Social Network Analysis and Mining (2021) 11:36
https://doi.org/10.1007/s13278-021-00746-y

REVIEW PAPER

Misleading information in Spanish: a survey

Eliana Providel[1,2] · **Marcelo Mendoza**[2]

Received: 12 December 2020 / Revised: 18 March 2021 / Accepted: 27 March 2021 / Published online: 9 April 2021

Abstract
Misleading information spread on social networks is often supported by activists who promote this type of information and bots that amplify their visibility. The need for useful and timely mechanisms of credibility assessment in social media has become increasingly indispensable. Efforts to tackle this problem in Spanish are growing. The last years have witnessed many efforts to develop methods to detect fake news, rumors, stances, and bots on the Spanish social web. This work leads to a systematic review of the literature that relates the efforts to develop this area in the Spanish language. The work identifies pending tasks for this community and challenges that require coordination among the leading investigators on the subject.

Keywords Misleading information · Rumor verification · Stance classification · Bot detection · Fake news

1 Introduction

The rise of social networks has led millions of users to read and comment on the news acquired in these media. Given the speed of information dissemination that characterizes social networks and the lack of timely credibility assessment mechanisms, the proliferation of untrue information has been exacerbated (Vosoughi et al. 2018; Zhang and Ghorbani 2020). Many times, this information is deliberately propagated to manipulate public opinion. The scope of this phenomenon is still studied.

Several researchers have tackled the task of detecting misleading information, avoiding the harmful effects of this on social media users (Mendoza et al. 2010; Castillo et al. 2011; Zubiaga et al. 2016). These methods are based on machine learning algorithms, which, based on a labeled data set, adjust a predictive model's parameters. The most frequently used data sources in this task consider the text of the original post and their comments and data from the users' profiles involved in the spread of the information (Mendoza et al. 2010). Other methods incorporate temporal characteristics of the propagation phenomenon or structural characteristics of either the social network or the network of interactions between users (Zubiaga et al. 2018). Some of these methods combine various sources of information, using complex neural network architectures (Ma et al. 2020).

The design of misleading information detection methods in Spanish has shown a growing interest from the community during the last years. A main challenge for the development of this topic is probably because the community of researchers is smaller than the English-speaking community. Other factors that may explain how difficult is to research in this area can be attributed to the lack of NLP resources available to tackle these tasks. Despite these factors, research on Spanish misleading information detection has shown interest for the past five years. This work summarizes the efforts to develop effective methods to tackle this challenge on the Spanish-speaking social web.

The evidence shown in this work will allow us to conclude that we are being witnesses of many relevant efforts in the study of misleading information in Spanish. However, there are many challenges and pending tasks that must be addressed by researchers and stakeholders in the upcoming years. These tasks will require coordination, both for the development of language resources that support the implementation of these methods in Spanish and for the consolidation of benchmark data that evaluates and compares the community's different methods.

The contributions of this work are:

✉ Marcelo Mendoza
marcelo.mendoza@usm.cl

Eliana Providel
eliana.providel@uv.cl

[1] School of Informatics Engineering, Universidad de Valparaíso, Valparaiso, Chile

[2] Department of Informatics, Universidad Técnica Federico Santa María, Santiago, Chile

- We conducted a systematic review of the literature in misleading information in Spanish. To the best of our knowledge, this is the first work that addresses this task systematically in our language.
- We identify three areas of work in this community, with more significant development in one than others. These areas are fake news and rumor verification, bot detection, and stance classification. While bot detection and stance classification show some recent progress, the fake news and rumor verification show a high interest in the last year. Unfortunately, works that holistically address the phenomenon, incorporating all these variants of the problem in a coherent framework, are practically nonexistent.
- We identify pending tasks and challenges that must be urgently addressed by the community. Among them, the need to promote the use of NLP resources in Spanish and the consolidation of benchmark data to evaluate these methods' performance in our native language stand out.

The work is organized in the following sections. Section 2 introduces the review methodology. In Sect. 3, we present the results of the systematic review of the literature, organizing the papers according to the tasks through which they address the problem of misleading information in Spanish, detailing the methods, tools, and datasets used for empirical validation. In Sect. 4, we discuss the results found, comparing their findings and detecting gaps in the literature. Several challenges emerge from this section that the community must urgently face in the upcoming years. A short work agenda summarizes the main tasks that must be addressed to produce further progress in the subject. Finally, we conclude in Sect. 5, providing concluding remarks.

2 Review methodology

2.1 Subject of the study

Misleading information detection is a highly active research subject (Zubiaga et al. 2018). The detection of misleading information concerns the identification of information created to manipulate or deceive people. Without losing too much generality, many of the last years' efforts focus on the effects of misleading information in social media. For the study, we bound the scope of this phenomenon to this ecosystem.

The study of misleading information encompasses several concepts that are often used interchangeably but have different meanings, such as fake news, rumors, or hoaxes, among others (Zhang and Ghorbani 2020). Clear definitions help to understand the scope and implications that each of these concepts has. We distinguish between these concepts, providing definitions with wide acceptance in the community:

- *Misinformation* We adopt a definition provided by Swire-Thompson and Lazer (2020). According to these authors, misinformation is *any item of information whose content contradicts the epistemic consensus reached by the systematic application of a methodology*. Consensus can occur through different methods. By applying the scientific method, a scientific agreement is reached. An investigative methodology in a judicial process drives towards a legal consensus. While all consensus are transitory given that are subject to the community's investigation, misinformation contradicts the available evidence. Examples of misinformation are the statements made by groups such as the anti-vaccines (Germani and Biller-Adorno 2020), who refuse the scientific evidence.
- *Disinformation* Disinformation is a specific type of misinformation whose purpose is to manipulate and mislead public opinion (Zhou and Zafarani 2020). Disinformation includes all those forms of information dissemination whose objective is to return power or money to its author. Disinformation is materialized, for example, in deceptive news and hoaxes (a plan to deceive someone in specific).
- *Fake news* fake news are false news published in the news media (Zhou and Zafarani 2020). The definition includes all types of news that contain one or more false facts or statements. Fake news includes deceptive news, that is, news designed to deceive and manipulate, and also unintentional false news, that is, the publication of unverified false information. Both types of news have different causes. While deceptive news are aimed at manipulating public opinion, unintentional fake news are related to the need to capture the audience using unreliable sources.
- *Rumor* Zubiaga et al. (2016) defines a rumor as an *item of information whose veracity status is yet to be verified at the time of posting*. This widely accepted definition generally points to posts spread in social networks from no authoritative sources, as social media users or untrusted news outlets. Accordingly, while fake news originates from traditional media and can eventually spread on a social network, rumors generally originate within a social network. There are many types of rumors, the most commonly used categories being the following: true rumor (a rumor that was confirmed), false rumor (a rumor that was denied with factual evidence), and unverified rumor (a rumor that has not been verified or denied yet). These types of rumors have been operationalized in automatic detection tasks, and there are widely used datasets in the area that use this categorization (Ma et al. 2020).
- *Stance* stance is the posture of a person towards a given target or claim. Stances toward targets (e.g., legal abor-

tion) comprise for and against postures. There are more types of stances for claims, being the most widely used supporting, questioning, denying, and commenting. Stance and false rumors are related concepts, as social media users tend to doubt or deny false rumors, while confirmed facts are commonly supported (Mendoza et al. 2010).

- *Bot* a bot is a software program that handles a social network account to perform repetitive actions such as repeated postings or reposts of other user messages (Cresci 2020). Bots, a.k.a. bot spammers, have the objective of amplifying the impact of specific content, increasing its visibility. While the first generations of bots focused on spamming, the last generations imitate social behavior, mimicking humans within networks and being very difficult to detect (Mendoza et al. 2020). These bots, known as social bots, evolve, incorporating sophisticated rules and behavior patterns. The relationship between bots and political campaigns has been in the spotlight in recent years (Kollanyi et al. 2016). Their ability to help spread false information is one of the main threats on social media.

The concepts defined above are strongly interconnected. On social media, rumors and fake news can be amplified by users whose belief system is confirmed by information. This phenomenon, known as confirmation bias, strengthens the spread of information. The participation of users in its propagation can be deliberate or involuntary. Detecting the stance of users concerning information can reveal the type of information that is spread.

The visibility of some rumors and fake news can be amplified by bots (Varol et al. 2017). The relationship between rumor detection and bots has been studied, showing that false information often gains greater visibility with support for bot account activity (Mendoza et al. 2020). Bot detection is an important task that can help in misleading information detection. Accordingly, the misleading information analysis has many faces. The most important are rumors and fake news and their initial effect on the network, bots and their amplifying effect on misinformation, and users' stance in front of this information (Cücük and Can 2020). We highlight the relationship between these three faces of misleading information in this survey, understanding that they address views of the same phenomenon from complementary perspectives, as we show in Fig. 1.

