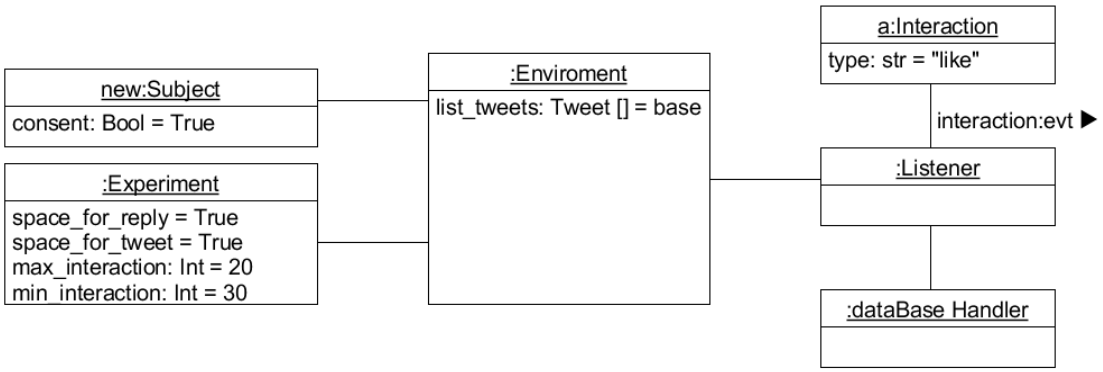


Figure 13: Object diagram: Conduct experiment scenario

The class diagram in Figure 14 shows the nature of the relationship between classes. The classes comprise of *interaction, tweet, tweet frame, configuration*, and *feed*. Interact with the platform can be done through one of the buttons whose visibility is set to True. But, different from the Twitter beyond liking, retweeting, and replying, the actions of blocking, following, and sharing are still considered interaction with a post. Moreover, these actions are incremental and are only used to measure engagement of the Subject participant with the post. A configuration is defined by the researcher and can be composed by the visibility of the interaction buttons, the engagement of each post, the length of the feed, and the characteristics of the post author.

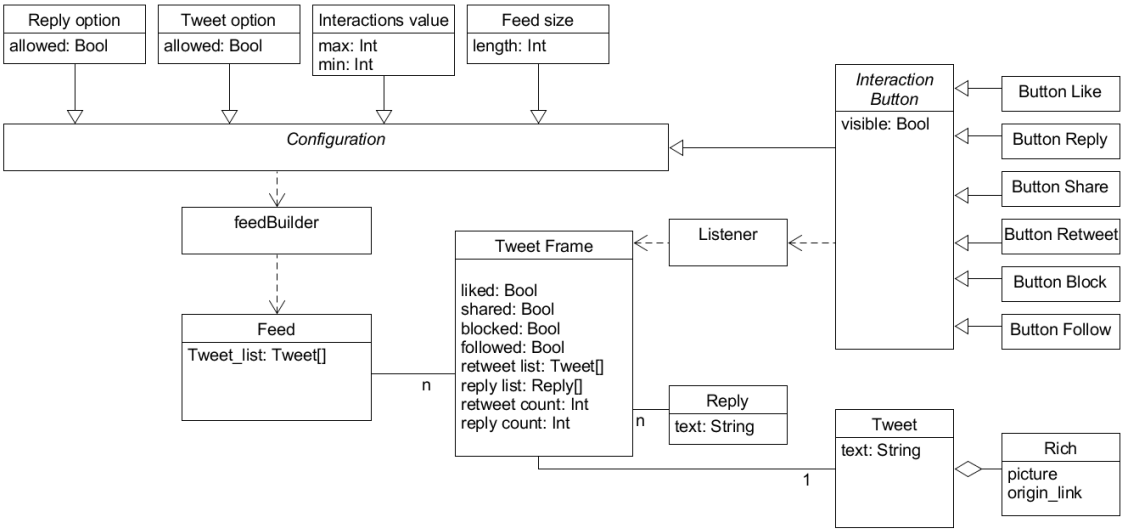


Figure 14: Class diagram of the system

The platform has been developed using the framework Django. This choice is justified by the fact that it's inherent characteristics meet the previously defined requirements. It stands out for its high scalability, which allows adding more resources at any time, no matter at what stage of development. The frontend uses templates to provide a user interface that allows participation in this study. The backend is responsible for processing the requests sent by the frontend and providing a response.

As for the database, a relational one was used because, Django, is designed to simplify the use of relational databases. Furthermore, although social media interactions are often stored in graph databases, they can easily be entered into a relational one. The database consists of 10 tables being all generated by Django migrations and the respective attributes conceived in the Django models. There are three main tables, *Interaction*, *Configuration*, and *Post*, which deal respectively with the actions performed by users, the variables presented in the environment and the narratives (the posts) to be presented to the user. The user table will record data related to the user participation in the study namely time of participation and personality. This last field, it's kept with the intention to allow the research to look for patterns related to the user's personality and post content, this can be done, for example applying a personality test to the user. The *post_by_participant* table stores the content published by the user in the feed. Also, the *action_type* table stores the type of interaction performed by the user. A record is stored in the *action_reply* table, if the *action_type.name* is a *reply*, with the content written by the user. The *hashtag* and *image* tables, as the name suggests, contain the hashtags and images associated to a specific post. Figure 15 presents a better understanding of the data record by the system in a partial data model view. In Table 7 a summary of the tables is presented.

Table	
Post	Content obtained by the posts generator.
Image	Location of an image belonging to a post.

Hashtag	Associate each hashtag to a post.
Participant	It stores the time of entry and exit for each participant and their personality trait (if a personality test is apply).
Action_type	It stores the seven types of actions: share, like, reply, block, follow, retweet and make_a_reply.
Action_reply	Keeps the content of the reply, whenever the type of action is *make_a_reply*.
Post_by_participant	Keeps the content of the posts made by the participants.
Configuration	A set of variables that define a configuration for the study. It is possible to define the interactions available (like, comment, share, make a post...), the number of posts and the number of iterations by default in each one.
Environment	Defines the configuration present in the study.
Interaction	Relates an post, with a participant, an action, and a configuration.

Table 7: Tables that compose the database.

It is important to reinforce that different metrics can be collected depending on the direction the researcher wants to take the experiment. Table 8 shows the variables that are currently configurable. Applying a new configuration to the study is only possible for users with administrator rights. Additionally, Environment table allows changing a configuration that is applied throughout the study.

Entries	**Explanation**
space_for_comment	If True display a space to comment.
space_for_create_post	If True allows the participant to post.
user_picture	If True posts are made by users with a profile picture.
[reply,like,share, block, follow, retweet]_button	If True it is possible to perform this interaction.
posts_number	Define the number of posts to present at the feed.
[lower,upper]_limit_interaction	Defines the minimum and maximum number of interactions that the post must present.

Table 8: Configuration Table attributes

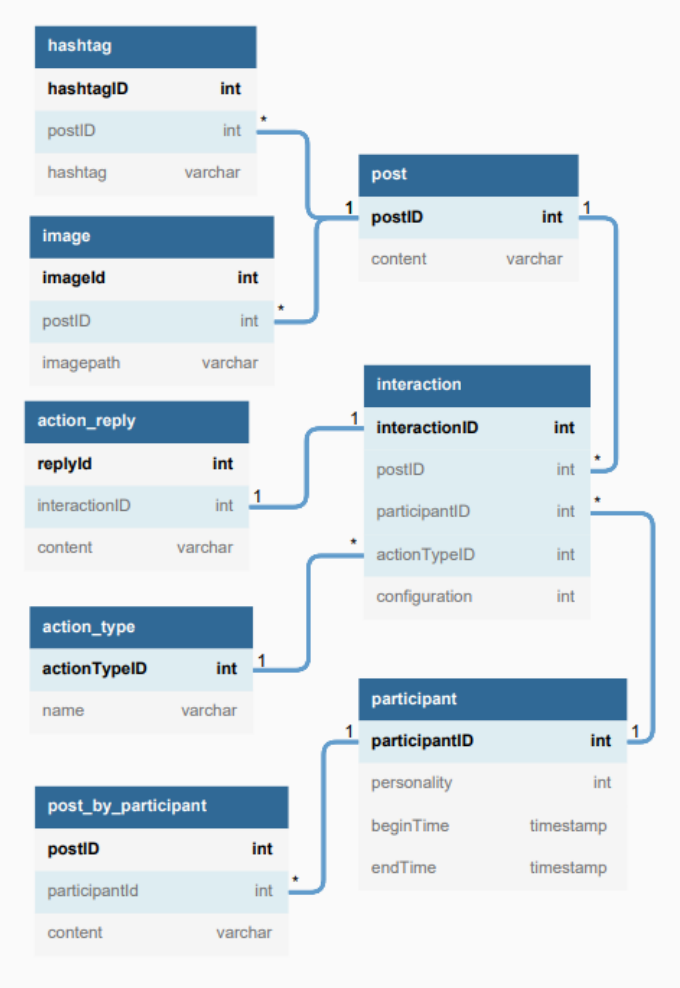


Figure 15: Partial data model in First Normal Form

3.4 Final Result

The product created by the development of the previous components is the web application presented in this section. As previously stated, this platform can be utilised to perform research and collect data on user behaviour. It is critical to emphasise that no personal data is saved, hence the platform is compliant with GDPR (General Data Protection Regulation). This project was created in Python (open source) and can be customized to meet the requirements of any research. The platform is available at: https://github.com/maria85290/MicrobloggingSimulator. Below are presented some pages of the environment, along with a brief description.

3.4.1 Home Page

The home page of the application is presented to the user when first visit the website and is displayed in Figure 16. To create a clean and modern design, a simple homepage was chosen. In order to access other pages, the user must use the *navbar* that presents all the actions that can be performed.

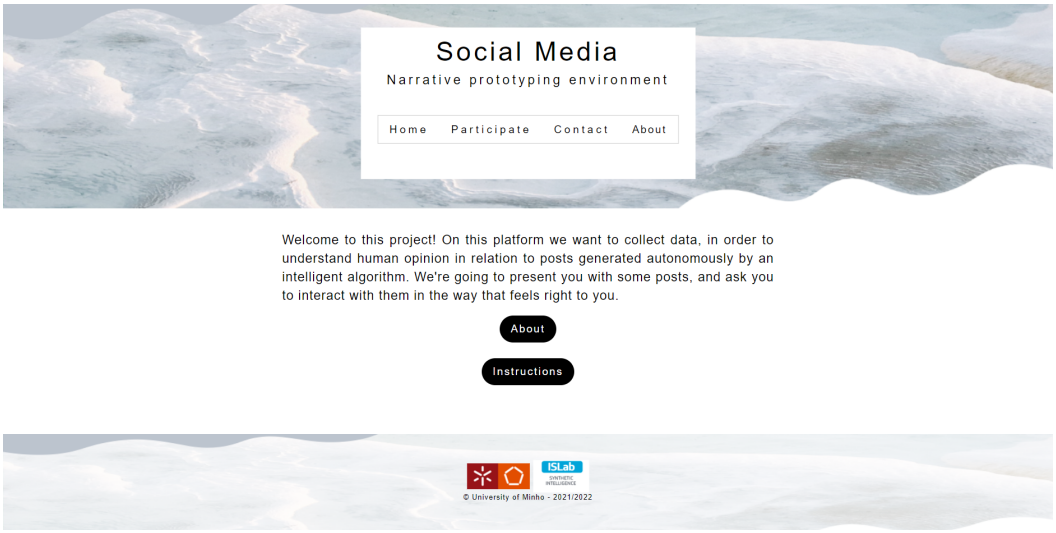


Figure 16: Home page

3.4.2 New Participation

When a user first accesses the section *Participate*, a pop-up window is displayed asking the user to consent to the tracking of their actions and to confirm that they are at least eighteen years old, as shown in Figure 17.

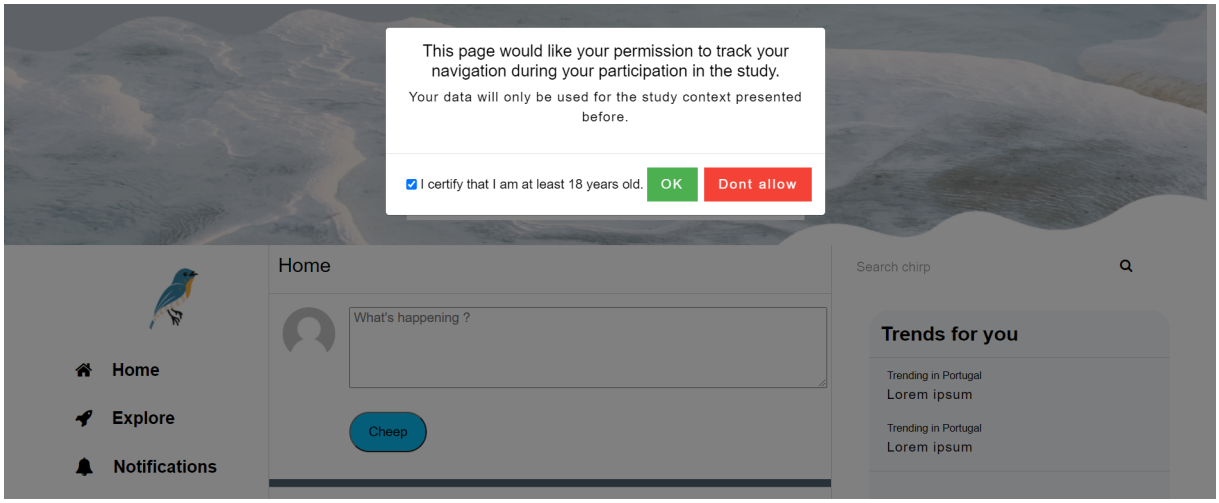


Figure 17: User consent.

The participation page, in Figure 18, presents an design of a social microblog, showing to the user a set of posts with which it is possible to interact.

On this page clicking on the button: "Do you need help", a window will appear explaining how to interact with the platform in all available actions. As can be seen in Figure 19, each of the selectable options is explained to the user.

It is worth noting that some microblogging platform features, such as the user profile or private messaging section, were not implemented since they don't are relevant; nonetheless, they are marked.

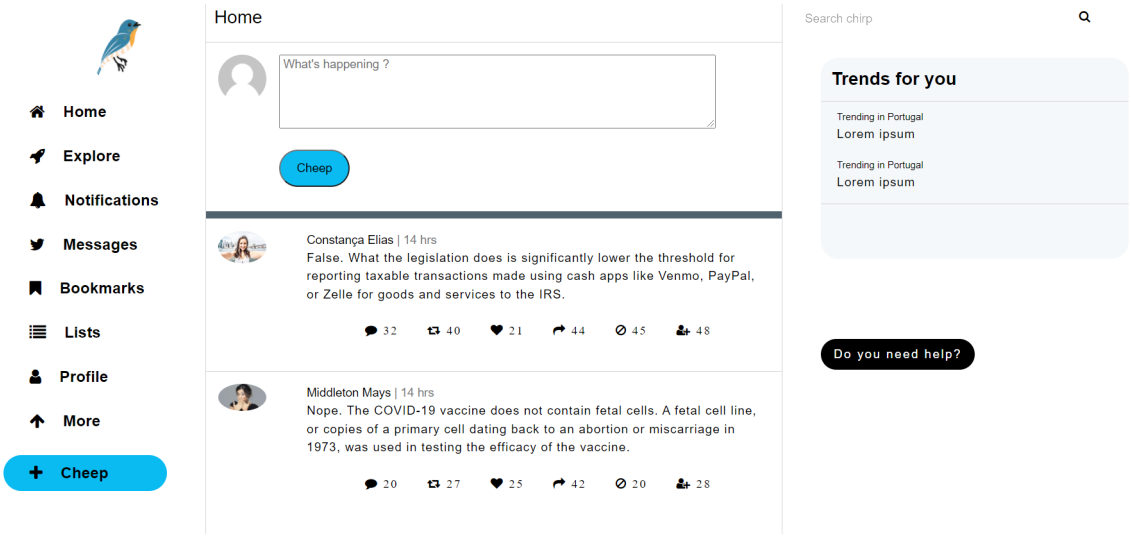


Figure 18: Participation Page

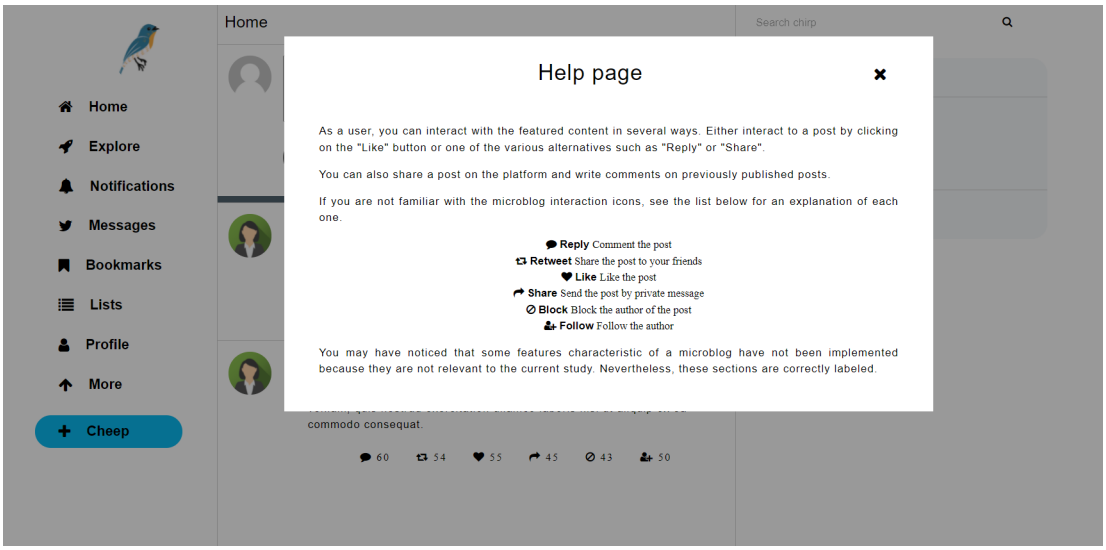


Figure 19: Help Page

3.5 Conclusion

This Chapter presents a simulator for a microblogging environment that provides an ethical approach to prototyping and research. This tool was created in response to the realisation that using social media for research raises a variety of ethical difficulties that are challenging for investigators to resolve. The final result is a web application, developed using Django framework, with a microblogging aspect that can be used by researchers interested in conducting online studies in the context of a laboratory environment.

Chapter 4

MentaLex: a Mental Processes Lexicon

How people act in different situations, especially on social media websites, is greatly influenced by their personalities. These characteristics can be observed in a persons writing or speaking style as well as other facets of their social interactions [47]. Similar to language, writing is a characteristic that is particular to each individual. In another perspective, an analysis of the frequency of terms can identify letters sent by soldiers, speeches given by politicians, and authors of literary works [10].

For the Post Generator it will be necessary to add or enhance personality in the posts for a greater connection between social media users and posts. To this end, an analysis of personality-related works was conducted. Since the results found did not meet the requirements of the project, the construction of a personality lexicon was undertaken.

4.1 Related work

The current state of the art in Big Five (big5) [1] [39] personality trait detection can be illustrated by Table 9. Similar numbers are presented in other surveys [48]–[50] and papers [51]–[53]. In short, the average scores ranges on 60% except for Ren *et al.* [54] that appears as an outlier with an strategy based on BERT with SenticNet on CNN (it was not possible to easily find the code for reproduction).

The Essays dataset [10] is being studied for 20 years and the prediction values have not yet increased consistently [49]. It is intended to do a step back and, instead of trying to develop a new predictor, to build a lexicon. This is inspired by the EmoLex [55], [56] that can be considered a gold standard for emotion detection and has been used for around 10 years with more than 1000 citations on each paper. It is expected that a lexicon as such may provide, at least, useful insights for advancing the state of the art.

[1] The Big Five personality theory is a hierarchical organisation of personality traits based on five basic dimensions: Extroversion, Amiability, Consciousness, Neuroticism, and openness to experience [39]

Authors	Opn	Con	Ext	Agr	Neu	AVG	Proposals
Xue *et al.*	63.16	57.49	58.91	57.49	59.51	59.312	GLOVE and BiGRU model
Ren *et al.*	**80.35**	**80.23**	**79.94**	**80.30**	**80.14**	**80.192**	BERT with SenticNet on CNN
Ramezani *et al.*	56.30	59.18	64.25	60.31	61.14	60.24	HAN based on: term frequency vector, ontology, enriched ontology, LSA and BiLSTM.
Wang *et al.*	64.80	59.10	60	57.70	63	60.92	Personality GCN
Jiang *et al.*	65.86	58.55	60.62	59.72	61.04	61.158	Five models developed: ABCNN, ABLSTM, HAN , BERT, RoBERTa and BERT + RoBERTa
Mehta *et al.*	64.6	59.2	60	58.8	60.5	60.62	SVN + MLP
Kazameini *et al.*	62.09	57.84	59.30	56.52	59.39	59.028	SVM
Majumder *et al.*	62.68	57.30	58.09	56.71	59.38	58.832	Word2Vec and CNN
Tighe *et al.*	61.95	56.04	55.75	57.54	58.31	57.918	PCA

Table 9: The state of the art in personality detection. For full references about the mentioned studies see [50]

In 2010, Tal Yarkoni [57] conducted a thorough examination of the personalities and words used in 694 blogs posts. He performs a big5 questionnaire to the authors in order to categorise their personalities. The outcomes included a summary of the relationships between the big5 and the 20 words that more represent each trait. However, there is no information regarding a potential lexicon developed from the study results. Since then, no significant research on the creation of personality lexicons has been found.

4.2 Procedure

The Essays dataset [10] is composed of 2468 open topic essays written by people whose personality was already assessed. Therefore, together with each essay there are five personality traits (also called big5 as proposed by [39]) binary flagged.

A big5 assessment is made through a form from which the subject chooses a set of words that are considered most related to himself or herself [39]. Analogously, the hypothesis explored is that, the personality trait of a person would be revealed by the words chosen for creating a text. For testing this hypothesis, the tf/idf *cf.* [58] value is computed for each lemmatized word on each personality trait, excluding those whose tf/idf value is 0.0. That procedure, shown in algorithm 1, results in a lexicon with the words commonly chosen and not used words by each personality.

The deducing hypothesis is that a set of words is meaningful enough to provide at least a personality insight. For that claim to be true, the identified words shall be “insistently”used on different texts as a sort of “trail”. For testing that deduced hypothesis, the Twitter Personality dataset [9] with 9879 tweets is used to verify how in-depth the resulting lexicon is. That dataset was chosen because in little texts the probability of a word happening is reduced compared to larger texts.

Algorithm 1 Meaningful Words per Trait.

```
Require: Essay_Dataset ← pipe([lemmatize | remove stop words | standardize in lowercase | extract adjectives, verbs and nouns]).
  procedure getWordsFor(trait)
      trait_texts ← retrieveTextFor(trait, Essay_Dataset)
      TF/IDF ← computeTF/IDF(trait_texts)
      for word, tf/idf in TF/IDF do
          if tf/idf > 0 then
              Words ←+ word
          end if
      end for
    return Words
  end procedure
  for trait in big5 do
      Used_Lexicon ←+ getWordsFor(trait)
      Avoided_Lexicon ←+ getWordsFor(big5 - trait)
  end for
```

4.3 Results and Discussion

By the assessment made upon the Essays dataset it was not possible to find all the 5 personality traits. It was possible to distinguish agreeableness and neuroticism with almost disjoint word sets. However, the word-set for extroversion, openness and conscientiousness are overlapping (despite being almost disjoint with the other two). That experimental result may suggest that these three traits may be considered one when personality is abducted from open topic essays. An alternative cognition theory that suits better with these findings is the one proposed by Vann Joines [59] called “Personality Adaptations“ (the relation between big5 and Personality Adaptations is presented in [39]).

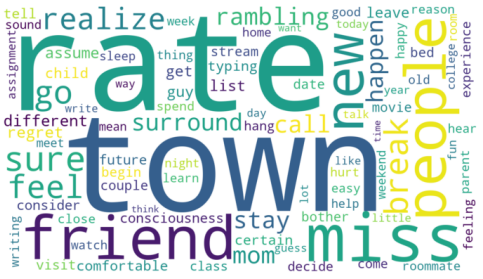
Personality Adaptations theory suggests the existence of six mental processes, three of them primary and the other three secondary. Each person has at least one primary process and may or may not have a secondary one [59]. What is suggested is that the primary processes that are being revealed when someone writes a text. Accepting that suggestion, adapting from the Personality Adaptations theory, the primary processes would be: paranoid, schizoid and neuroticism. For a reference, the terms associated with the respective mental process and the unrelated words are presented as word-clouds in Figure 20, constructed directly by calculating the tf-idf value described in algorithm 1.

As can be noticed the number of words related to each mental process is actually small. To evaluate its suitability of the lexicon, the Twitter Personality dataset is used to verify how many tweets are composed by words that are in the lexicon (coverage analysis).The generated lexicon, composed by 276 words, presented a coverage of 69% in that dataset. As an attempt to increase the coverage the wordset, synonyms were retrieved resulting in a dataset with 3432 words covering 81% of the dataset. That coverage is considered sufficient due to the tweets’ peculiarities; *i.e.* very short unstructured texts with misspelled words and slangs, eventually, with only a link or emoji. The frequency of the most commonly used words in the lexicon in the Twitter dataset, as shown in Figure 21, was within a suitable distribution. It can be seen that although some words are more frequent, the amount of lexicon words present in the Twitter dataset is large.

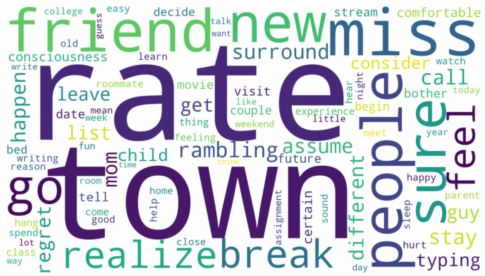
As an analogous procedure, upon the 316 words found to be avoided, their synonyms was gathered



(a) Neuroticism



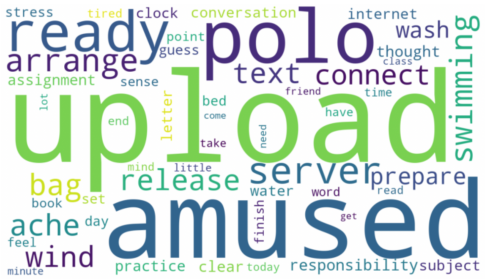
(b) ¬Neuroticism



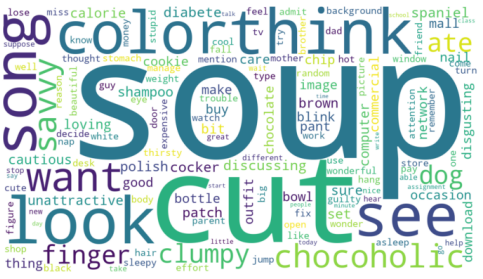
(c) Paranoid



(d) ¬Paranoid



(e) Schizoid



(f) ¬Schizoid

Figure 20: Word-cloud that includes both the words connected to each mental process and the words that are not.

resulting into a lexicon of 3639 words.