This study's focus addresses the three faces of misleading information mentioned above as variants of the same problem. The use of machine learning methods to address these variants of the problem is the main subject of this study. Accordingly, we review different approaches that have addressed these tasks in Spanish. We show that both rumor and fake news detection, bot detection, and stance classification are related tasks. This study will gather information about the study of these tasks in Spanish, either using methods based on this language or working on data collected on the Spanish-speaking social web.

Fig. 1 The three faces of misleading information in social media. Malicious users fed fake news and rumors into social media to manipulate public opinion or generate revenue on their sites (e.g., clickbait). Bots amplify the effect of fake news and rumors, increasing its reachability, and accelerating its spread. Finally, users, in this dizzying scenario, adopt a stance in front of the information. Many of them voluntarily (e.g., hyper-partisans) or involuntarily spread misleading information on social networks

2.2 Overall view of the study

We develop our study by conducting a systematic literature review (SLR). The purpose of an SLR is to provide explicit search criteria and the definition of criteria for exclusion and inclusion of papers so that the study results can be reproduced and updated. The criteria used to identify relevant works must be clear and explicit so that all the works included in the study address the focused problem. To organize the study, we follow the guidelines of Brereton et al. (2007), who have systematically studied the literature review methodologies. In this paper, we expand these guidelines by including a final search stage in which we check papers that cite works detected using the SLR search criteria. The same exclusion and inclusion criteria are applied to these works to ensure that all the works fit the studied problem. The general framework of the process is in the diagram in Fig. 2.

The diagram in Fig. 2 shows that the first task of the SLR is to plan the review. In this planning, the search terms and the inclusion/exclusion criteria are defined. The review continues with searching the related literature (Fig. 2-(1)), applying the search criteria based on keywords. We use Scopus as a service for literature search (Fig. 2-(2)) because Scopus offers extensive coverage of sources, including journal articles and conferences. Scopus records citations and articles that cite found works. These data will be of vital importance for the review process in its later stages. Once the search process for related literature has finished, the first body of literature is available (Fig. 2-(3)), which is thoroughly reviewed in a literature screening process guided by

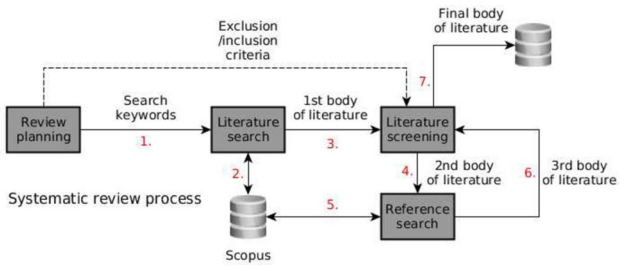

Fig. 2 General process of the SLR used in this study. First, we apply search strings to identify an initial document body. These documents are examined using inclusion/exclusion criteria. A second stage of the SLR consists of analyzing the works that cite the papers detected in the first phase. After applying the inclusion/exclusion criteria, the references are reviewed, looking for more relevant documents. Finally, the documents are consolidated into a single body of literature

the exclusion/inclusion criteria. The second body of literature is obtained (Fig. 2-(4)) on which we use Scopus incorporating the works that cite the works included in the first body of literature (Fig. 2-(5)). The third body of literature is obtained (Fig. 2-(6)), which is subjected to the literature screening process, applying the same inclusion/exclusion criteria and reviewing the references of the selected articles. Once this process has finished, the final body of literature is obtained (Figs. 2-(7)).

Further details of the review planning, literature search and screening processes are provided in "Appendix".

3 Results

We organize the results found in the SLR using different dimensions of analysis. First, we pay attention to the representations used to characterize the events. Then, we focus on the type of learning algorithms used to solve the task, particularly the classification between machine learning and deep learning-based methods. Another relevant dimension to organize the body of literature relies on the datasets used to validate each proposal. We pay special attention to the methods used to collect and label the data. Within each dimension of analysis, we distinguish between the three tasks that are subject to this study.

3.1 Representations

We distinguish different types of approaches depending on the representations used to characterize the events. Methods based on the content of the posts are named text-based approaches. This kind of approach pays attention to the text and stylistic features. User-based approaches make use of features extracted from user profiles. Features like account age, followers, and followees are essential for these approaches. Finally, hybrid approaches make use of both representations to address a task.

3.1.1 Text-based representations

Fake news and rumor detection

The number of papers in fake news and rumor detection in Spanish has increased in the last years. A recent competition at PAN 2020 (Pardo et al. 2020) generated much attention from the community. Besides, other initiatives such as MEX-A3T at IberLEF 2020 (Aragón et al. 2020) have helped increase interest in this topic in Spanish. We start this section summarizing efforts in fake news and rumor detection to later present progress on stance and bot detection tasks.

In Posadas-Durán et al. (2019), the detection of fake news in a Spanish corpus was addressed using text-based features extracted from the news headlines. The authors studied the usefulness of bag-of-words (BOW) and part-of-speech (POS) representations, the latter obtained using a Spanish model provided in the Spacy library. The authors evaluated n-grams versions of these representations. Lexical features were used in Boididou et al. (2018) to distinguish between credible and misleading tweets in Spanish. Tweets in this work include text and images. The authors used the text of the tweet to infer the veracity of the post. The approach considers aggregated features at the text level, such as the number of positive and negative words, number of exclamation marks, and emoticons. The authors also incorporated features extracted from the comments, such as the number of tweets that include hashtags or mentions. Abonizio et al. (2020) evaluate textual features that are not tied to a specific language for detecting fake news. The authors studied the problem in three languages: American English, Brazilian Portuguese, and Spanish. The Spanish partition used for the study corresponds to the Fake News Spanish Corpus. Several stylometric and psychological text features were used to support the detection of fake, legitimate, and satirical news. Miranda et al. (2020) analyzed the headlines and downloads of a relevant newsgroup in Costa Rica, classifying them as true or false news according to the source that disseminates them. The news was represented according to its content using lexical and psycholinguistic characteristics.

The author profiling task at PAN 2020 (Pardo et al. 2020) proposed the fake news spreaders detection task, distinguishing between fake promoters and trusted users using tweet track records. The challenge provides two data partitions, one in English and one in Spanish. Many of the submissions to this challenge used representations based on text features. Vogel and Meghana (2020) used term frequency-inverted document frequency (TF-IDF) char n-grams and char n-gram counts. Ikae and Savoy (2020) uses word selection based on relative differences in TF rankings between

both classes. Majumder and Das (2020) used the Universal Sentence Encoder Multilingual (Yang et al. 2020), an extension of the Universal Sentence Encoder created for English (Cer et al. 2018). The encoder allowed them to code both partitions of the dataset to tackle the task. Giglou et al. (2020) uses ConceptNet Numberbatch (Speer et al. 2016), a hybrid word embedding built using an ensemble approach. It combines data from ConceptNet (Liu and Singh 2004), Word2Vec (Mikolov et al. 2013), GloVe (Penningtonet al. 2014), and OpenSubtitles 2016 (Tiedemann 2012) using a variation of retrofitting. ConceptNet Numberbatch is a multilingual word embedding, where words in different languages share a common semantic space, and all languages inform that semantic space. They concatenate the average feature vectors computed using ConceptNet Numberbatch with Latent Semantic Analysis (LSA) vectors computed from a matrix of token counts related to n-grams. Shashirekha and Balouchzahi (2020) use ULMFiT (Universal Language Model Fine-Tuning) (Howard and Ruder 2018) trained on the Spanish Wikipedia and fine-tuned in the PAN dataset. Lichouri et al. (2020) combine three TF-IDF features (word 5-grams, char 5-grams, and char with boundary 5-grams) to create a feature vector per user. Shashirekha et al. (2020) combine unigram TF-IDF, n-gram TF, and Doc2vec trained on the PAN 2020 fake news spreaders dataset. Then, they apply feature selection on unigram TF-IDF and n-gram TF. Fernández and Ramírez (2020) use bigrams and trigrams considering the top 1000 features, ordered by frequency from the whole corpus. Then, they compute one text vector per profile, comprising the whole tweet record of each user. Koloski et al. (2020) use char n-grams and word n-grams with low dimensionality reduction through singular value decomposition (SVD) term-document matrix decomposition. Pinnaparaju et al. (2020) use TF-IDF to create a word vector for each user. Bakhteev et al. (2020) use FastText (Bojanowski et al. 2017) at tweet level to train a recurrent neural network (RNN). Espinosa et al. (2020a) use word n-grams and char n-grams to compute a feature vector at the user-level. López and Martí (2020) use the multilingual extension of the universal sentence encoder to compute text embeddings at the tweet level. These embeddings are fed into many dense layers to obtain a final representation of each user record. Manna et al. (2020) combine stylometric features, categories of emojis, and lexical features related to news headlines to create a unique feature vector per user. Pizarro (2020) combines char n-grams and word n-grams to create a text-based feature per user. A similar approach was followed by Buda and Bolonyai (2020), who only use word n-grams. Das et al. (2020) use ELECTRA-based embeddings (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) (Clark et al. 2020), a variant of Bidirectional Encoder Representations based on the Transformer (BERT) (Devlin et al. 2019) that is pretrained using a replaced token detection task. This training strategy allows them to work with longer sequences of text. They sample each tweet collection, building 15 different base models, which are processed by a dense layer to fed a softmax classifier. Shrestha et al. (2020) use many features based on writing style, word n-grams, char n-grams, and BERT semantic embeddings. Each of these feature types is used to train a single classifier. Labadie et al. (2020) use char, and word level encoders trained on-the-fly and fused to fed a classifier. Also, the representation is combined with stylistic features. Vogel and Meghana (2020) extended their work submitted to the fake news spreaders task at PAN 2020, studying characteristics and tendencies in the dataset used during the competition. The study shows that the task was challenging in two ways. Many tweets mixed false with factual information. Also, tweets were inherently noisy, short, and incorporate many platform-specific features and many spelling mistakes and grammatical errors. Accordingly, word-level-based approaches perform worse than character n-gram approaches.