For a summary, the lexicon characteristics are presented in tables 10 and 11.

	words	synonym	words & synonym	%	nouns	verbs	adj.
paranoid	87	1003	1090	32%	523	443	124
schizoid	53	762	815	24%	407	335	75
neuroticism	136	1391	1527	44%	793	560	174
total	276	3156	3432	100%	1723	1338	373
coverage	69%	–	81%	–	–	–	–

Table 10: Summary of commonly used words Lexicon

	words	synonym	words & synonym	%	nouns	verbs	adj.
¬paranoid	85	766	851	23%	399	125	327
¬schizoid	144	1554	1698	47%	827	235	636
¬neuroticism	87	1003	1090	30%	523	126	449
total	316	3323	3639	100%	1749	482	1408

Table 11: Summary of avoiding prone words Lexicon

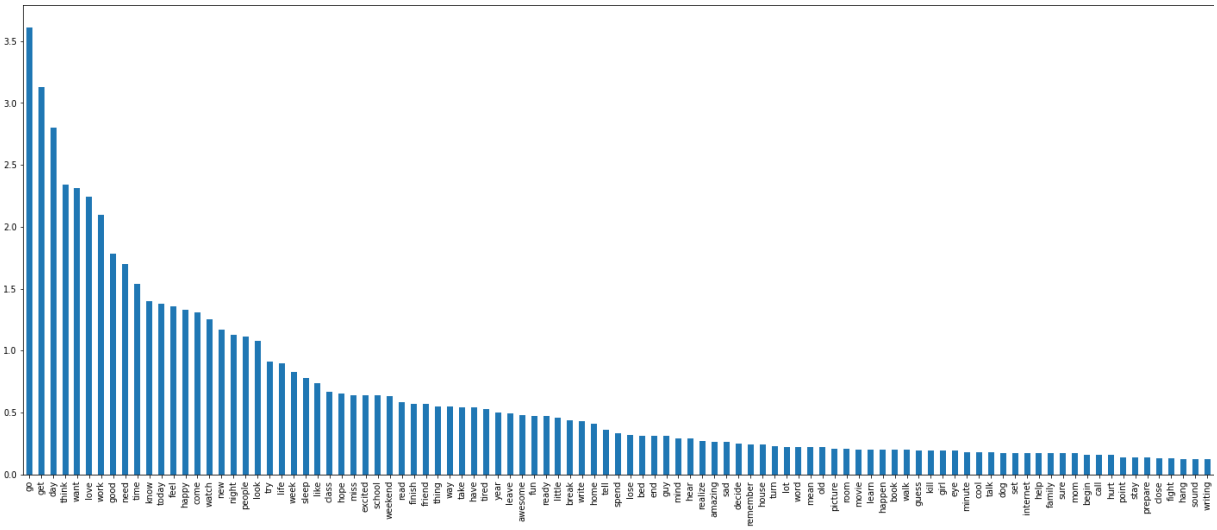


Figure 21: The frequency of partial mentaLex words in the Twitter dataset

4.4 Conclusion

This chapter examines how individual personality traits link to writing style and presents an mental process lexicon. Considering that the performance of personality predictors are not consistently increasing throughout the years, as an alternative, the idea was to provide big5 personality traits lexicon based on the Essay dataset. The weight of relevance for every trait of each word is calculated by tf/idf on the words in the Essay dataset. However, it was eventually realized that three personalities wordset were overlapped requiring a change of course. The research changed the underlying model from big5 to Vann Joines' mental processes as it appear to suit better with the empirical findings. The resulting lexicon dataset is composed with 3432 commonly used words and 3639 avoided prone words for each mental process. The commonly used words are capable of covering 81% of the Twitter Personality dataset.

Chapter 5

Post Generator

This chapter present the approaches and strategies developed to obtain the Post Generator. As a remainder, a post is an online publication that usually consist of four different elements: *Text Core*, *Emojis*, *Hashtags* and *Links*, as shown in the Figure 22. The *Text Core*, remains the textual part of the post that contains the relevant content/message. An *Emoji* is a pictogram that can be embedded in the *Text Core*. *Hashtags* are search terms on social media that can be included in the post. And finally, *Links* are URLs to websites.

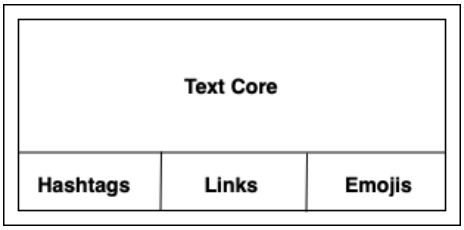


Figure 22: Social media post representation

5.1 Text Core, Information Retrieval

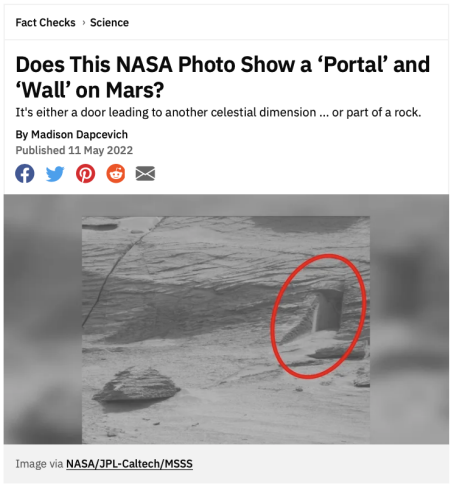
Looking at the Figure 22, this section goal is to obtain the post element *Text Core*. And, the following requirements are established for this:

- Consistent. This involves avoiding contradictions and establishing a connection between the textual elements.
- Coherence. Have a meaningful text flow.

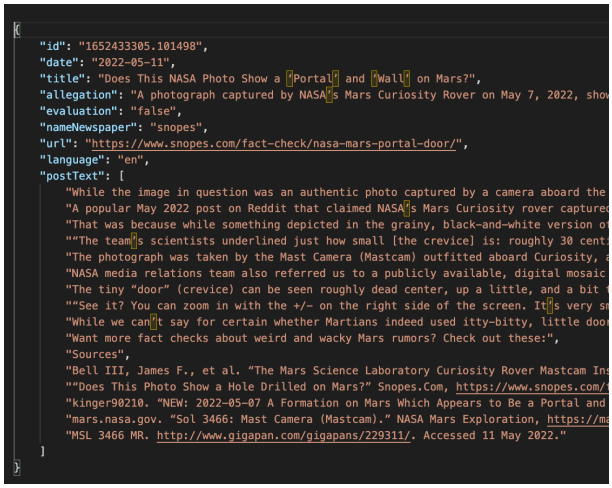
- Informative. Provide context with relevant information.
- Don't produce fake-news.
- Have less than 280 characters (to meet Twitter's requirements [60]);

Fact-checking news from the newspaper Snopes was used as an information source. To build a dataset of fact-checking news an extractor for this website was been developed. The tool scrapes the web page using the *beautifulsoup* library and extracts the following fields from the news: URL, date, newspaper name, title, allegation, classification and the news content. The result is stored in a JSON file that will be used as input by the Post Generator.

Figure 23 shows an example of the result of scraping a news item in SNOPES. In Figure 23b are the extracted information and the names of the variables in the created JSON file. In addition to the assigned *id*, publication *date*, *title*, *url* and *language*, the *allegation* and *evaluation* fields stand out, which respectively refer to the theme or content that will be verified and the conclusion of the newspaper for its veracity. The *postText* field stores the body of the news with the information presented in the newspaper website to conclude on the true of the claim.



(a) Headline of a SNOPES news.



(b) Result of SNOPES news scraping

Figure 23: Extraction example

Having created a dataset, it is now time to describe how to obtain the post component, *Text Core*. To this end, the summarisation methods used are described in section 5.2.

5.2 Summarisation

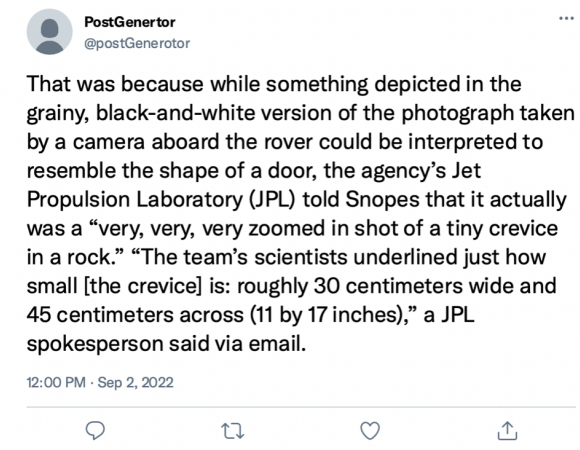
To obtain the *Text core*, Summarisation techniques are used. Summarisation is about the computational process of text length reduction with the aim of producing a condensed set of text that contains the most important information from the original material [61]. Text summarisation is broadly divided into two classes

[61]: **Extractive Summarisation** in which the summary is consisting entirely of content extracted from the input [62] and 2) **Abstractive Summarisation** which use Natural Language Generation (NLG) for paraphrasing the main concepts of a given text using a different set of vocabulary [62]. As an observation, the aim is to summarise the content of a set of sentences that make up a news in a maximum of 280 characters. The difficulties experienced in each of the techniques will be described throughout this section.

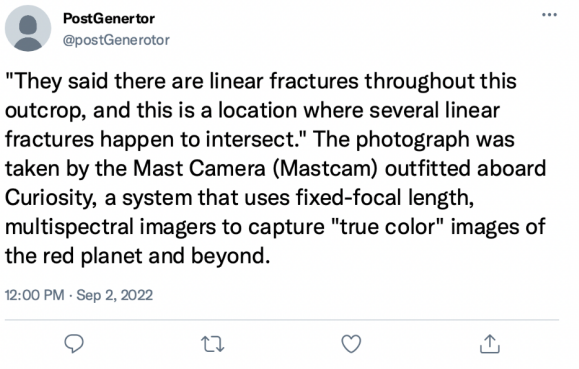
For the **Extractive Summarisation**, *TextRank* algorithm was first used. This unsupervised method of text Summarisation is based on the idea that words that occur more frequently are significant [63]. Hence, the sentences containing highly frequent words are important. It uses a ranking algorithm that assigns a score to each sentence in the text [63]. The highest scoring sentences are then included in the summary. The algorithm that implements the *TextRank* method was developed using Python's *NLTK* libraries to perform a count of the number of words in the text.

Other algorithm tested was *Luhn's Heuristic Method* that uses TF-IDF to scoring the importance of each sentence [64]. The idea is that any sentence with maximum occurrences of the highest frequency words are less important to the meaning of document than others. The significant words are those which are more often in the text but at the same time, they don't belong among stopwords [64]. The most significant sentence make the summary.

Figure 24 presents an example of the results obtained by the previously Extractive summarisation algorithms.



(a) Output from the TextRank algorithm.



(b) Output from the Luhn algorithm.

Figure 24: Extractive Summarisation example. The models receive as input the JSON with the information of the news presented in 23b

Still in the *Extractive Summarisation*, other methods were applied to obtain the *Text Core*. The method presented next derives from a result obtained in the Systematic Review carried out in the Chapter 2: tweets that present sadness and surprise are more likely to attract users. To apply this result, the emotions of the news phrases (previously extracted to the *postText* field of the JSON file) were analysed. Then

EmotionsMeter.com was used to score the emotions of each phrase. The sentence with the highest emotional value for sadness and surprise is chosen as the summary. An example, is show in Figure 25

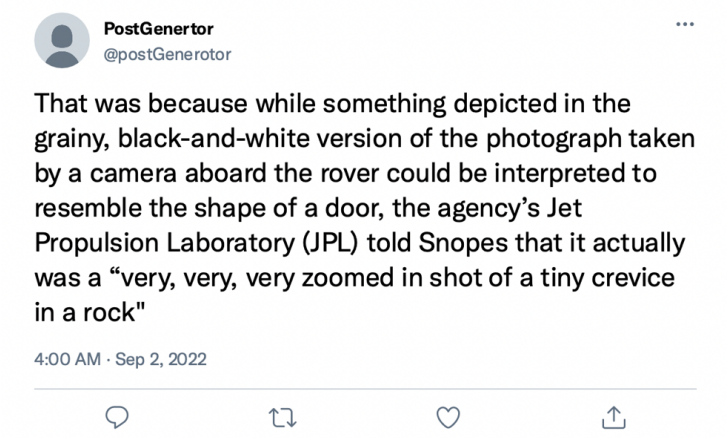


Figure 25: Text core as a result of selecting the sentence with the highest value of surprise and sadness.

The last *Extractive Summarisation* approach tested to getting the *Text Core* is the least complex and the most conservative. No new content is created, the *Text Core* result from the connection of the fields *allegation* and *classification* extracted from the news. That is, information about the claim to be debunked and the classification assigned by SNOPES. For the example that has been used, the application of this approach obtains the content in Figure 26.



Figure 26: *Text core* created by connecting the fields: allegation and news classification.

For the **Abstractive Summarisation** different pre-trained models are used: GTP-2, BART, BERT, T5 and XLNet. A simple overview of these models is presented in the Table 12.

Model	**Version**	**Description**
GTP-2 [1]	gpt2-medium	Generative Pre-trained Transformer 2, open source, developed by OpenAI. Is a unsupervised language model without task-specific training [65].
BART	facebook/bart-base (Model with 6 encoders and decoders layers).	BART employs seq2seq translation with both the BERT (bidirectional encoder) and GPT (left-to-right decoder) architectures [66]. The model is pre-trained on CNN/Daily Mail data.
BERT	bert-large-uncased	BERT (Bidirectional transformer) is a pre-trained bidirectional transformer used to overcome the limitations of RNN and other neural networks. [67].
XLNet	xlnet-base-cased	The XLNet model, which incorporates permutation language modelling in its architecture, is an improved version of the BERT model. XLNet is a bidirectional transformer as well, and it predicts the subsequent tokens in a random order [68].
T5	t5-base	T5 is a transformer model from Google that is trained in an end-to-end manner with text as input and modified text as output. This model achieved the state of the art results in summarisation taks [69].

Table 12: Abstractive summarisation models

The Python modules *transformers* and *bert-extractive-summarizer* were used to create these models. An English word has 4.7 characters and a requirement for the *Text Core* is that there are no more than 280 characters. Therefore, it was calculated that the model can issue about 60 tokens ($\frac{4.7}{280}$). However, maximum number of tokens to be created was set at 50 for all models *(max_length = 50)*. This choice was made to avoid potential words not conforming to the rule (± 5 characters per word) and to make

[1] The GPT-3 is more robust than the previous version as it uses a large amount of data in the pre-training phase. However, it was not used, as it is not available as open source.

room for more information to be included in the post in later stages. The *postText* field of the extracted JSONs files (explained in detail in Section 5.1), served as input for the models.

5.2.1 Summarisation approaches Evaluation

To analyse the results obtained by the previous models, one hundred news items were randomly selected from the dataset and each of the previous approaches was applied. The outputs were accessed with a complementary evaluation: Automated evaluation and Human evaluation.

Automated evaluation is about ensuring that no fake-news content is generated and that the maximum length of the post is respected. The fake-news verification was performed using the tool *Fake news classifier* [70]. This classifier uses the BERT model, which during validation reached 98% of accuracy and 99% of Recall and Precision values.

Human evaluation was carried out through a questionnaire applied to a group of volunteers to analyse the coherence, consistency and informative character of the results. The need to this validation is associated with the subjective character of the evaluated parameters.

Firstly, the automated evaluation was performed, and the results organised in Table 13. Analysing the size of the summary produced, *bert, gtp-2* and *xlnet* models generate 40% of posts with more than 280 characters. It was therefore necessary to check if the equality of this number was just a coincidence. Analyse the text produced by each of these models show that 95% of the summaries are the equal. So it does not make sense the 3 models move to the next stage, since they produce the same result. GTP-2 was chosen to represent the results of this group. However, the GTP-2 model was next discarded, since leads to meaningless sentences (96% of the generated tweets are irrelevant, without knowledge or information related to the fact-checking news), as shown for the previous example, Figure 27.



Figure 27: Result of news summarisation by GTP-2 model.

For extractive summarisation, procedures based on *Luhn's Heuristic Method* and *Text-Rank-NLTK* were discarded because the maximum post size is not respected in most of the examples. The emotion selection approach was also abandoned as it created 37% of posts longer than 280 characters and almost half of them were classified as fake-news. A possible explanation for this may have to do with the highest values of surprise and sadness being associated with the presentation of fake-news to be classified by the journal.

Approach	Size (0<N<280)	True context	Comment
T5	86%	96%	
BERT	60%	–	repeated
BART	100%	94%	
XLNet	60%	–	repeated
GTP-2	60%	–	meaningless
Luhn's Heuristic method	23%	–	exceeds size
Text-Rank-NLTK	15%	–	exceeds size
Emotion-selection	63%	53%	fake-news
Allegation sentence	100%	100%	

Table 13: Summary of auto-evaluation results on a sample of 100 fact-checking news extracted from Snopes. The red colour is associated with disregarded approaches and the green colour with approaches approved by automatic evaluation. The **-** means that the evaluation was not carried out since the approach had already been rejected.

Next, a human validation was performed in the content generated by the models approved in the automatic evaluation: *T5*, *BART* and *Allegation sentence*. Based on [11], eight people were recruited to review the results of each remaining model. These participants are students from the University of Minho - Braga, 5 men and 3 women, aged between twenty-two and twenty-five. For each of the posts, they were asked to rate in a binary way, giving 1 if the post is informative, coherent and consistent or not. The results obtained in each model are summarised in Table 14.

Approach	Acceptance (%)
T5	79%
BART	25%
Allegation sentence	83%

Table 14: Summary of Human evaluation. Acceptance was calculated as a fraction of the number of 1 given by the user (n) and the number of posts rated for each approach: $\frac{n}{100} \times 100\%$

The results show greater acceptance of the T5 model and Allegation sentence. By other side, the results of the BART model have low acceptance among voluntaries in the study. The main reason is that in many cases the generated sentence is not complete and part of the message remains incomplete, as show Figure 28. That means, the size is respected but consistency and informative character is lost.

As a final remark, T5 models and Allegation sentence were selected as approaches to obtain the *Text Core* element of the post. Although the T5 model has a 14% chance of generating a post with a size greater than 280 characters, the content produced has only 4% of possibility of create a fake-news. This model

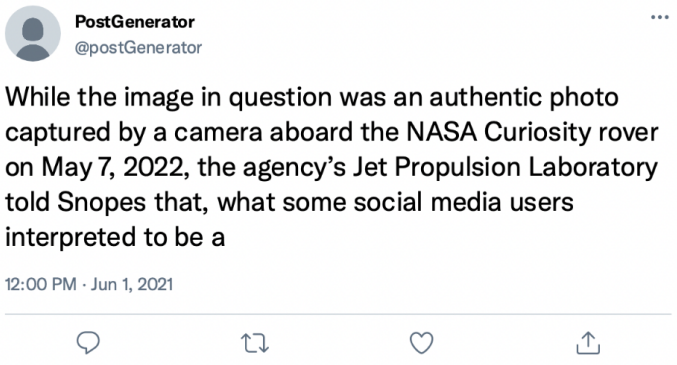


Figure 28: Result of news summarisation by BART model.

has 79% of acceptance between the study participants. The use of the Allegation sentence, in addition to not producing fake content and the size being always within the expected, the acceptance rate in the human test performed was 83%.

5.3 Emotional Reinforcement

The systematic review conducted in Chapter 2 conclude that the use of emotions increases engagement and sharing of posts on social media. The post elements: *Emojis* and *Hashtags* are added to the previous *Text Core* creating a emotional appeal. The way each of these elements are included is explained next.

Emojis allows make the text fun and interactive and add an extra layer of emotion and personality to the post. Tweets with emojis receive 25% more interaction than those without [71]. They go beyond mere images, they are a new language that facilitates the expression of emotions in environments where face-to-face interaction is not possible. Its insertion humanises a post and increases its dissemination with the audience by creating an emotional connection with the reader. Every day, 95% of social media users send 10 billion emojis [72].

Kralj Novak's *et. al* in [73] presents an Emoji Sentiment Lexicon that associates each emoji with a sentiment score ([-1,1]). Based on this lexicon, emojis were divided into positive and negative categories. For this work, the first category has been associated with the news classified as True. The remaining emojis were associated with posts related to news classified as fake or untrue. The content of Table 15 shows an example of this relationship, where the choice of the emojis (from the positive or negative group) is done using the *classification* field of the news JSON files.

Emojis will be added to the generator in different number depending on the amount of space available (respecting the 280 character limit). However, the maximum number of emojis that can be added is four, to prevent the user from getting tired [74]. The post is updated with emojis before hashtags because they are more successful in increasing engagement rates [75].

Hashtags are other post element. Using them, help people locate particular tweets when they search

Sentiment	News Classification	Emojis
Positive	true, research-in-progress, mostly-true	💕, ❤️, 👋, 🔥, ☺️, 😛 ...
Negative	false, misattributed, unproven, labeled-satire, mostly-false, scam, mixture, outdated	😩, 😡, 😫, 🚫, 👎 ...

Table 15: Example of how emotions, news classification (based on SNOPES), and emojis relate.

for them. Specially if they are current and relevant. However, employing too many of them, especially the irrelevant ones, could have the opposite effect and encourage users to keep scrolling through their feed.

First, for adding *Hashtags* to the *Text core*, the *postText* field of the JSON file was cleaned: Stopwords and punctuation were removed, and the words were lemmatised. Then, the unsupervised machine-learning model LDA (Latent Dirichlet Allocation) is applied, receiving as input the previously cleaned text. The topics identified by the models are extracted and emotion score evaluated. Finally, every topic found with emotion is added to the post as a hashtag.

To heighten the emotional side of the post, the most important aspects were highlighted using capital letters. This allows the user to identify the content of the post more quickly, which can ensure greater engagement. For this, the post's *Text core* is submitted to the *KeyBERT* model to identify the keywords. The result, are highlighted in the text in capital letters. This approach doesn't add new information to the post, it just reinforces the most relevant one.

For better understanding, a result of application the previous components is shown in Figures 29 and 30. As can be seen, the Figures 29b and 30b are obtained from Allegation sentence and T5 approaches (Figures 29a and 30a), enhanced by emojis, hashtags and capital letters. The emotional assessment performed by *emotionsMeter.com* proves an increase in the percentage of emotions in posts.

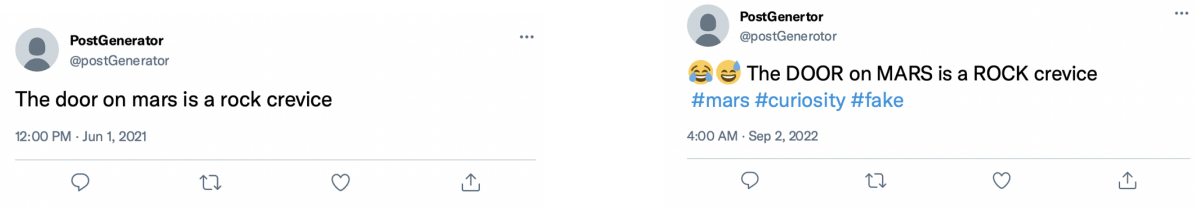


(a) Results obtained by *emotionMeter.com*: 'fear':0.0, 'anger':0.0, 'anticip':0.0, 'trust':0.0, 'surprise':0.0, 'positive':0.0, 'negative':0.0, 'sadness':0.0, 'disgust':0.0, 'joy':0.0



(b) Results obtained by *emotionMeter.com*: 'fear':0.0, 'anger':0.0, 'anticipation':0.0, 'trust':0.0, 'surprise':0.0, 'positive':0.0, **'negative':1.0**, 'sadness':0.0, 'disgust':0.0, 'joy':0.0

Figure 29: a)Post generated using Allegation Sentence approach. b) Post after emotional reinforcement.



(a) Results obtained by *emotionMeter.com*: 'fear':0.0,'anger':0.0, 'anticipation':0.0, 'trust':0.0, 'surprise':0.0, 'positive':0.0, 'negative':0.0, 'sadness':0.0, 'disgust':0.0, 'joy':0.0

(b) Results obtained by *emotionMeter.com*: 'fear':0.0, 'anger':0.0, **'anticipation':0.25**, 'trust':0.0, **'surprise':0.5**, 'positive':0.0, **'negative':0.25**, 'sadness':0.0, 'disgust':0.0, 'joy':0.0

Figure 30: a) Post generated using T5 model. b) Post after emotional reinforcement.

5.4 Personality Reinforcement

So far it has been explained how to get the elements of the post and reinforce them emotionally. In this section, the attention is on the personality reinforcement. For this, it is used the lexicon created, *mentaLex* (explained in Chapter 4), which gives an association between a set of words and primary mental traits of Vann Joines. The development of this lexicon change the course of work and personality reinforcement is adapted to a mental process reinforcement.

Before starting the reinforcement, it is necessary to identify the mental process that stands out in the current state of the post. For this, a simple classifier was developed, inspired by emoTFIDF [76], which integrates the TF-IDF value to obtain the post relevance score for the three primary mental process (schizoid, neurotic and paranoid). The post is submitted to mentTFIDF, and the output will be the predominant mental processes. Next, the predominant mental process is reinforced by replacing, whenever possible, words that do not belong to that mental process with synonyms that do. Figure 31 shows the application of the previous method.

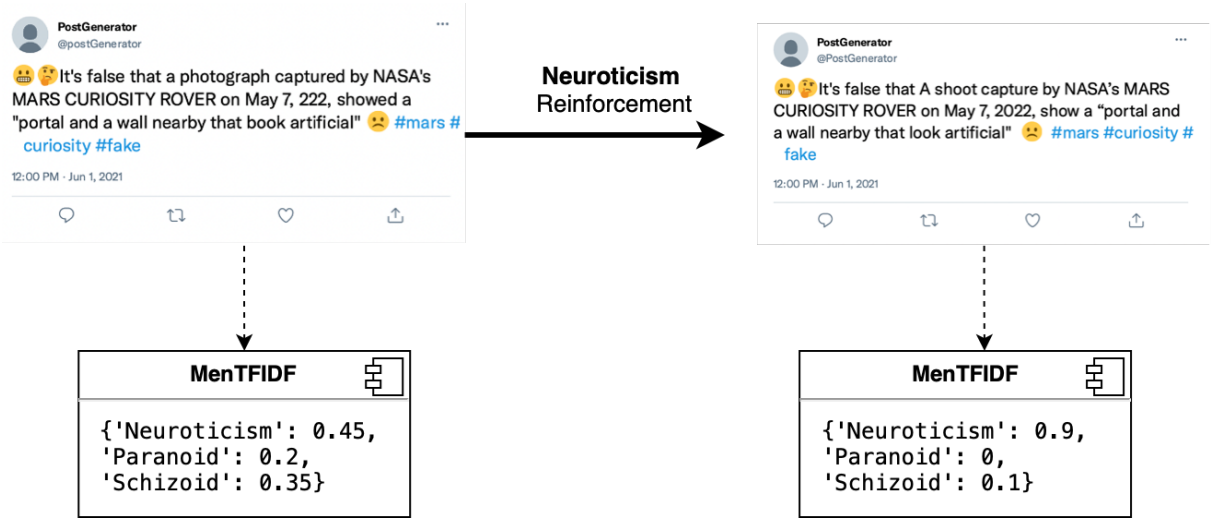


Figure 31: Example of Neurotic mental process reinforcement.

5.5 Conclusion

This chapter presents the design of the Post Generator algorithm to achieve an emotional microblogging post that is amplified by a mental process. In this way, the Post Generator is the result of the strategies tested with better results.

The Post Generator algorithm, receives a JSON file, containing the news content. Then the post components are obtained. The algorithm starts by using the *postText* field of the JSON file to generate the *Text core*, which as discussed earlier, is obtain by the *T5 model* for summarisation or selecting the *Allegation Sentence*. Then emojis and hashtags are added to the post in order to reinforce its emotional side. In addition, the *Text Core* keywords are highlighted in capital letters. Next, the predominant mental process in the post is identified using the classifier *menTFIDF*. The final post, results from the reinforcement of the predominant mental trait in the previous post.