Fake news analysis in Mexican Spanish was addressed in MEX-A3T (Aragón et al. 2020) at IberLEF 2020. The Fake News Detection Track consists of classifying a given set of news written in Mexican Spanish between true and fake. Villatoro-Tello et al. (2020) used three different sets of features for this task, namely word n-grams, char n-grams, and BERT for Spanish (BETO). These vector features were concatenated and fed into a supervised autoencoder (SAE) to obtain a low-dimensional vector with 300 components. Zaizar-Gutiérrez (2020) use TF-IDF term weighting to create a vector representation of each news. A similar approach for text representation was adopted by Arce-Cardenas et al. (2020), who also use TF-IDF term weighting to represent the news.

Stance classification Stance classification was one of the tasks considered in the Ibereval contest (Taulé et al. 2018). The challenge releases two datasets related to the 2017 Catalan Referendum, considering both Catalan and Spanish tweets. This challenge recorded many papers with different approaches. A group of papers addressed the task using word embeddings. For example, Vinayakumar et al. (2017) used word embeddings to represent the tweets. Word embeddings were defined as learnable parameters related to the stance classification task. The study shows that the approach is promising. González et al. (2017) also explored the use of word embeddings to tackle the task. However, the authors decided to use embeddings pretrained over the Spanish version of Wikipedia using word2vec (Mikolov et al. 2013). The work showed that the use of pretrained embeddings decreased the effectiveness of the stance classifier. Barbieri (2017) also evaluated the performance of word embeddings in this task. Vectors pretrained using FastText (Bojanowski et al. 2017) on a 5 million tweets corpus geo-localized in

Spain were used. The reported results show better performance in the dataset in Spanish than in Catalan. FastText was also used by Wojatzki and Zesch (2017), who showed that stance classifiers trained on these representations obtained slight improvements compared to word n-grams and char n-grams. Ambrosini and Nicolo also used FastText to solve the task obtaining better results in Spanish than in Catalan (Ambrosini and Nicolò 2017).

Another widely used representation in IberEval was based on word n-grams or char n-grams. For instance, Swami et al. (2017) evaluated char n-grams and word n-grams showing that both representations got the same performance in Spanish, but char n-grams are slightly better in Catalan tweets. An approach based only on word n-grams showed that stance detection performed better when using Spanish and Catalan tweets than when using only Catalan (Taulé et al. 2018). Segura-Bedmar (Taulé et al. 2018) evaluated the effectiveness of word n-grams using TF-IDF vectors. Although the results obtained in the validation partition were satisfactory, the results in the testing partition were much lower, evidencing the presence of overfitting in the training of these classifiers. TF-IDF vectors constructed on word uni-grams representations were used to address the stance detection task (García and Larriba Flor 2017). The authors claim that combining Catalan and Spanish tweets in a single dataset allows obtaining a classifier with better results than those obtained when training separating both languages. The winning study of the stance competition also followed a text-based approach (Lai et al. 2017). That work defined several characteristics, among them stylistic characteristics (e.g., bag-of-words, bag-of-lemmas, POS and char n-grams), structural characteristics of the text (e.g., frequency of hashtags, amount of words that starts with capital letters), and context characteristics of the tweet (e.g., words included in anchor and outlinks). A more sophisticated representation approach was adopted in Taulé et al. (2018), where word embeddings and lexicons were combined. The authors adopted a word2vec representation trained on the task data. Various lexicons were used to detect affective words, including ELHPolar (Saralegi and Vicente 2013), ISOL (Molina-González et al. 2013), MLSenticon (Cruz et al. 2014), and a Spanish version of NCR (Mohammad and Turney 2013). The study shows that the combination of word2vec and lexicon-based word n-grams offers better results in the validation partition of the task that the one obtained using char n-grams or word n-grams.

Bot detection

Author profiling for bot detection was proposed as a challenge in PAN at CLEF 2019 (Rangel and Rosso 2019). The task considers two folds of Twitter accounts, one in Spanish and the other in English. The Spanish partition records 2400 bot accounts and 2400 human accounts, with splits of 1500/900 tweets per class. For each account, there is a history of 100 tweets per author. The competition received several papers with different approaches. Many papers explored the use of word n-grams and char n-grams representations. Srinivasarao and Manu (2019) addressed this task using char n-grams and word n-grams. The authors represented each user with a sequence of tweets. The authors concluded that word unigram and bigram perform well when a sequence of tweets in Spanish is analyzed. A similar approach was studied by Jimenez-Villar et al. (2019), who showed competitive results in this task by combining word n-grams and 3–5 character n-grams. Espinosa et al. (2019) studied character n-grams to solve the task. The authors showed that the results obtained using this approach are similar in both Spanish and English tweets. Unigrams and char n-grams were also evaluated in this task, showing good results in the validation fold but lower results in the testing partition, evidencing overfitting during classifier training (Van Halteren 2019). Other proposals also used char n-grams and word n-grams in this task (Pizarro 2019; Vogel and Jiang 2019; Fagni and Tesconi 2019), adding other characteristics such as the polarity of each tweet and the pointwise mutual information (PMI) of terms (Giachanou and Ghanem 2019) or POS n-grams (Gishamer 2019).

Fernquist (Fernquist 2019) defined ad hoc characteristics for author profiling tasks in PAN at CLEF 2019, such as the monotonicity and lexical diversity of the posts. The objective of these characteristics is to capture repetitive lexical patterns that show the use of algorithms to generate text. These characteristics' regularity was measured using text compression ratios, showing that bots tend to produce low entropy texts. Bacciu et al. (2019) also used lexical-based features to solve the bot detection task in PAN 2019. In addition to stylometric features, the authors evaluated the usefulness of a text masking process, which obfuscates some words from the text to increase the focus of the method on the unmasked words. The study shows that the characteristics extracted using masking text preprocessing have advantages over more straightforward stylometric characteristics such as the number of positive and negative words. In the same competition, Ashraf et al. (2019) define several stylometric characteristics to tackle the task. The characteristics include, but are not limited to, URL counts, hashtags, emojis, capitals, and other special symbols. The authors claim that the set of features is language-independent, favoring its use in other domains. Przybyła (2019) also used an extensive list of stylometric features to address the task, showing better results in predicting bots in English than in Spanish tweets. Stylometric and n-grams were combined by Valencia et al. (2019) to solve the same task. The authors showed that the combination of both approaches outperformed n-grams. Goubin et al. (2019) defined many stylometric features that were used to solve this task. Among other features used for this competition, the authors introduced distributional features,

such as word entropy and POS tagging distributions. The authors showed that their classifier's performance obtained competitive results in both English and Spanish tweets. In addition to using stylometric characteristics, Gamallo and Almatarneh (2019) used lexicons consisting of specific words belonging to the language of bots and humans. These lexicons were built on-the-fly using the PAN 2019 dataset, identifying frequent words in each of the classes defined for the task. The authors report better results in detecting bots in Spanish than in English tweets. Stylometric features were also used in other studies of this contest (Johansson 2019; HaCohen-Kerner et al. 2019).

Some papers submitted to the author profiling task in PAN at CLEF 2019 used word embeddings. For instance, in Onose et al. (2019) word embeddings pretrained on a corpus with around 1.5 billion words created from multiple Spanish web resources were used to solve the task. The authors showed that a deep learning model trained on these vectors was outperformed by simpler models, such as char n-grams or word n-grams. The authors attribute this result to the little thematic coverage offered by the Spanish corpus. Petrik and Chuda (2019) also used word embeddings to represent each sequence of tweets in this contest. However, they evaluated a variant where the embeddings were trained on-the-fly using word2vec on the same dataset. The authors showed that this approach performs better in English than in Spanish tweets. Halvani and Marquardt proposed a similar approach (Halvani and Marquardt 2019). They also confirmed that when training word embeddings on-the-fly, better results were obtained in detecting bots in English than in Spanish. Word embeddings pretrained using FastText (Bojanowski et al. 2017) in Spanish were used in Polignano et al. (2019). The paper claims that the results obtained using FastText outperforms word n-grams, char n-grams, and word2vec. Word2vec and text-based vectorizations as TF-IDF were combined to address the bot detection task (López-Santillán et al. 2019). The authors decided to combine several vector representations of text using genetic programming to obtain a global combination at the user's level. The authors report better results when analyzing tweets in English than in Spanish.