Chapter 6

Results and Discussion

After develop the Post Generator algorithm it was possible to design an experiment to answer the research hypothesis presented in 1.3, which considers the possibility that emotions and personality, when added to a post, increase microbblogging users engagement. The following sections present the aspects of the experiment: the design, results and the discussion.

6.1 Preparation

The experiment was attended by twenty participants, between September 12 and 18, 2022. They were Twitter users recruited at the University of Minho and their participation was voluntary. All participants were informed of the context of the work in which this study was conducted and that no inferences could be drawn about them from the data collected.

The first step of the experiment was to provide participants with access to the microbblogging environment (detailed in Chapter 3) in which the study was conducted. To do this, they were given access to the platform through a computer owned by the researcher. Participants were then asked to interact with the platform as they do when they scroll through their Twitter feed. It is important to note that the Platform configuration for the study has the interactions: like, block, follow, retweet, reply, share and post. Furthermore, the default number of interactions for each post was randomly set in a range between 0 and 20. The posts, included in the study and presented on the platform are extracted from the SNOPES Twitter page and others generated using the Post Generator Algorithm (detailed in Chapter 5). The voluntaries received posts relate to ten news items. In total, there are 30 posts: 10 generated by T5 model, 10 generated by allegation sentence and 10 posts extracted from the SNOPES Twitter page. The list of posts used in the experiment can be found in the Appendix A. All this posts are presented to the user on the platform without him knowing what was created by the Generator and what came from the Snopes Twitter page.

It is important to mention that the order of presenting the posts was random. After that, each participant should submit their participation. With this submission, the study is completed.

6.2 Results

With the procedure completed, it is possible to analyse the voluntaries data. For that, all the interactions associated to each post were structured in tables, included in Appendix B. From those data come the findings detailed in this section.

First, some of the demographic data of the participants profile were examined to provide some context about them. Of the twenty participants, seven are Female and thirteen are Male, all between the ages of twenty and twenty-eight. The bar chart in Figure 32 shows the number of men and women by age. It is clear that the mans have a wider age range compared to the woman's.

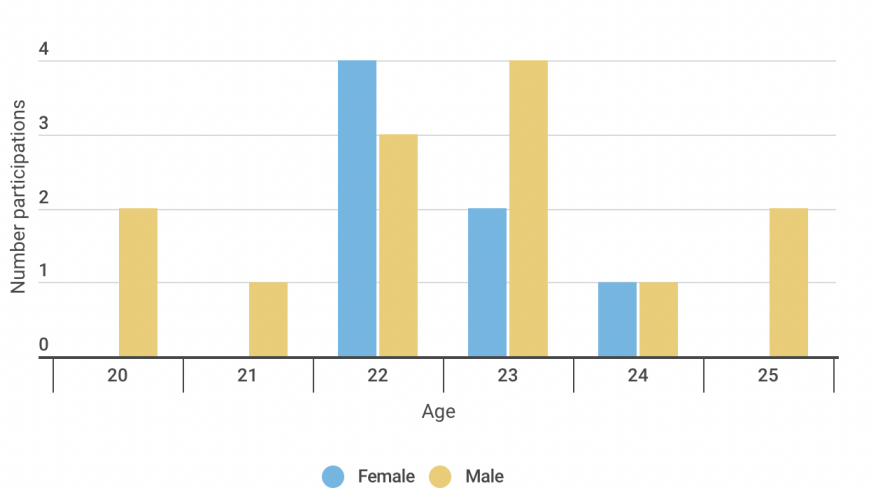


Figure 32: Voluntaries gender distribution by age.

These participants are Portuguese students from four courses at the University of Minho, as can be seen from the diagram in Figure 33, which shows the number of participants from each course. From the diagram, it can be concluded that 60% of the participants are students of the Master's degree in Computer Engineering and they come from the same circle of friends.

After analysing the participants profile and demographic data, it is time to interpreted the results of the experiment to understand the reaction of the participants to the different social media posts.

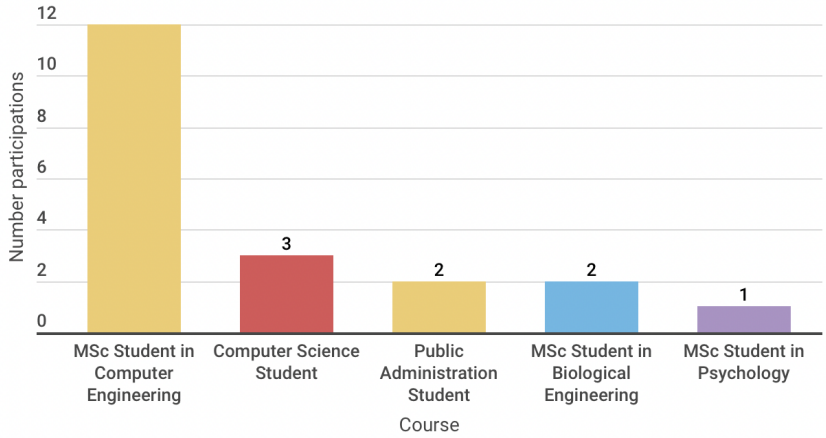


Figure 33: Voluntaries distribution by course.

The bar chart presented in Figure 34 shows that *like* was the most frequently used interaction by participants, followed by *follow* and *retweet*. On the other hand, the interactions *reply* and *post*, which imply writing, were the least used by participants. In total, there were 434 interactions with the platform, which corresponds to an average of 21 interactions per participant and 14 interaction per post. The study [77], conducted on Twitter, collected 120026 tweets (from 70173 unique profiles) about fact-checking news and analysed the interactions these tweets received. The results are consistent with those obtained in the experiment presented in this document: likes and retweets are the most common interactions performed by users.

As an observation, in this study, unlike Twitter, the *block* interaction is associated with the post and not with the author's page. This means that if the participant *blocks* some post has a negative connotation towards it. It takes a position contrary to the like, and its use shows a lack of interest in the post.

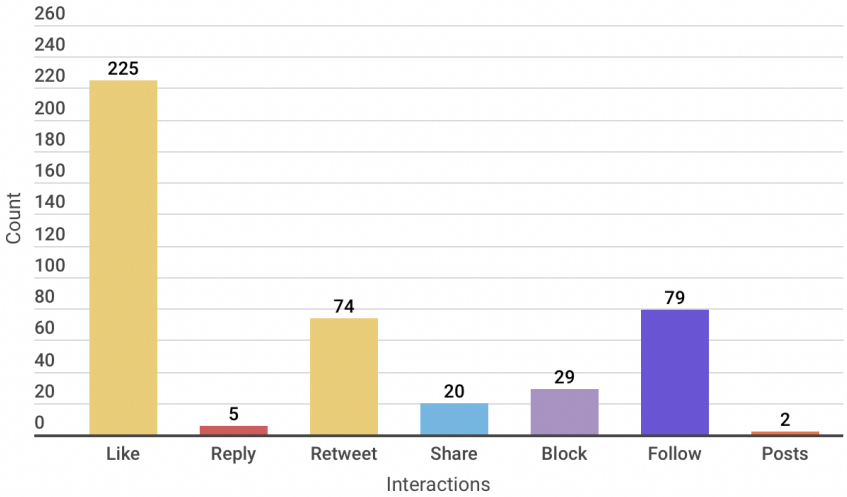


Figure 34: Global Interactions count.

The bar chart in Figure 35 shows the relationship between the approaches used (T5 model, Allegation

Sentence and Snopes tweets) and the number of interactions receive. The number of interactions with the tweets extracted from Snopes Twitter page is very low and a huge part of them are negative interactions (*block*). This low number could be related to the structure of the post: formal and without emojis or hashtags. The exception was the post in Figure 36, which refers to a manipulated image (a plane crashing into a bridge) with positive interactions (*likes, retweets and follows*). In the posts created by the Post Generator, the Allegation Sentence approach, despite being less sophisticated than the T5 model (for text summarisation), had the highest number of interactions.

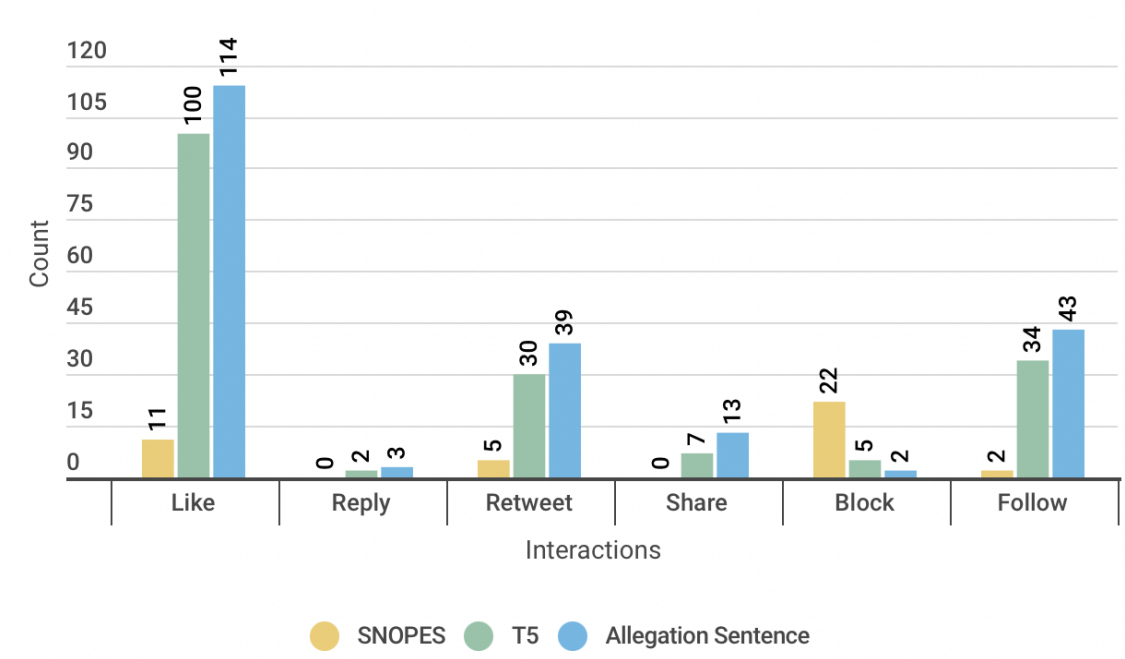
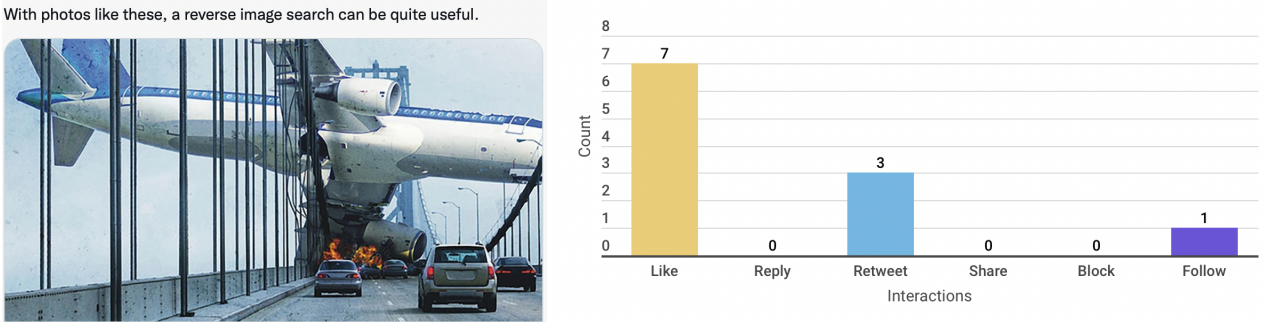


Figure 35: Interactions count per approach



(a) (b)

Figure 36: a) Post 21: Extracted from SNOPES twitter page; b) Interactions count.

The data from Figure 35 allows calculate the average engagement rate for the three types of posts. The quality of social media engagement can be accurately determined by a function of $\frac{interactions_count}{visualisations}$. In a non-lab environment determining the number of visualisations is a difficult task, because visualisations

refer to all users who've seen the post (including users who've seen it and not taken any action). But in a controllable environment, such as this study, it was correct to assume that the number of visualisations is equal to the number of participants (n = 20). For this reason, the next formula was followed to determine the engagement value of each post:

$$engagement = \frac{(likes + replies + retweets + followings + shares) - blocks)}{20} \times 100\%$$

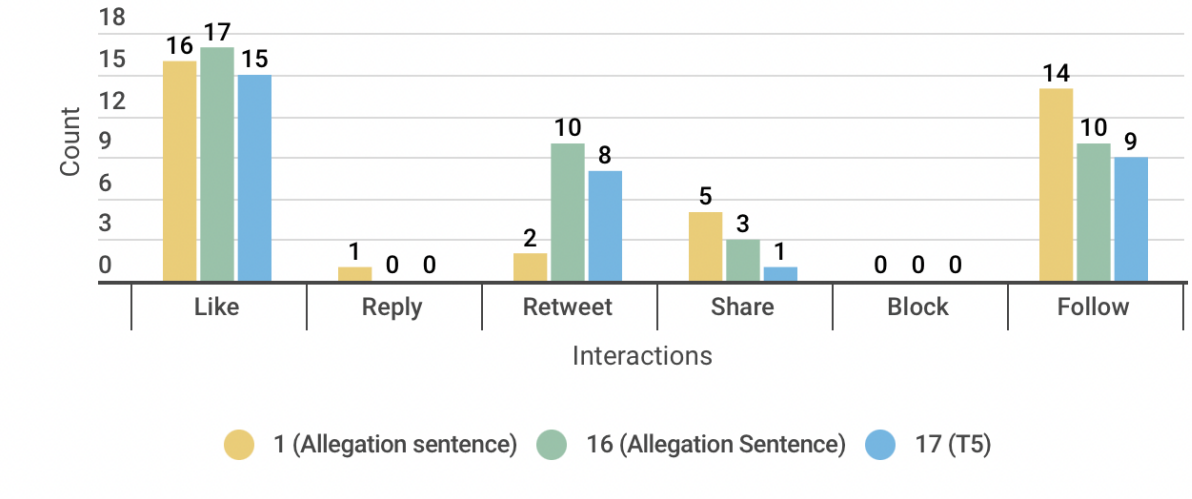
The results of applying the above formula are summarised in Table 16 and proves that *Allegation Sentence* method has higher engagement. In contrast, the engagement score of tweets produced by fact-checking SNOPES is negative.