The SLR detected two papers for stance classification in Spanish not related to the PAN competition that uses text-based representations to address the task. In a paper authored by Lai et al. (2020), the authors explore a large number of lexical features to address stance detection. The authors distinguish between four types of characteristics. These are (1) stylistic features including word n-grams, char n-grams, POS tagging-based features, and lemma-based features, (2) stylometric features, (3) affective features, based on both sentiment lexicons [e.g., Affective Norms for English Words (AFINN)] or emotions [e.g., Linguistic Inquiry and Word Count (LIWC)], and (4) domain knowledge-based characteristics. The latter is of particular relevance for stance detection since the authors include the descriptors' characterization of the target topic on which a post expresses a stance. The authors evaluate these characteristics in different settings, which allows them to obtain several relevant conclusions. They stand out that stylistic features perform well in supervised learning environments, while affective and stylometric features perform better in semi-supervised learning. The study also highlights that affective characteristics work better when the target of the stance points to a celebrity. At the same time, stylometric features offer better results when the target is a campaign or institution. Graells-Garrido et al. (2020) study how demographic groups are represented in the discussion about abortion legislation in Argentina and Chile. When analyzing users' stances related to this discussion, the study shows that the stance can vary dramatically over time, even in users with strong opinions about an issue. Text-based representations (word n-grams) were used to address this task.

In summary, in this section, we showed that many of the papers detected by our SLR that use text-based representations are related to three competitions: IberEval 2017, PAN 2020, and PAN 2019. Many papers have focused on fake news spreaders detection and author profiling using text-based representations in Spanish. This fact highlights the importance of these competencies in the investigation of this problem.

3.1.2 User-based representations

Our SLR detected two papers based on users' representations to address tasks related to the subject of this study, one related to claim verification and the other to bot detection. User-based features are used in Boididou et al. (2018) to distinguish between credible and misleading tweets in Spanish. Tweets in this work include text and images provided in a dataset named Verification Corpus. The authors make use of user-based features to infer the level of veracity of a tweet. Features considered in this work include the number of followers and followees, account age, and many binary features retrieved from the user profile as if the account is verified or includes biography. Al-Zoubi et al. (2018) define several user-based features to train a Twitter bot account classifier. The approach combines characteristics extracted from the user's profile (e.g., number of followers/followees, account age, the existence of actual image in the profile, living place) and characteristics of user behavior. Experiments on a multilingual dataset that includes Spanish accounts show that some of the most relevant features for this task are the number of followers and the characteristics derived from the user profile's completeness.

3.1.3 Hybrid representations

Fake news and rumor detection The author profiling task at PAN 2020 (Pardo et al. 2020) that considers a challenge for fake news spreaders detection in Spanish and English, also received submissions that used hybrid representations of their data. Agirrezabal (2020) used each user's complete tweet track to build a feature vector computed with the average word embedding from all words used. Word embeddings were computed on-the-fly. This vector was combined with features vectors based on user, stylometric, and POS tags features. Cardaioli et al. (2020) combined stylometric text features as diversity, readability, and five personality features: agreeableness, conscientiousness, extraversion, emotional range, and openness. Espinosa et al. (2020b) combined psychological features based on emotions and write style, linguistic features (e.g., POS and Named Entity Recognition (NER)), user profile features (e.g., mentions, RTs, and hashtags), and headline text features. Hashemi et al. (2020) combined word-based TF-IDF and word embeddings pretrained on the Spanish Wikipedia with user profile features: agreeableness, conscientiousness, extraversion, emotional range, and openness. Bello et al. (2020) combined user behavioral features as URL, hashtag, retweet, and user counts; lexical features as word and char counts, POS tags, NER tags, and psycholinguistic features based on lexicons as LIWC (Chung and Pennebaker 2012). Russo (2020) combined stylometric and user behavioral features as the number of RTs, mentions, replies, URLs, hashtags, and emoticons combined with emotion words.

Bot detection Hybrid representations combine features extracted from texts and users' profiles. Papers that use services based on hybrid features, such as Botometer, also fall into this category. Botometer (Davis et al. 2016) is a classifier of Twitter accounts based on many features. Initially trained in English texts, our SLR has found many papers that use Botometer to study bots' presence in Spanish-speaking countries. Our SLR also detected papers based on other hybrid representations that address some of the tasks related to this survey subject.

Botometer (Davis et al. 2016) contains a classifier that uses various data sources, combining user, content, and network-based data. Although Botometer was initially designed for tweets in English, it has been used on foreign language networks. Botometer was used to detect bots on the Honduran Twitter network during the 2017 presidential election (Gallagher et al. 2019). The authors show bots' presence during the campaign, with more than one hundred separate communities of social bots being identified. Botometer was also used to detect bots in a trending topic in Mexico (Suárez-Serrato et al. 2018). The study shows a significant presence of bots on the network. The types of features used to analyze the detected bots include lexical and network features, determining that some bots do not have an adversarial behavior, suggesting the need to define a class of benign bots. Botometer was also used to detect bots in four trending topics related to natural disasters in Mexico and USA (Khaund et al. 2018). The study performs a structural analysis of the bot and human networks involved in these trending topics and leads to a lexical-based analysis of the tweets. The study concludes that while humans have more focused interests in specific communities, bots follow accounts without a clear sense of community belonging. Another work that has used Botometer to detect bots in Spanish-speaking networks is Castillo et al. (2019). The work shows a significant presence of bots during the Chilean 2017 presidential elections. The authors compare Botometer results with other classifiers trained by them using lexical and user-based features without detecting significant performance differences. Botometer was also used to detect bots in Guatemala (Richards et al. 2019), showing a high presence of automated accounts. The authors started with a small number of bot accounts, from which they build a network of followers and followees to carry out their analysis. After retrieving the Botometer scores for all the accounts, the authors applied kernel density estimation on the score distributions, finding evidence of bimodality. The authors take advantage of these features to tag bots in the highest bot-account density mode. Pastor-Galindo et al. (2020a) studied the detection of bots in the context of the 2019 Spanish general election. The authors gathered more than 5.8 million tweets related to the election authored by almost eight hundred thousand users involved in political discussions. Tweets and user profiles were tagged using automatic tools, computing topics, and keywords-based features for the tweets. Botometer-based features were also retrieved for this task. The study shows how to use the data to reveal trending topics and the intervention of bots detected using Botometer in these topics. The study reveals that bots are used during political debates to amplify the social impact of social media candidates.

Other hybrid-based representations not based on Botometer are also devoted to author profiling tasks. Van Halteren (Volkova and Bell 2017) follows a hybrid approach for predicting deleted accounts in many languages, including Spanish. The study evaluates user, network, and lexical-based features to characterize the behavior of several Twitter accounts. The study shows that a combination of these features is helpful in the prediction of deleted accounts. The authors conclude that the analysis can help credibility assessment since many deleted accounts are used to amplify misleading information. Our SLR detected three papers in PAN at CLEF 2019 based on hybrid representations. User and text-based approaches are combined in Oliveira et al. (2019). The authors showed that although the combination of all the characteristics allows obtaining better results in

PAN 2019, the user-based features achieved a very competitive performance. Other work (Bounaama and Abderrahim 2019) in the same contest pay attention to stylometric features, combining them with user-based features. The study claims that bot detection using this hybrid approach gets similar results in English and Spanish tweets. A similar approach was proposed by Bolonyai et al. (2019), who combined stylometric and user-based features to solve the task in PAN 2019. The paper shows that the combination of features allows better results when detecting bots than when using only lexical features.

3.1.4 Synthesis of the section

This section shows that different representations have been explored in any of the three tasks related to this survey. We summarize them in Fig. 3. The bars show the number of works reviewed in this section according to the type of representation and task. By far, word embeddings, text stylography and word n-grams are the most used representations, followed by char n-grams and BOW. Most of the papers that use these representations are related to fake news spreaders, bot detection or stance classification due to the IberEval 2017 and PAN 2019/2020 competitions. This section also shows several case studies based on Botometer. Features based on users' profiles or user behavior are less explored. The lack of work related to graph-based representations is surprising. One reason that explains the lack of works based on graphs is the lack of datasets that considers graphs of social connections or interactions between users. The difficulties of acquiring this type of data are probably related to the challenges that current platforms have to access this type of data, either due to data volume constraints or privacy issues. Regarding word embeddings, papers based on word-2vec, GloVe (Global Vector for word representation), FastText, and the original version of BERT are detected. While FastText and BERT are used because there are pretrained versions in Spanish, other embedding-based approaches are less explored, revealing the lack of Spanish resources in representation learning.

3.2 Learning algorithms

We distinguish different types of approaches depending on the learning algorithms used to address the tasks. We separate the algorithms between classical machine learning algorithms that depend on handcrafted features or learn directly from the input data and deep learning algorithms. Classical algorithms as support vector machines (SVM), random forests (RF), and logistic regression (LR) are dominant in these tasks, while convolutional neural networks (CNN) and long short-term memory (LSTM) architectures are explored when deep learning is chosen. Approaches based on attention are less explored, and graph-based architectures are unexplored in Spanish for these tasks.

3.2.1 Classical algorithms

Fake news and rumor detection Many works of the SLR are devoted to fake news and rumor detection in Spanish using classical machine learning algorithms. Posadas-Durán et al. (2019) introduce a dataset for the detection of fake news in

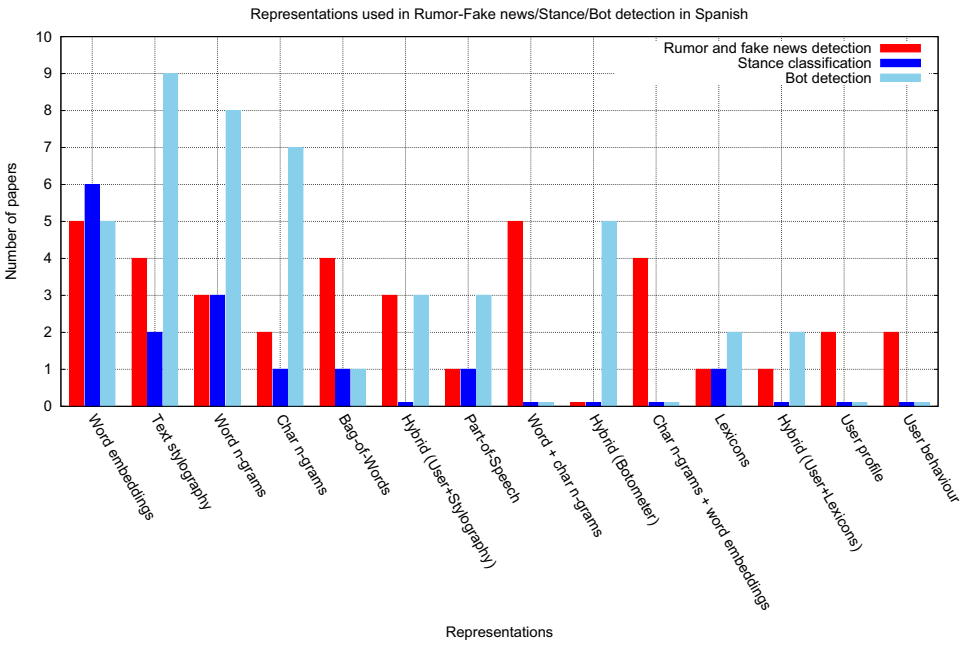

Fig. 3 Representations used in rumor-fake news/stance/bot detection in Spanish and number of papers in which they were used. From left to right, the representations are shown in decreasing order of occurrence. The colors indicate the presence of the representation in each of the tasks examined in this work

Spanish. To distinguish between true and false news, the authors study the performance of four classical classifiers. These are SVM with linear kernel, LR, RF, and adaptive boosting (ADABOOST). These classifiers' performance depends on the type of representation used, obtaining better results SVM and ADABOOST when using char n-grams and RF when including POS n-grams. Tweets with false and real information were used to train classifiers of misleading content (Boididou et al. 2018). The dataset used by the authors includes the text and some descriptive images of the content. Two algorithms were studied to distinguish between true and false tweets; these are LR and RF. Experimental results showed that although both classifiers obtained comparable performances, LR had some advantages over RF in this specific task. Abonizio et al. (2020) evaluate textual features that are not tied to a specific language for detecting false news in the Fake News Spanish Corpus. Classical machine learning algorithms were used to address the task with an average detection accuracy of 85.3% using RF. Miranda et al. (2020) analyzed the headlines and downloads of a relevant newsgroup in Costa Rica, classifying them as true or false news according to the source that disseminates them. After training a set of classical binary classifiers such as RF, LR, and SVM on text-based features, they obtained similar results for these methods, with performance around 63% in precision. Zaizar-Gutiérrez (2020) evaluated many classical algorithms in the fake news analysis in Mexican Spanish task addressed in MEX-A3T (Aragón et al. 2020), based on TF-IDF text representations, achieving the best results using an SVM, with 77.9% in F1 score. Arce-Cardenas et al. (2020) also evaluated several classical machine learning algorithms based on TF-IDF text representations, achieving the best results using an feed-forward neural networks (FFNN), with 76.1% in F1 score.

Many of the submissions of the fake news spreaders task at PAN 2020 used classical machine learning techniques to build their classifiers. While LR was chosen by Vogel and Meghana (2020) (79% in testing accuracy), Pinnaparaju et al. (2020) (70%), Manna et al. (2020) (72.5%), and Koloski et al. (2020) (79%), RF was used by Cardaioli et al. (2020) (74%), Agirrezabal (2020) (72%), and Espinosa et al. (2020b) (64%). Boosted aggregation (BAGGING) also attracted the attention of several researchers in this endeavor. Ikae and Savoy (2020) built several base classifiers and then tackled the task using majority voting, reaching 72% on testing accuracy. Shashirekha and Balouchzahi (2020) constructed two base classifiers using linear SVM on unigram TF-IDF and n-grams, respectively. A third base classifier was trained using LR at tweet-level doc2vec representations. The task was also approached using majority voting, reaching 67.5% in testing accuracy. Buda and Bolonyai (2020) used an ensemble of classical classifiers (LR, SVM, RF, and gradient boosting (XGBOOST)), processed to obtain meta-level features fed into an LR, reaching one of the competition's highest results, with 80.5% on testing accuracy. Shrestha et al. (2020) also used BAGGING, training base classifiers on different types of features. The task was approached using majority voting, reaching 76% accuracy on testing. Other classical algorithms were also tested in this task. Bello et al. (2020) used XGBOOST, achieving a 72% testing accuracy. Russo (2020) used RF regressors (regression as classification) for the task with only 51.5% accuracy testing. Neural networks were also studied in this task. Giglou et al. (2020) used FFNN for the task, achieving 74% in testing accuracy. López and Martí (2020) used Universal Sentence Encodings (Cer et al. 2018) at the tweet level to encode each user's tweet tracks. The encodings were fed into an FFNN with four layers to get an encoding processed by a fully dense layer to address the task, reaching 75% accuracy in testing. Of the classical algorithms, the ones with the best performance were those based on SVMs. Pizarro (2020) used SVM reaching 82% in testing, the best result of the competition. Another very competitive result was obtained by Espinosa et al. (2020a), who, with a linear SVM, reached 81.8% in testing accuracy. Other submissions also used SVMs. While linear SVM was chosen by Lichouri et al. (2020) (76%), linear SVM was used by Fernández and Ramírez (2020) (73%), and Hashemi et al. (2020) (78.5%).

Stance classification Stance classification records many approaches in our SLR that make use of classical machine learning algorithms. SVMs with radial basis functions were used to address the stance classification task in IberEval 2017 (Swami et al. 2017). The study shows the performance of SVMs on text-based representations. Experimental results illustrate that SVMs solve the task better in Catalan than in Spanish. Text-based features were used to train different classifiers on the stance detection task in IberEval 2017 (Lai et al. 2017). The study shows the performance achieved using SVM, RF, LR, and multinomial Naive Bayes (MNB). Furthermore, the authors show a committee machine's performance, applying majority voting on the previous four classifiers' outputs. Experimental results show that the best performance is achieved using SVM, with a slight advantage in the Catalan dataset over the Spanish dataset. Lai et al. (2020) introduce multiTACOS, a multilingual system for stance detection. The proposed system considers many features of different types to characterize tweets' contents. Among the most relevant characteristics defined in the system are lexical, syntactic, and stylographic features. Some features that reveal cross-lingual patterns would allow the system to be more robust to local linguistic variations. The system uses classical text classification methods to detect stances such as SVM or LR. The authors report results in various stance datasets, highlighting the result obtained in the IberEval 2017 contest in which they report the best performance. Almendros and Cervantes (Taulé et al. 2018) use

text-based features to detect stance in the MultiStanceCat task at IberEval 2018. The challenge consists of detecting tweets related to the 10 October Catalan Referendum, including tweets in Catalan and Spanish. The authors train a linear SVM to address the task. Experimental results show performance in testing instances with F_1 around 30% for Catalan and 27% for Spanish tweets. Segura-Bedmar (Taulé et al. 2018) uses BOW TF-IDF vectors to represent the tweets of the IberEval 2018 contest. Using these vectors, the author trains many classical machine learning classifiers, showing that the best results are achieved using RF. Experimental results show that the classifier reaches F_1 around 28% for testing instances both in the Spanish and Catalan folds.