Approach	Likes	Replies	Retweets	Shares	Blocks	Follows	Engagement
Allegation Sentence	3.8	0.1	1.3	0.4	0	1.43	**35%**
T5	3.33	0	1	0.23	0	1.13	**28%**
Snopes tweets	0.37	0	0.16	0	0.73	0	**-1** %

Table 16: Engagement score for each approach.

As far as engagement is concerned, let's look the three posts with the highest engagement scores. As can be seen in Figure 37, all interactions associated with these posts are positive and the content is relate to the Queen Elizabeth II, who died in the same week that the experiment was conducted.

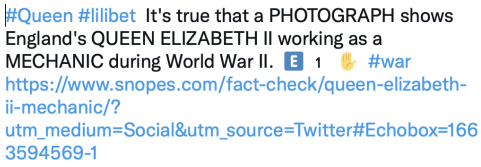
In terms of interaction through posting in the feed, only two users used this option. The content posted by one of them *"Just to test if works"* proves that the intention of the participant was check if all components of the developed prototype environment works. The other post consists of a laughing emoji, with the same finality. Furthermore, when analysing the replies, it was not possible to extract relevant value from the content.



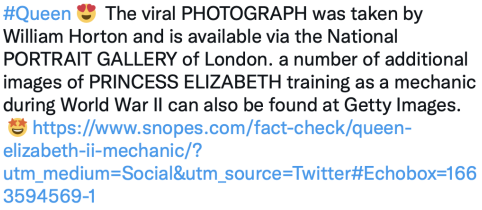
(a) Interactions count for the top 3 posts with more interactions.



(b) Post 1



(c) Post 16



(d) Post 17

Figure 37: Top 3: Posts with more interactions

6.3 Discussion

In order to validate the research hypothesis of this dissertation (presented in section 1.3) and at the same time test the appropriateness and relevance of the developed approaches to produced social media posts, a experiment was conducted with 20 volunteers. Prove this hypothesis is essential to define a pathway that fact-checking agencies or brands can follow to share with success their content on social media. The analysis of the results presented in the previous section revealed findings of interest. These findings establish a link between participants demographic data, posts and the interactions received. The correlation found confirm that there are patterns between the type of posts and the interactions they receive.

The posts created by the Generator have on average 28% (for T5 model) and 35% (in case of the

Allegation Sentence) engagement rates. From the two approaches, it can be concluded that the use of Allegation Sentence is highly accepted by users compared to the others. So, the less technical approach has an advantage over the others, confirmed by the 35% engagement rate. In this experiment, the posts retrieved from the Snopes Twitter feed received negative engagement. With this results the post Generator strategy, presented in Chapter 5, seems to be, at least, more efficient than the strategy followed by the Snopes.

Furthermore, the inclusion of images, especially manipulated ones and with strong content that appeals to the emotional side of the participant, influence the engagement rate, as can be seen in post of Figure 36. In addition, the patterns identified suggest that issues or personalities currently in the media, such as the Queen Elizabeth II (who passed away in the week the study was conducted), are directly related to the increase in interactions number.

Overall, the experiment result is able to validate the research question. This is, the reinforcement of emotions and personality in the post had a positive effect on the participants engagement. Nevertheless, the experiment had some limitations. Most of these limitations are due to the same problem, which is the number of participants and the lack of diversity. As mentioned earlier, the participants come from the same university and many of them are close friends. Furthermore, the mother tongue of the participants is Portuguese. To overcome these limitations, the experiment would need to be conducted with a larger number of participants and in a more diverse environment.

6.4 Conclusion

To prove the research hypothesis, an experiment was conducted with twenty participants, the results are presented and analysed in this chapter. In the first section, the experiment is described, including the planing with all the steps performed by the participants. Then, the results are presented and supported with charts. Finally, in discussion, the results are analysed and limitations described. The experiment validate the research hypothesis, as emotional and personality reinforcement increased engagement rates in posts, on average, 35%.

Chapter 7

Conclusion

This master project in Computer Engineering intend to create a strategy for a Post Generator and contribute to validate the hypothesis "Microblogging posts, built with emotions and personality, are able to improve users engagement". To understand these terms and obtain an algorithm for the Post Generator the DSR research methodology was applied, which states that after identifying a problem, it is necessary to define objectives to reach a solution.

In the Systematic Review it was possible examine the different patterns used by fact-checkers in social media. It turned out that including emotional appeals and personality in tweets can be an efficient way to increase the reach and engagement of readers. Although some studies have been conducted in the area of natural language generation (recognition of language patterns in fact-checking posts and words used), there is a gap in using post generators to create engagement social media posts. What stands out in fake-news are the strong and emotional headlines used to increase consumption, circulation and hit rates on social media. In addition to the spread of fake-news, the systematic review reveals another problem. Social media has been used as a research environment for other scientific studies that require "personal"data or research on humans. Developing research on social media can raise a variety of complicated ethical issues involving data, consent, traceability or even the morality of individuals.

In order to fulfil the goal of this project, an emotional Post Generator with personality reinforcement was created. Several heuristics were developed to obtain the main component of the post (the *Text Core*). The approaches T5 sumarisation model and the Allegation sentence showed good results and were used. Relevant emojis and hashtags are embedded in the post reinforcing the emotional side. For personality, a mental process lexicon was created, based on the theory of Adaptive Personality. The lexicon identify the primary processes: paranoid, schizoid and neurotic and contain the 3432 frequently used words and 3639 avoiding words for each mental process (including noun, verb and adjective). This lexicon, was been used to reinforce the predominant mental trait in the post.

Finally, to validate the engagement achieved through the content created by the Post Generator, a experiment was conducted with twenty participants. This required an immersive platform where participants are in a similar environment to microbblogging platforms without compromising the ethical issues identified through the systematic review. For that, a configurable lab environment inspired by the Twitter layout was developed, with the aim of creating an environment where social media-focused studies can be conducted without compromising ethical values. The experiment results validate the research hypothesis, since the algorithm implemented for the Post Generator increased 35% of engagement score in social media posts with the inclusion of an emotional appeal and personality reinforcement.

7.1 Contributions

The research that was conducted and described in this dissertation produced the following contributions:

- An emotional and personality based Post Generator that creates posts with 35% more engagement compared to current approaches of fact-checking agencies on social media.
- A systematic review on news dissemination behaviour, inspired by the Kitchenham guidelines [15] that allowed identify a set of patterns on fact-checking and fake-news posts. In addition, it showed a research deficit related to fact-checking news sharing behaviour.
- MentaLex, a mental process lexicon which allows identify the primary mental processes: schizoid, neurotic and paranoid.
- A public web application, the Microblogging Environment Simulator, that allows to perform studies in a in-lab-environment having attention to ethical issues that can appear on social media research;

7.2 Publications

The developed work was disclosed by:

- A paper for Practical Applications of Agents, Multi-Agent Systems (PAAMS) 2022 [78], were the Microblogging Environment was discussed as a ethical approach for prototype and demonstrated during the conference. More references in Appendix C.1.
- A paper for the SGAI International Conference 2022 [79], resulting from a collaboration with members of Natural Language process laboratory, of the University Portucalense (Porto) and a psychologist from Digimedia, University of Aveiro. In this paper, MentaLex was submitted. More references in Appendix C.2.
- A paper to be submitted with the Systematic Review. More references in the Appendix C.3.
- A paper, in development, detailing the Post Generator, to be submitted at the IE2023 conference.

7.3 Future work

The research presented in this thesis provides a basic analysis of the high-level approaches that could be used to achieve engagement posts. For future work, several considerations should be made, including:

- **Post Generator**

  Define a more robust strategy for including mental processes in posts. That is, validate the results obtained and ensure cohesion of posts.

- **In-lab-environment**

  In order to make it more immersive, it is recommended to implement some areas of the platform that could be relevant to participant engagement, namely the profile pages of post authors. In addition, other social networks can be explored and modelled by developing an appropriate interface.

- **Psychological validation**

  Emotions and personality are two terms from the psychological sciences. To ensure that no scientific errors have been made, especially with regard to the lexicon developed and explain in Chapter 4, it would be important the work developed be analysed by a research psychology team in the future. With their input, improvements could certainly be made in this study.

- **Experiment**

  Conduct a more in-depth study with the participants. For example, include a questionnaire that allows to detect participants primary mental processes. This could take the study to a point where it would be possible to obtain personalised posts for each individual, according to their mental process.

Bibliography

[1] Q3 2021 Letter to Shareholders, [Accessed 5. Dec. 2021]. [Online]. Available: https://s22.q4cdn.com/826641620/files/doc_financials/2021/q3/Final-Q3'21-Shareholder-letter.pdf (cit. on p. 1).

[2] *Cambridge dictionary*, [Online; accessed 1. Sept 2022], 2022. [Online]. Available: https://dictionary.cambridge.org/pt/dicionario/ingles/post (cit. on p. 1).

[3] S. Lal, L. Tiwari, R. Ranjan, A. Verma, N. Sardana, and R. Mourya, "Analysis and classification of crime tweets", *Procedia Computer Science*, vol. 167, pp. 1911–1919, 2020, International Conference on Computational Intelligence and Data Science, issn: 1877-0509. doi: https://doi.org/10.1016/j.procs.2020.03.211. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920306761 (cit. on p. 1).

[4] N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota, and A. Soomro, "A systematic literature review on text generation using deep neural network models", *IEEE Access*, vol. 10, pp. 53 490–53 503, 2022. doi: 10.1109/ACCESS.2022.3174108 (cit. on p. 1).

[5] C. Ireton and J. Posetti, *JOURNALISM, 'FAKE NEWS' DISINFORMATION United Nations Educational*. Scientific and Cultural Organization, 2018 (cit. on p. 2).

[6] F. S. Marcondes, J. J. Almeida, and P. Novais, "A Short Survey on Chatbot Technology: Failure in Raising the State of the Art", in *Distributed Computing and Artificial Intelligence, 16th International Conference*, Springer, Jun. 2019, pp. 28–36, isbn: 978-3-030-23886-5. doi: 10.1007/978-3-030-23887-2_4 (cit. on p. 2).

[7] A. Hevner, A. R, S. March, *et al.*, "Design science in information systems research", *Management Information Systems Quarterly*, vol. 28, pp. 75–, Mar. 2004 (cit. on p. 3).

[8] K. Peffers, T. Tuunanen, M. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research", *Journal of Management Information Systems*, vol. 24, pp. 45–77, Jan. 2007 (cit. on pp. 3, 4).

[9] N. Ahmad and J. Siddique, "Personality assessment using twitter tweets", *Procedia computer science*, vol. 112, pp. 1964–1973, 2017 (cit. on pp. 5, 33).

[10] J. Pennebaker and L. King, "Linguistic styles: Language use as an individual difference", *Journal of personality and social psychology*, vol. 77, 2000. doi: 10.1037//0022-3514.77.6.1296 (cit. on pp. 5, 32, 33).

[11] P. J. Alroobaea R. Mayhew, "How many participants are really enough for usability studies?", *Science and Information Conference*, 2014. doi: 10.1109/sai.2014.6918171 (cit. on pp. 4, 43).

[12] *DIGITAL 2022: PORTUGAL*, [Online; accessed 1. Sept 2022], 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-portugal (cit. on p. 4).

[13] M. G. H. Allcott, "Social media and fake news in the 2016 election", *Journal of Economic Perspectives*, vol. 31, pp. 211–236, 2017. doi: 10.1080/13803611.2013.805656 (cit. on p. 6).

[14] S. Park, J. Y. Park, H. Chin, J.-h. Kang, and M. Cha, "An Experimental Study to Understand User Experience and Perception Bias Occurred by Fact-checking Messages", in *WWW '21: Proceedings of the Web Conference 2021*, New York, NY, USA: Association for Computing Machinery, Apr. 2021, pp. 2769–2780, isbn: 978-1-45038312-7. doi: 10.1145/3442381.3450121 (cit. on p. 6).

[15] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering", vol. 2, 2007 (cit. on pp. 6, 56).

[16] T. Galvao, T. Andrade Pansani, and D. Harrad, "Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA", *Epidemiol Serv Saude*, vol. 24, Jun. 2015. doi: 10.5123/S1679-49742015000200017 (cit. on p. 6).

[17] M. Gusenbauer and N. R. Haddaway, "Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources", *Res. Synth. Methods*, vol. 11, no. 2, pp. 181–217, Mar. 2020, issn: 1759-2879. doi: 10.1002/jrsm.1378 (cit. on p. 7).

[18] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election", *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017, issn: 0895-3309. doi: 10.1257/jep.31.2.211 (cit. on p. 8).

[19] A. S. Vosoughi S Roy D, "The spread of true and false news online", *Science*, vol. 359, pp. 1146–1151, 2018. doi: 10.1126/science.aap9559 (cit. on pp. 10, 14, 18, 19).

[20] N. Vo and K. Lee, "The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News", in *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 275–284, isbn: 978-1-45035657-2. doi: 10.1145/3209978.3210037 (cit. on pp. 10, 12, 16).

[21] N. Vo and K. Lee, "Learning from Fact-checkers: Analysis and Generation of Fact-checking Language", in *SIGIR'19: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 335–344, isbn: 978-1-45036172-9. doi: 10.1145/3331184.3331248 (cit. on pp. 10, 13, 16).

[22] B. D. Horne and S. Adali, "This Just in: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News", 2017 (cit. on pp. 10, 15, 18, 19).

[23] S. Soroka and S. McAdams, "News, Politics, and Negativity. Political Communication", *Transmission Propensity. Organizational Behavior and Human Decision Processes*, vol. 32, pp. 1–22, 2015. doi: 10.1080/10584609.2014.881942 (cit. on pp. 10, 15, 18).