Bot detection The task that most classical methods of machine learning records in our SLR is bot profiling. This is due to many works based on these techniques received in the PAN at CLEF 2019 competition. The prevalence of SVMs among these techniques is remarkable. For instance, an SVM was trained on char and word n-grams representations addressing this challenge (Srinivasarao and Manu 2019). The authors showed that the SVM outperformed classifiers based on MNB and LR. The study also shows that the SVM performs better in detecting bots in English than in Spanish accounts. A linear SVM was trained on word and char n-grams representations to solve the same task (Giachanou and Ghanem 2019). The study shows that the SVM outperforms other classical algorithms as LR and RF. The study also shows that the detection of bots in Spanish is a bit more difficult than in English. Bacciu et al. (2019) and HaCohen-Kerner et al. (2019) also used char n-grams to train an SVM, with promising results in this task. An SVM-based ensemble was studied by Gishamer (2019). The proposal creates two SVM classifiers, one based on word n-grams and the other on POS tags. Both classifiers are assembled using an SVM-based meta-learner to make the final prediction. The study shows that bot detection in English is more straightforward than in Spanish, showing that the ensemble outperforms each base classifier. An SVM was trained on features related to the account's behavior, sentiment, and variety of posts (Oliveira et al. 2019). The study reports better results in detecting bots in English than in Spanish. An SVM classifier was trained on the 3-grams to solve the same task (Bounaama and Abderrahim 2019). The study reports poor results in detecting bots in both Spanish and English. The poor performance can be attributed to the fact that the 3-grams are incapable of generating a representation with sufficient descriptive capacity to solve the task. In the same competition, a linear SVM classifier was trained to detect bots (Vogel and Jiang 2019). The classifier was trained on TF-IDF vector representations of word unigrams and bigrams as also char n-grams. The study reports promising results in detecting bots both in Spanish and English. Another study (Jimenez-Villar et al. 2019) evaluates text representations based on word n-grams and char n-grams, looking for the optimal feature selection technique to solve the task using SVMs. The authors evaluate three feature selection strategies, document frequency selection (DF), frequently co-occurring entropy (FCE), and information gain (IG). Experimental results show that DF works well in detecting bots in English, but IG works better in Spanish. The best results in this competition were obtained using SVMs. For example, Pizarro (Pizarro 2019) used char and word n-grams to train an SVM in this task. The proposal uses a TF-IDF representation of the tweets. The study shows outstanding results, both detecting bots in Spanish and English. The competition reported that this work obtained the best performance, with an $F_1 = 93.33\%$ in detecting bots in Spanish. Espinosa et al. (2019) also used SVMs on char bi-grams in the same competition. The authors studied other classical machine learning methods, such as J48, NB, and RF. The results obtained by the SVMs outperformed the other methods, obtaining accuracies around 92% in bot detection tasks in both English and Spanish.

The PAN at CLEF 2019 competition also received some papers that included a wide spectrum of algorithms' in this task. For instance, many learning algorithms were studied in Al-Zoubi et al. (2018). The focus of the paper is in the selection of the most useful characteristics to solve this task. The authors evaluate three selection strategies, whale optimization (WO), genetic algorithms (GA), and particle swarm optimization (PSO). Experimental results show that WO outperforms the other techniques in this specific task. Another work with a wide spectrum of methods is that of Ashraf et al. (2019) showed that RF outperforms other classical algorithms when using stylometric features to solve the task. Among the algorithms included in the study are MNB, LR, SVM, and linear-SVM, being the last two variants of SVMs. The study shows that RF performs better at detecting bots in English than in Spanish. LR was also a prevalent approach used in this competition. LR was used in Valencia et al. (2019) using representations based on tweet content's statistical features. The study reports good results detecting bots both in Spanish and English. Several stylometric and user-based features were used to create an aggregated representation of each tweets' track addressing the bot detection task in the same competition (Bolonyai et al. 2019). The proposal shows that a classifier based on LR achieves better results in detecting bots in English than in Spanish. Przybyła (2019) models the message sequence of each user in the bot profiling task. The proposal is based on stylographic handcrafted features that are used to measure the predictability of the messages generated by each user. The author measures distributional features of all the messages, computing mean, and standard deviation of the model's outcomes. These features are then fed into an LR classifier, whose outcome is

a bot profiling confidence score. The proposal gets better results in detecting bots in English than in Spanish.

RF was less explored in this task. Goubin et al. (2019) used features based on text to train an RF bot classifier. The features include word entropy, the ratio of tweets that contain emojis, and part-of-speech (POS) distribution. The study reports results based on more classifiers, as bagging and an SVM that works on TF-IDF text representations. The best results in bot detection were achieved using RF, showing critical performance differences between training and testing partitions. Johansson (2019) uses text and stylographic features to train an RF classifier. The classifier gets better results detecting bots in English than in Spanish.

Finally, FFNN, decision trees, and MNB were also less explored in this competition. An FFNN was used in Halvani and Marquardt (2019). For each account, the architecture was fed with a matrix input formed by the word embeddings of each of the words used in the tweets' track. The embeddings were trained from scratch using 200 dimensions. A global max pooling operator gets a 1D vector representation of each account, feeding it into the FFNN network. The output is implemented with a softmax layer. The study reports better results in detecting bots in English than in Spanish. Many features were used to train CatBoost, a gradient boosting algorithm based on decision trees (Fernquist 2019). Among the features used to train CatBoost, the study highlights the use of Tf-Idf vectors of char and word n-grams and some aggregated features at the user-level. Experimental results show that CatBoost outperforms other ensemble algorithms as RF, showing promising results in detecting bots in English and Spanish. Gamallo and Almatarneh (2019) use text and lexical-based features to train a MNB classifier. The model performs well in the bot profiling task, achieving better accuracy in Spanish (88%) than in English (81%).

3.2.2 Deep learning-based algorithms

Fake news and rumor detection The fake news spreaders task at PAN 2020 received some submissions based on deep learning architectures. Majumder and Das (2020) trained an LSTM using sequences of tweets to represent each author's tweet track, achieving 72% testing accuracy. Shashirekha and Balouchzahi (2020) used ULMFiT (Howard and Ruder 2018) to address the task, reaching 64% in testing accuracy. Bakhteev et al. (2020) fed a sequence of tweets into an RNN, averaging their hidden state vectors to represent each tweet track. These encodings were fed into an FFNN to address the task, achieving 78% testing accuracy. Das et al. (2020) used 15 ELECTRA-based text models (Clark et al. 2020) to fed a fully dense layer, combining these representations. The encodings obtained were used to fed a softmax layer, achieving 69% of testing accuracy. Labadie fed word, and char-based embeddings trained on-the-fly to feed an LSTM with attention, achieving 72% accuracy. In the fake news analysis in Mexican Spanish task addressed in MEX-A3T (Aragón et al. 2020), Villatoro-Tello et al. (2020) used a fully dense layer to classify the news encoded using a supervised autoencoder (SAE). These encodings were used to address the classification task, achieving 85.6% in F1 score in the testing partition. This result was the best performance achieved in the contest.

Stance classification For stance classification, our SLR found more works in this line related to the IberEval 2017 competition. In Vinayakumar et al. (2017), an LSTM and a gated recurrent unit (GRU) networks are used to address this task. The study shows that LSTM networks offer slightly better performance than GRUs when detecting stance in Catalan, while GRU networks outperform LSTM when solving the same task in Spanish. Convolutional neural networks (CNN) and LSTM networks were combined to solve the stance detection task in the same competition (González et al. 2017). The hybrid architecture was trained using one-hot term encodings. The results show that this approach outperforms other variants of the same architecture trained on word2vec embeddings. Taulé et al. (2018) use word2vec to address the MultiStanceCat task at IberEval 2018. The authors used word embeddings pretrained on a large-scale Spanish corpus of tweets to train a CNN. The classifier was also fed with a lexicon-based feature vector that allows the representation to include affective features. Although the model performed well in the development partition, the results in the testing instances were disappointing.

Bot detection The bot detection task in PAN at CLEF 2019 received more works based on deep learning than the other tasks surveyed in this study. Hierarchical attention networks (HAN) were used in this competition in Onose et al. (2019). The authors created a representation based on word-2vec for the tweets of each account. The word encodings for the Spanish partition of the dataset were trained using around 1.5 billion words created from multiple Spanish web resources. HAN networks were defined using two layers based on words and sentences. Both layers were implemented using bidirectional gated recurrent units (Bi-GRU). The study reports better results in English than in Spanish accounts. A hybrid architecture that combines two CNNs and a bidirectional LSTM (Bi-LSTM) was trained by Petrik and Chuda (2019). The model was trained on word n-grams from a sequence of words defined by a time distributed input sequence. The competition reports better results in detecting bots in English than in Spanish. In the same competition, a 2D-CNN was evaluated to solve the bot detection task (Polignano et al. 2019). The authors proposed to represent the track of tweets of each account using word embeddings. Each tweet was truncated to the first 50 words, achieving a matrix with 50 columns whose entries correspond to

word embeddings. A sequence of 2D convolutional filters was applied, alternating them with max-pooling operators. Finally, a flatten layer was applied to get a vector representation of each account, feeding four fully dense layers trained to solve the task. The study tests three word embeddings, word2vec trained on Google News, GloVe (Penningtonet al. 2014), and FastText (Bojanowski et al. 2017). Only FastText was used to solve the task in Spanish. Experimental results show that FastText helps detect bots in both English and Spanish. Fagni and Tesconi (2019) combined two text embedding architectures to learn a joint representation of each tweet, modeling lexical and syntactical text features. The authors use a word2vec word embedding architecture pretrained on the Spanish Billion Word Corpus. A second text encoding architecture mixes a gated recurrent unit (GRU) and a CNN neural network to create another text embedding. The second architecture's motivation is to model syntactical features, as the combination of GRUs and CNNs may capture sequence embeddings. Both embeddings are fed into an SVM to address the bot profiling task at CLEF 2019. Experimental results show good performance in both English and Spanish folds.

Finally, only one work in bot detection in Spanish was found in our SLR without a relationship with the PAN at CLEF 2019 competition. In this work, LSTM networks were used to build models distinguishing between deleted, suspended, and non-deleted accounts on Twitter (Volkova and Bell 2017). LSTM networks were trained in a supervised way, using sequences of tweets from each account. The study shows that this model outperforms classical classifiers based on SVM and RF, highlighting the need to model the tweet sequence to solve this task adequately.

3.2.3 Synthesis of the section

This section shows that classical machine learning techniques are predominant in the tasks analyzed by this survey. We summarize them in Fig. 4. Among them, the use of SVMs is widely adopted by the community as other methods such as RF and LR. These tasks are less explored for deep learning algorithms, despite the fact that many recent papers use this type of architectures. Among these, the LSTM and CNN networks have a higher adoption by the community. Attention-based approaches are less explored, and work based on GNNs was not detected. The use of graph-based representation learning in Spanish is unexplored according to our SLR.

3.3 Datasets and linguistic resources

3.3.1 Linguistic resources in Spanish

We highlight some linguistic resources that can or should be useful for fake news and rumor detection, stance classification, and bot detection in Spanish. Representation learning in Spanish is a key building block of many of the deep learning architectures proposed in the literature. Some methods have already used these resources. Other resources should be relevant in the upcoming years.

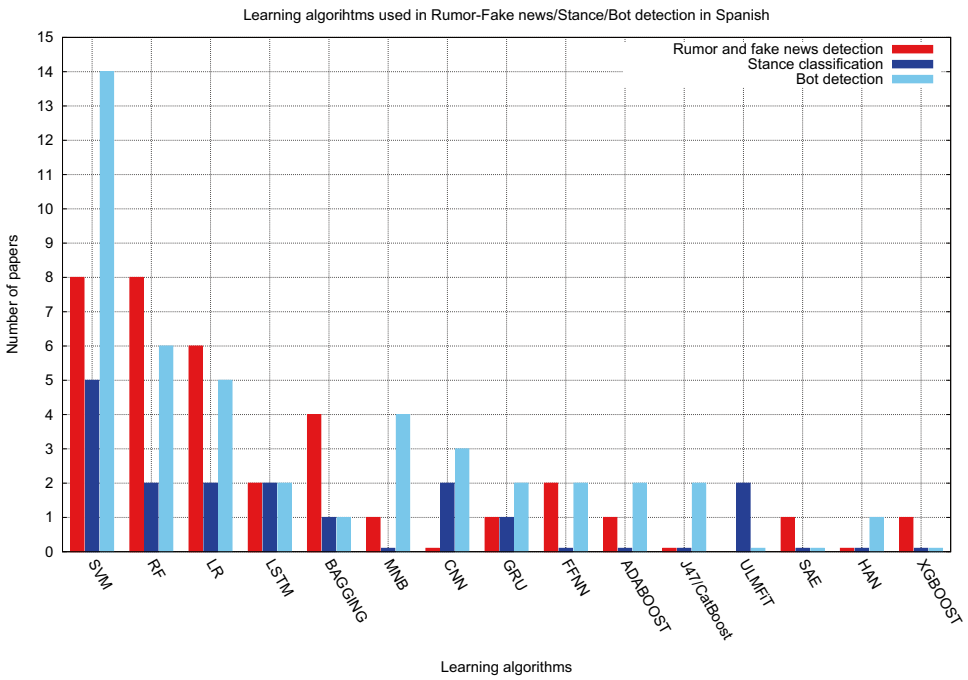

Fig. 4 Learning algorithms used in rumor-fake news/stance/bot detection in Spanish and number of papers in which they were used. From left to right, the algorithms are shown in decreasing order of occurrence. The colors indicate the use of the algorithm in each of the tasks examined in this work

- *Spanish Billion Word Corpus and Embeddings (SBWCE)* This resource consists of a corpus of the Spanish language of nearly 1.5 billion words, collected from different corpora gathered from the web. It also includes a set of word embeddings created from this corpus using the word2vec algorithm. These vectors have been used by some of the methods outlined in this survey when using deep learning approaches (Fagni and Tesconi 2019; González et al. 2017). The corpus is available at https://crscardellino.github.io/SBWCE/.
- *Spanish word Embeddings* This resource leverages many Spanish word embeddings trained on large corpora. The resources include embeddings trained using FastText (Bojanowski et al. 2017) and GloVe (Penningtonet al. 2014). The embeddings are available at https://github.com/dccuchile/spanish-word-embeddings.
- *BERT for Spanish* This resource release contextual embeddings for Spanish based on BERT (Devlin et al. 2019). The resource, named BETO, was trained on a big Spanish corpus that includes the SBWCE corpus. BETO embeddings are available at the Huggingface Transformers library. The project is available at https://github.com/dccuchile/beto.
- *SpanBERTa* This resource releases a RoBERTa (Liu et al. 2019) language model for Spanish, namely SpanBERTa. SpanBERTa has the same size as RoBERTa-base. The authors followed RoBERTa's training schema to train the model. SpanBERTa embeddings are available at the Huggingface Transformers library. The project is available at: https://skimai.com/roberta-language-model-for-spanish/.

3.3.2 Datasets for fake news and rumor detection in Spanish

Even though several papers have addressed this task, many of them use datasets that are not public, making it difficult to reproduce their results. We identified four datasets for rumor verification and/or fake news detection in Spanish that are available on the web.

- *The Spanish fake news corpus* This dataset contains a collection of news gathered from various web repositories, including newspapers, digital media, news verification sites, and webpages that publish fake news. It has 971 news stories, of which 491 are true, and 480 are false. The dataset is divided into 9 topics: Science, Sports, Economy, Education, Entertainment, Politics, Health, Security, and Society. The dataset describes the news item, including the title, the subhead, and the URL in which the original publication can be accessed. It is available at: https://github.com/jpposadas/FakeNewsCorpusSpanish.
- *Verification corpus* This dataset contains true and false tweets initially collected for the MediaEval 2015 Verifying Multimedia Use (VMU) task. The dataset contains tweets with references to images, of which 193 correspond to real images and 218 to false images. It also has two examples of misused videos. The dataset comprehends 6225 true tweets and 9404 false tweets posted by 5895 and 9025 users. Both the images and the related tweets are associated with events, some of them real and other hoaxes. There are two versions of the dataset, one from 2015 and the other from 2016, both with training and testing partitions. The dataset provides the tweets and the images, all of them with veracity labels and the event they are associated with. It is available at: https://github.com/MKLab-ITI/image-verification-corpus.
- *FTR-18* This dataset comprehends a multilingual rumor dataset on football transfer news. The dataset was collected during the 2018 Summer Transfer Window. The FTR-18 dataset comprises 3,045 news articles and more than 2,064K tweets. The news considers transfer news covered by online sports media written in English (1517), Spanish (747), or Portuguese (781) by 96 different news organizations. This collection includes 304 transfer moves associated with 175 target football players. The authors annotated the rumor veracity by adding the evidence supporting or refuting the rumor. The dataset contains tweets and retweets written in English (1130K), Spanish (677K), or Portuguese (257K) that are related to the news. The dataset is available at: https://github.com/dcaled/FTR-18.
- *PAN-AP 2020* Fake news spreaders detection at PAN 2020 was proposed as a binary classification task (Pardo et al. 2020). The task organizers released PAN-AP-2020, a dataset that comprises revisions from fact-checking websites such as Politifact or Snopes. The collection of tweets related to these news was classified as supporting or not fake news. Users that share at least one fake news were labeled as spreaders. The dataset comprises 500 tweet track records at the user-level, with 100 tweets per user-profile, with training and testing partitions considering 300 and 200 instances. Data access can be required at: https://zenodo.org/record/4039435#.YFDCl_5-E5Q.

3.3.3 Datasets for stance classification in Spanish

We detected three datasets for stance classification in Spanish. This observation reveals that this task appears as less explored than rumor verification. There are some efforts that have extended these datasets, incorporating other Latin languages such as Italian and French (Lai et al. 2017).