[24] C. Heath, "Do People Prefer to Pass along Good or Bad News? Valence and Relevance of News as Predictors of Transmission Propensity", *Transmission Propensity. Organizational Behavior and Human Decision Processes*, vol. 68, pp. 79–94, 1996. doi: 10.1006/obhd.1996.0091. (cit. on pp. 10, 15, 18).

[25] B. D. Britt MA Rouet J-F and M. K., "A Reasoned Approach to Dealing With Fake News. ", *Policy Insights from the Behavioral and Brain Sciences*, vol. 6, pp. 94–101, 2019. doi: 10.1177/2372732218814855 (cit. on pp. 10, 15, 19).

[26] I. Singh, A. Barkati, T. Goswamy, and A. Modi, "Adapting a Language Model for Controlled Affective Text Generation", *arXiv*, Nov. 2020. eprint: 2011.04000. [Online]. Available: https://arxiv.org/abs/2011.04000v1 (cit. on pp. 10, 14).

[27] C. Shao, P.-M. Hui, P. Cui, X. Jiang, and Y. Peng, "Tracking and Characterizing the Competition of Fact Checking and Misinformation: Case Studies", *IEEE Access*, vol. PP, no. 99, p. 1, Nov. 2018, issn: 2169-3536. doi: 10.1109/ACCESS.2018.2881037 (cit. on pp. 10, 13).

[28] K. S. Autry and S. E. Duarte, "Correcting the Unknown: Negated Corrections May Increase Belief in Misinformation", *Applied Cognitive Psychology*, vol. 35, no. 3, Apr. 2021, issn: 0888-4080. doi: 10.1002/acp.3823 (cit. on pp. 10, 15).

[29] J. Lee, S. Kanthawala, B. C. Britt, D. F. Deavours, and T. Ott-Fulmore, "Prevalence of anger, engaged in sadness: engagement in misinformation, correction, and emotional tweets during mass shootings", *Online Information Review*, Aug. 2021. doi: 10.1108/OIR-03-2021-0121 (cit. on pp. 11, 13, 17).

[30] Z. Wang, J. Wang, H. Gu, F. Su, and B. Zhuang, "Automatic Conditional Generation of Personalized Social Media Short Texts", *arXiv*, Jun. 2019. doi: 10.1007/978-3-319-97310-4_7. eprint: 1906.09324 (cit. on pp. 11, 13, 17).

[31] F. S. Marcondes, J. J. Almeida, D. Dures, and P. Novais, "Fact-Check Spreading Behavior in Twitter: A Qualitative Profile for False-Claim News", in *Trends and Innovations in Information Systems and Technologies*, Springer, Jun. 2020, pp. 170–180, isbn: 978-3-030-45690-0. doi: 10.1007/978-3-030-45691-7_16 (cit. on pp. 11, 12, 16).

[32] L. Piccolo, A. C. Blackwood, T. Farrell, and M. Mensio, "Agents for Fighting Misinformation Spread on Twitter: Design Challenges", in *CUI '21: CUI 2021 - 3rd Conference on Conversational User Interfaces*, New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 1–7, isbn: 978-1-45038998-3. doi: 10.1145/3469595.3469628 (cit. on pp. 11, 14, 16).

[33] S. Dathathri, A. Madotto, J. Lan, *et al.*, "Plug and Play Language Models: A Simple Approach to Controlled Text Generation", *arXiv*, Dec. 2019. eprint: 1912.02164. [Online]. Available: https://arxiv.org/abs/1912.02164v4 (cit. on p. 14).

[34] *Strategic Influence: Public Diplomacy, Counterpropaganda and Political Warfare*, [Accessed 10. Nov. 2021], Nov. 2021. [Online]. Available: https://www.lulu.com/shop/j-michael-waller/strategic-influence-public-diplomacy-counterpropaganda-and-political-warfare/ebook/product-1vkz7kr8.html?page=1&pageSize=4 (cit. on p. 16).

[35] D. M. J. Lazer, M. A. Baum, Y. Benkler, *et al.*, "The science of fake news", *Science*, Mar. 2018. doi: 10.1126/science.aao2998 (cit. on p. 16).

[36] Q. Zhang, X. Li, Y. Fan, and Y. Du, "An sei3r information propagation control algorithm with structural hole and high influential infected nodes in social networks", *Engineering Applications of Artificial Intelligence*, vol. 108, p. 104 573, 2022, issn: 0952-1976. doi: https://doi.org/10.1016/j.engappai.2021.104573. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197621004061 (cit. on p. 16).

[37] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion", in *WWW '12: Proceedings of the 21st international conference on World Wide Web*, New York, NY, USA: Association for Computing Machinery, Apr. 2012, pp. 519–528, isbn: 978-1-45031229-5. doi: 10.1145/2187836.2187907 (cit. on p. 16).

[38] M. K. F. Mori M. and K. N., "The uncanny valley [from the field]", pp. 98–100, 2012 (cit. on p. 17).

[39] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications", *J. Pers.*, vol. 60, no. 2, pp. 175–215, Jun. 1992, issn: 0022-3506. doi: 10.1111/j.1467-6494.1992.tb00970.x. eprint: 1635039 (cit. on pp. 17, 32–34).

[40] A. Marwick and R. Lewis, "Media Manipulation and Disinformation Online", *New York: Data Society Research Institute*, vol. 359, pp. 1146–1151, 2017 (cit. on p. 18).

[41] *Why We Fall for Fake News*, [Online; accessed 7. Jun. 2022], 2018. [Online]. Available: https://www.cits.ucsb.edu/fake-news/why-we-fall (cit. on p. 19).

[42] M. Henderson, N. F. Johnson, and G. Auld, "Silences of ethical practice: Dilemmas for researchers using social media", *Educational Research and Evaluation*, vol. 19, no. 6, pp. 546–560, 2013. doi: 10.1080/13803611.2013.805656 (cit. on p. 22).

[43] *Social Media and Mental Health*, [Accessed 25. Oct. 2021]. [Online]. Available: https://www.brainforestcenters.com/news/social-media-and-mental-health (cit. on p. 23).

[44] A. K. Przybylski, K. Murayama, C. R. DeHaan, and V. Gladwell, "Motivational, emotional, and behavioral correlates of fear of missing out", *Computers in Human Behavior*, vol. 29, no. 4, pp. 1841–1848, Jul. 2013, issn: 0747-5632. doi: 10.1016/j.chb.2013.02.014 (cit. on p. 23).

[45] Something unexpected went wrong, [Accessed 30. Oct. 2021], 2021. [Online]. Available: https://www.reddit.com/r/deboostedtwitter/comments/jhgym2/something_unexpected_went_wrong (cit. on p. 24).

[46] A. Cockburn, *Writing Effective Use Cases*. Boston, MA, USA: Addison-Wesley, 2001, isbn: 978-0-20170225-5 (cit. on p. 25).

[47] A. R. Feizi Derakhshi, M. R. Feizi Derakhshi, M. Ramezani, *et al.*, "The state-of-the-art in text-based automatic personality prediction", Oct. 2021. doi: 10.48550/arXiv.2110.01186 (cit. on p. 32).

[48] E. Kerz, Y. Qiao, S. Zanwar, and D. Wiechmann, "Pushing on personality detection from verbal behavior: A transformer meets text contours of psycholinguistic features", *arXiv preprint arXiv:2204.04629*, 2022 (cit. on p. 32).

[49] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection", *Artificial Intelligence Review*, 2020. doi: 10.1007/s10462-019-09770-z (cit. on p. 32).

[50] A.-R. Feizi-Derakhshi, M.-R. Feizi-Derakhshi, M. Ramezani, *et al.*, *Text-based automatic personality prediction: A bibliographic review*, 2021. doi: 10.48550/ARXIV.2110.01186. [Online]. Available: https://arxiv.org/abs/2110.01186 (cit. on pp. 32, 33).

[51] X. Sun, B. Liu, J. Cao, J. Luo, and X. Shen, "Who am i? personality detection based on deep learning for texts", in *2018 IEEE international conference on communications (ICC)*, IEEE, 2018, pp. 1–6 (cit. on p. 32).

[52] L. Arambašic, M. Bicanic, and F. Rajic, "Essays are a fickle thing", *Text Analysis and Retrieval 2021 Course Project Reports*, p. 1, (cit. on p. 32).

[53] M. Ramezani, M.-R. Feizi-Derakhshi, and M.-A. Balafar, "Text-based automatic personality prediction using kgrat-net; a knowledge graph attention network classifier", *arXiv preprint arXiv:2205.13780*, 2022 (cit. on p. 32).

[54] Z. Ren, Q. Shen, X. Diao, and H. Xu, "A sentiment-aware deep learning approach for personality detection from text", *Information Processing & Management*, vol. 58, no. 3, p. 102 532, 2021 (cit. on p. 32).

[55] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon", *Computational intelligence*, vol. 29, no. 3, pp. 436–465, 2013 (cit. on p. 32).

[56] S. Mohammad and P. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon", in *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, 2010, pp. 26–34 (cit. on p. 32).

[57] T. Yarkoni, "Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers", *Journal of Reserach in Personality*, 2010. doi: 10.1016/j.jrp.2010.04.001 (cit. on p. 33).

[58] D. Jurafsky and J. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (Prentice Hall series in artificial intelligence). Pearson Prentice Hall, 2009, isbn: 9780131873216. [Online]. Available: https://books.google.pt/books?id=fZmj5UNK8AQC (cit. on p. 33).

[59] V. Joines and I. Stewart, *Personality Adaptations: A New Guide to Human Understanding in Psychotherapy and Counselling*. Lifespace, 2002, isbn: 9781870244015. [Online]. Available: https://books.google.pt/books?id=LJgPAAAACAAJ (cit. on p. 34).

[60] *Counting characters* (cit. on p. 38).

[61] A. Mahajani, V. Pandya, I. Maria, and D. Sharma, "A comprehensive survey on extractive and abstractive techniques for text summarization", in *Ambient Communications and Computer Systems*, Y.-C. Hu, S. Tiwari, K. K. Mishra, and M. C. Trivedi, Eds., Singapore: Springer Singapore, 2019, pp. 339–351, isbn: 978-981-13-5934-7 (cit. on pp. 38, 39).

[62] A. Farzindar, "Social network integration in document summarization", vol. 2, pp. 746–769, Jan. 2014. doi: 10.4018/978-1-4666-6114-1.ch035 (cit. on p. 39).

[63] R. Mihalcea and P. Tarau, "TextRank: Bringing order into text", in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 404–411. [Online]. Available: https://aclanthology.org/W04-3252 (cit. on p. 39).

[64] P. Verma, S. Pal, and H. Om, "A comparative analysis on hindi and english extractive text summarization", *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 18, pp. 1–39, May 2019. doi: 10.1145/3308754 (cit. on p. 39).

[65] *Better language models and their implications* (cit. on p. 41).

[66] M. Lewis, Y. Liu, N. Goyal, *et al.*, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension", Jan. 2020, pp. 7871–7880. doi: 10.18653/v1/2020.acl-main.703 (cit. on p. 41).

[67] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423. [Online]. Available: https://aclanthology.org/N19-1423 (cit. on p. 41).

[68] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding", in Red Hook, NY, USA: Curran Associates Inc., 2019. doi: 10.5555/3454287.3454804 (cit. on p. 41).

[69] C. Raffel, N. Shazeer, A. Roberts, *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer", *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020. [Online]. Available: http://jmlr.org/papers/v21/20-074.html (cit. on p. 41).

[70] R. M. K. Prakriti Garg Ria Gupta, "Rumour/fake-news detection", Jan. 2022 (cit. on p. 42).

[71] *How to use emoji in marketing to drive engagement* (cit. on p. 44).

[72] *95% of internet users have used an emoji. over 10 billion emojis are sent daily* (cit. on p. 44).

[73] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis", *PLoS ONE*, vol. 10, no. 12, e0144296, 2015. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0144296 (cit. on p. 44).

[74] *Using Emojis in Social Media Marketing*, [Online; accessed 31. July 2022], 2022. [Online]. Available: https://www.socialmediatoday.com/marketing/7-tips-using-emojis-social-media-marketing (cit. on p. 44).

[75] *How emojis affect twitter*, [Online; accessed 31. July 2022], 2022. [Online]. Available: https://mention.com/en/reports/twitter/emojis/ (cit. on p. 44).

[76] *EmoTFIDF*, [Online; accessed 1. Sept 2022], 2022. [Online]. Available: https://github.com/mmsa/EmoTFIDF (cit. on p. 46).

[77] D. D. Francisco S. Marcondes and A. Galab, "Principles for a smart tweeting technology: Trl-1 observations on the fact-check spreading", Oct. 2022 (cit. on p. 50).

[78] M. A. Barbosa, F. S. Marcondes, D. Durães, and P. Novais, "Microblogging environment simulator: An ethical approach", *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation*, 2022 (cit. on p. 56).

[79] M. A. Barbosa, F. S. Marcondes, R. Queiroz, L. Brito, A. Gala, and D. Durães, “Mentalex: A mental processes lexicon based on the essay dataset”, *Lecture Notes in Artificial Intelligence*, 2022 (cit. on p. 56).

Appendix A

Detailed posts

Id	Approach	Post
1	Allegation sentence	👉 🎉 ☑️ ✨ #ceremony #coronation aww #official It's dependable that A colored PHOTOGRAPH shows 27-year-old QUEEN ELIZABETH II on the day of her CORONATION in 1953. https://www.snopes.com/fact-check/queen-elizabeth-ii-coronation-day/?utm_medium=Social&utm_source=Twitter#Echobox=1663611185-1
2	T5	#queen #coronation #UK 🫶 💂 After a week of mourning in the united kingdom and beyond, a number of images of the QUEEN went viral. In particular, a PHOTOGRAPH from the day of her CORONATION in 1953. https://www.snopes.com/fact-check/queen-elizabeth-ii-coronation-day/?utm_medium=Social&utm_source=Twitter#Echobox=1663611185-1
3	SNOPES Tweet	✅ True.
4	Allegation sentence	#horse 😣 It's miscaptioned that a picture evidence a HORSE named “Sugar” who is known to lie down and pretend to SLEEP in order to avoid work 🤠 😐 😶 https://www.snopes.com/fact-check/sugar-the-antiwork-horse/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Ec

5	T5	#smartHORSE A picture of SUGAR the HORSE is online since at least 2017, follow seize to a caption claim that the HORSE would rest down and pretend to sleep in order to avoid work. 🧐 🤔 https://www.snopes.com/fact-check/sugar-the-antiwork-horse/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Ec'
6	SNOPES Tweet	The Twitter account for the Jim Rose Circus wrote: “Meet Sugar, she doesn’t like to be ridden. If Sugar is approached with a saddle she lyes down and pretends to be asleep. Sugar refuses to open her eyes until the riders leave.”
7	Allegation sentence	#cobra ✊ bingo ➡️ ☀️ It's true that A VIDEO of a long SNAKE “standing up” and RAISING its head high into the air is real. #venomous 💙 #snake ⚫ #black https://www.snopes.com/fact-check/snake-standing-up-video/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663406617
8	T5	A real VIDEO was shared on social media showing a “snake standing up” and “raising its head” 🐍 🐲 https://www.snopes.com/fact-check/snake-standing-up-video/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663406617'
9	SNOPES Tweet	A video was shared on social media with captions that claimed it showed a very long “snake standing up” and “raising its head.” Readers wanted to know whether it showed a king cobra, black mamba, or other kind of snake. We asked a wildlife expert.
10	Allegation sentence	#Lightyear #ChrisEvans 😲 😾 It's false that DISNEY refused to cast TIM Allen as the voice of BUZZ LIGHTYEAR in the 2022 “Lightyear” movie because of the actor’s conservative beliefs. https://www.snopes.com/fact-check/did-disney-cancel-tim-allen/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663663801
11	T5	There was no prove to suggest that the reason he wasn’t included was his political views. the movie “Lightyear” did not feature the FICTIONAL TOY that ALLEN VOICED in the original “Toy Story” movies. 😬 🧐 https://www.snopes.com/fact-check/did-disney-cancel-tim-allen/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663663801

12	SNOPES Tweet	According to the rumor, Disney refused to cast actor Tim Allen in the 2022 “Lightyear” movie due to his conservative political views. Dan Evon investigates.
13	Allegation sentence	#prank eww It's false that all SIZES of Starbucks' PAPER CUPS hold the same amount of liquid. 😫 💩 https://www.snopes.com/fact-check/starbucks-cup-size-scam/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663379570
14	T5	#fake #coffeeTime A fake video went viral on facebook that supposedly showed a person POURING COFFEE from one CUP into another that appeared larger. The video appeared to show that the same amount of COFFEE filled both CUPS to the brim 🤣 🤔 https://www.snopes.com/fact-check/starbucks-cup-size-scam/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663379570
15	SNOPES Tweet	According to a popular internet rumor, Starbucks is scamming its customers with its trademark paper cup system, because, supposedly, all of its to-go cups hold the same amount of liquid. Dan Evon reports. 👇
16	Allegation sentence	#Queen #lilibet It's true that a PHOTOGRAPH shows England's QUEEN ELIZABETH II working as a MECHANIC during World War II. 🇪 1 ✋ #war https://www.snopes.com/fact-check/queen-elizabeth-ii-mechanic/?utm_medium=Social&utm_source=Twitter#Echobox=1663594569-1
17	T5	#Queen 😍 The viral PHOTOGRAPH was taken by William Horton and is available via the National PORTRAIT GALLERY of London. a number of additional images of PRINCESS ELIZABETH training as a mechanic during World War II can also be found at Getty Images. 🤩 https://www.snopes.com/fact-check/queen-elizabeth-ii-mechanic/?utm_medium=Social&utm_source=Twitter#Echobox=1663594569-1
18	SNOPES Tweet	✅ True. A photograph truly does show Queen Elizabeth II working as a mechanic during World War II.

19	Allegation sentence	It's miscaptioned that A picture shows Air FLORIDA FLIGHT 90 crashing into CARS on a BRIDGE in 1982. #crew eww #truth 💔 #crash #pilot https://www.snopes.com/fact-check/air-florida-airplane-bridge-crash/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663374907
20	T5	#crash #bridge The image that claim show the CRASH of Air Florida Flight 90, was a digital ILLUSTRATION created by a Canadian GRAPHIC designer named Steve McGhee. 😬 😩 https://www.snopes.com/fact-check/air-florida-airplane-bridge-crash/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663374907
21	SNOPES Tweet	With photos like these, a reverse image search can be quite useful. 
22	Allegation sentence	It's true that The SINGER Grimes, whose real name is CLAIRE Boucher, unsuccessfully tried to sail the Mississippi River from Minneapolis to New Orleans in a homemade houseboat. https://www.snopes.com/fact-check/grimes-mississippi-homemade-houseboat/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663639794
23	T5	When CLAIRE BOUCHER was 21, she paired up with 23-year-old William Gratz. the couple built a HOUSEBOAT from scratch in bemidji, mn. they planned to SAIL from BEMIDJI to new orleans, but it ended up being only a few miles downstream https://www.snopes.com/fact-check/grimes-mississippi-homemade-houseboat/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663639794
24	SNOPES Tweet	✅ True. The singer Grimes, whose real name is Claire Boucher, unsuccessfully tried to sail the Mississippi River from Minneapolis to New Orleans in a homemade houseboat.

25	Allegation sentence	#show #TikTok It's false that A TIKTOKER DISCOVERED a hidden purpose for CAR GRAB handles. 🤯 #violation #grab ouch #misleading 😳 💀 😶 https://www.snopes.com/fact-check/car-grab-handle-purpose/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663658696"
26	T5	#show #TikTok Main PURPOSE of CAR GRAB HANDLES is to help people climb into and out of the vehicle. Some drivers might also hang clothes from the GRAB HANDLES if a HOOK is not included with the car. 👌 https://www.snopes.com/fact-check/car-grab-handle-purpose/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663658696
27	SNOPES Tweet	An ad on Yahoo led to a slideshow article that lasted 220 pages. We clicked "next page" 220 times so you don't have to.
28	Allegation sentence	😯 😱 It is false that Amber Heard can be seen sniffing cocaine from a tissue in a video of her testimony during a defamation trial brought by Johnny Depp #cocaine #AmberHeard 👀 https://www.snopes.com/fact-check/amber-heard-sniff-cocaine-trial/?utm_term=Autofeed&utm_medium=Social&utm_source=Twitter#Echobox=1663677823
29	T5	#cocaine #nose #testimony #miscaptioned 🧐 😱 There is no EVIDENCE that AMBER Heard SNIFFED COCAINE in the Trial Pertaining to JOHNNY Depp https://www.snopes.com/fact-check/amber-heard-sniff-cocaine-trial/
30	SNOPES Tweet	Tweets and TikTok posts claimed to show Amber Heard sniffing cocaine as she paused, held a tissue to her nose, and rubbed her nose afterward. We posted the full clip below. 👇

Table 17: Generated posts presented to users.

Appendix B

Experiment results

The experiment presented in Chapter 6 was conducted with twenty participants, from whom interactions with the microblogging simulator (detailed in Chapter 3) were collected. All interactions were structured and included in this appendix. First, Table 18 lists the interactions made by the buttons (like, reply, retweet, share, block, follow). Next, Table 20 lists the posts that the user made during the study. Finally, Table 19 contains the comments related to a particular post.

Post	Approach	Likes	Replies	Retweets	Shares	Blocks	Follows
1	Allegation sentence	16	1	2	5	0	14
2	T5	14	0	1	3	0	10
3	SNOPES	0	0	0	0	0	0
4	Allegation sentence	10	2	4	1	0	3
5	T5	11	0	3	1	0	0
6	SNOPES	0	0	0	0	2	0
7	Allegation sentence	12	0	4	1	0	3
8	T5	8	0	0	0	0	0
9	SNOPES	0	0	0	0	5	0
10	Allegation sentence	13	0	2	1	0	1
11	T5	10	0	2	0	0	1
12	SNOPES	0	0	0	0	2	0

13	Allegation sentence	12	0	5	0	0	0
14	T5	11	1	4	0	0	2
15	SNOPES	1	0	0	0	5	0
16	Allegation sentence	17	0	10	3	0	10
17	T5	15	0	8	1	0	9
18	SNOPES	2	0	2	0	0	1
19	Allegation sentence	8	0	2	0	0	0
20	T5	5	0	3	0	2	0
21	SNOPES	7	0	3	0	0	1
22	Allegation sentence	2	0	0	0	2	0
23	T5	0	0	0	0	3	0
24	SNOPES	0	0	0	0	0	0
25	Allegation sentence	10	0	4	1	0	4
26	T5	11	0	4	2	0	3
27	SNOPES	0	0	0	0	5	0
28	Allegation sentence	14	0	6	1	0	8
29	T5	15	1	5	0	0	9
30	SNOPES	1	0	0	0	3	0

Table 18: Number of interactions with the posts

User	**Post**	**Reply**
User 3	Post 1	“God save the Queen”
User 3	Post 4	“very smart horse”
User 8	Post 4	“oh sometimes I do the same at work ☺️”
User 8	Post 14	“I’ve seen this video before in TikTok. If we look carefully, we can see that the cup is the same”
User 12	Post 29	“ No comments ... 😆 😆”

Table 19: Content related to the five replies made by the users.

User	Post
User 3	“Just to test if it works”
User 12	“😆”

Table 20: Posts that users made throughout the research.

Appendix C

Scientific production

C.1 Microblogging Environment Simulator: an ethical approach

- **Conference:** PAAMS (Practical Applications of Agents and Multi-Agent Systems) 2022
- **Rank:** B (CORE2021)
- **Year of publication:** 2022
- **Status:** Published

Microblogging Environment Simulator: an ethical approach

Maria Araújo Barbosa[0000−0002−0464−0991], Francisco S. Marcondes[0000−0002−2221−2261], Dalila Alves Durães[0000−0002−8313−7023], and Paulo Novais[0000−0002−3549−0754]

ALGORITMI Centre, University of Minho, Braga, Portugal
pg42844@alunos.uminho.pt, francisco.marcondes@algoritmi.uminho.pt, dalila.duraes@algoritmi.uminho.pt, pjon@di.uminho.pt

Abstract. Social media were originally created as a means of communication, but have eventually become an important way of producing and sharing information and news. Beyond that, the scientific community's interest in conducting studies in numerous fields using Twitter is growing daily. This paper presents a platform that simulates a microblogging, inspired by the Twitter layout, with the aim of creating an environment where social media-focused studies can be conducted without compromising ethical values. The end result is a in lab environment with the ability to present the content (in the format of post) to validate or investigate and record user interactions.

Keywords: Microblogging · Simulator · Prototyping · Research environment

C.2 MentaLex: a Mental Processes Lexicon based on the Essay dataset

- **Conference:** SGAI AI-2022
- **Rank:** National UK (CORE2021)
- **Year of publication:** 2022 (estimated)
- **Status:** Approved

MentaLex: a Mental Processes Lexicon based on the Essay dataset

Maria Aráujo Barbosa[1], Francisco S. Marcondes[1], Ricardo Queiroz[2], Luis Brito[2], Adelino Gala[3], and Dalila Durães[1]

[1] Algoritmi Centre, University of Minho, Braga, Portugal
[2] Portucalense University, Porto, Portugal
[3] Digimedia, University of Aveiro, Aveiro, Portugal
pg42844@alunos.uminho.pt, francisco.marcondes@algoritmi.uminho.pt, {44830, 44881}@alunos.upt.pt, adelino@ua.pt, dalila.duraes@algoritmi.uminho.pt

Abstract. Considering that the performance of personality predictors are not consistently increasing throughout the years, as an alternative, the idea was to provide big5 personality traits lexicon based on the Essay dataset. The weight of relevance for every trait of each word is calculated by tf/idf on the words in the Essay dataset. However, it was eventually realized that three personalities wordset were overlapped requiring a change of course. The research changed the underlying model from big5 to Vann Joines' mental processes as it appear to suit better with the empirical findings. The resulting lexicon dataset is composed with 3432 commonly used words and 3639 avoided prone words for each mental process. The commonly used words are capable of covering 81% of the Twitter Personality dataset.

Keywords: Mental Processes · lexicon · personality · NLP

C.3 A systematic review on news spreading strategies

- **Status:** To be submitted

A systematic review on news spreading strategies

Maria Araújo Barbosa[0000−0002−0464−0991], Francisco S. Marcondes[0000−0002−2221−2261], and Paulo Novais[0000−0002−3549−0754]

ALGORITMI Centre, University of Minho, Braga, Portugal
pg42844@alunos.uminho.pt, francisco.marcondes@algoritmi.uminho.pt, pjon@di.uminho.pt

Abstract. Since the US Presidential election in 2016, fake-news has multiplied on the internet and reached millions of people, becoming a persistent phenomenon on social media. In the fight against fake-news, numerous fact-checking systems have been developed to inform people about the truth. However, there does not appear to be a dissemination strategy for fact-checks, or if there is, it is poorly structured and ineffectual. This paper presents a systematic review, on fact-checks and fake-news spreading strategies to achieve user engagement, inspired on the Kitchenham guidelines [1]. A total of 14 articles, selected from Scopus and the ACM-DL, were included in the study. That low amount of research papers focusing in fact-check spreading strategies strengths the claim that this an underdeveloped matter. Fact-checking organisations appears to be unconcerned about developing tactics for effectively creating an disseminating fact-checking posts. And, the approaches used by fake-news to reach a larger audience, on the other hand, are usually not publicised. Based on these results, it was propose a new taxonomy of aspects to be considered related to the news spreading behavior. The taxonomy includes a variety of different facets, such as content, dissemination and engagement.

Keywords: Systematic Review · PRISMA · Social Media · Fake-news · Fact-checking · Engagement · Post generation