- *TW-10* This dataset, namely TW-10, was released at IberEval 2017 to address the stance classification task

towards the Catalan Independence. It includes tweets of social and political debates labeled in favor, against, and neutral. The data was collected during the year 2015. The dataset's training partition has 5400 tweets in Spanish, with 335 in favor, 1446 against, and 2538 neutral. The testing partition has 84 tweets in favor, 361 against, and 636 neutral. It is available at: http://stel.ub.edu/StanceIberEval2017/data.html.

- *MultiStanceCat* This dataset was built for the MultiStanceCat task released at IberEval 2018. The dataset contains 5545 tweets with stance labels in three classes, favor, against, and neutral, all of them regarding the stance of the tweet concerning the Catalan Referendum. This dataset's training partition records 1680, 1785, and 972 favor, against, and neutral tweets. The testing partition records 419, 446, and 243 favor, against, and neutral tweets. The tweets were collected from September 20, 2017, to September 29, 2017. The task and the dataset was published at: https://sites.google.com/view/ibereval-2018.
- *The Catalonia Independence Corpus (CIC)* Zotova et al. (2020) introduces a multilingual dataset for stance detection in Twitter for the Catalan and Spanish languages, intending to facilitate research on stance detection in multilingual and cross-lingual settings. The dataset is annotated with a stance towards the independence of Catalonia. The dataset is class balanced. The dataset comprises tweets from TW-10. It includes more tweets, including a collection of tweets from 12 days during February and March of 2019 posted in Barcelona and during September 2018 posted in Terrassa, Catalonia. The dataset comprises 680000 tweets in Catalan and 2 million tweets in Spanish. The annotation process was based on stance classification at the user level. In total, 25,510 users were manually categorized, reading their top salient tweets and their profiles. LDA was used to cluster more tweets related to underrepresented classes. Finally, the dataset has around 10000 tweets per language, with 4105 against, 4014 in favor, and 1868 neutral tweets. The corpus can be accessed at: https://github.com/ixa-ehu/catalonia-independence-corpus.

3.3.4 Datasets for bot detection in Spanish

Even though several works surveyed in this work are devoted to bot detection, they collect Twitter accounts without publishing these data. The only two datasets available for reproducible research dedicated to this topic are:

- *PAN AP 2019* The dataset was released for the Author Profiling Task of PAN (Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection) at CLEF 2019. This dataset collected a set of Twitter user accounts labeled as human or bot. For each account, there is a history of 100 tweets per author. The task considers two folds, one in Spanish and the other in English. The Spanish partition records 2400 bot accounts and 2400 human accounts, with splits of 1500/900 tweets per class for training and testing folds. The dataset is available at https://pan.webis.de/CLEF19/pan19-web/author-profiling.html.
- *2019 Spanish general election* Pastor-Galindo et al. (2020b) present a Twitter dataset collected from October 4th to November 11th, 2019, within the context of the Spanish general election. The collection contains almost eight hundred thousand users involved in political discussions, with 5.8 million tweets. The tweet collection presents the tweets' topic mentions and keywords (in the form of political bag-of-words) and the sentiment score. The users' collection includes one field indicating the likelihood of one account being a bot. The dataset combines features retrieved from Twitter and some features computed using Botometer. The dataset provides features at user and tweet level, favoring the analysis of topics related to the Spanish general election and evaluating some tasks as bot detection. The dataset can be accessed at: https://data.mendeley.com/datasets/6cmyyxswyp/1.

4 Discussion of results and challenges

This section answers some important questions related to studying misleading information in Spanish. Each of the parts of this section addresses a particular question. The questions aim to elucidate the pros and cons that we have when addressing this problem in our language, allowing us to show the gaps and challenges that we will have to face to study the problem. Finally, we propose a work agenda that allows us to address this topic's main challenges in the upcoming years.

4.1 How far our community has come exploring the problem?

Our survey shows that the study of this problem in Spanish has gained attention in recent years. Many of the works that our SLR has detected are related to competitions and challenges, highlighting among them the IberEval and PAN at CLEF competitions. The other works that address the problem and that are not related to these competencies are rather sporadic. In many cases, they are due to specific groups or even individual efforts rather than articulated the community's challenges. In Fig. 5, we show the relationships between this survey's works, indicating with colors the specific tasks addressed by each one and indicating the citations between them.

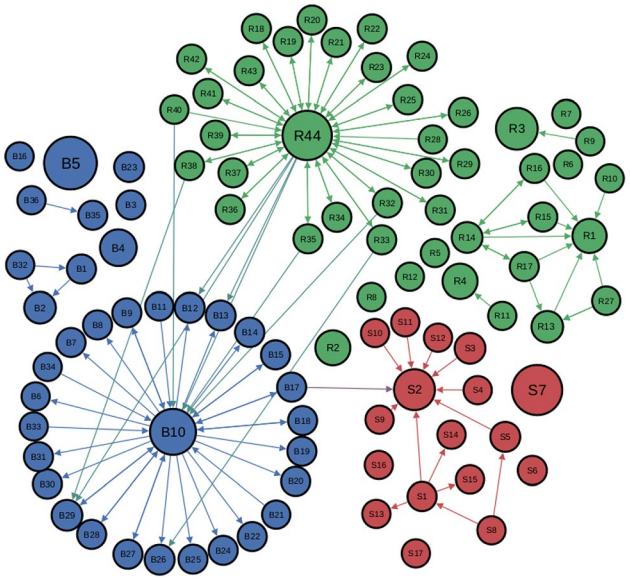

Fig. 5 Citation network of the work surveyed in this study. Each node's size indicates the number of citations that this work has received (including citations from work not surveyed in this SLR). Colors indicate specific tasks (green for fake news and rumor verification, red for stance classification, and blue for bot detection). Each node's label indicates a paper whose reference is indicated in Table 2, which can be reviewed in "Appendix"

The citation network shows weak cohesion between the works of our SLR, with very few citations between them. Most of the citations refer to the works in which the challenges are described, as nodes B10 (bot profiling task), S2 (stance classification task), and R44 (fake news spreaders detection task). The citation network also shows that stance classification is less explored in Spanish (see works in red), and these works have only a few citations between them. In light of this evidence, we can indicate that our community is weakly articulated around this problem. Despite some efforts to push the issue based on competencies, its development in Spanish shows many challenges.

4.2 How broad is the community that is concerned about this problem?

To answer this question, we analyzed the affiliations of each of the authors of the papers detected by our SLR. We related each of the affiliations to one country. Then, we counted the number of papers involved in the SLR attributable to each country, showing these results in Fig. 6. Note that some papers may count more than once, as in some cases, the papers are written by authors from distinct countries.

Figure 6 shows that Spain and Mexico lead this community. Curiously, non-Spanish-speaking countries such as Italy, India, and the USA also push the study of this problem. These countries' appearance may be due to the promotion that competitions have in big conferences such as CLEF. The contributions of authors from other countries are rather sporadic, showing that the attention that our community has paid to this problem in Spanish is not very broad.

4.3 What are the methodological biases that have limited the progress of our community in this problem?

Our study shows that classical machine learning approaches, basically those based on supervised classification, are predominant in our community. This finding shows that our community has been slow to adopt deep learning-based approaches. This is due to several factors, such as the lack of Spanish resources that allow working with Spanish word embeddings and the limited availability of datasets built in our language to address these problems. The main bias that

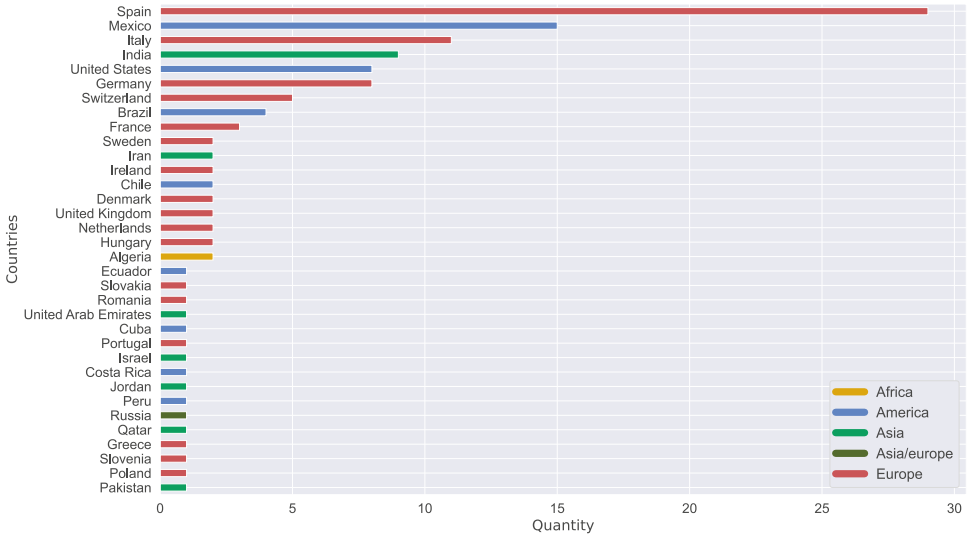

Fig. 6 Countries that lead the study of misleading information in Spanish. The plot shows a concentration of works led by authors whose affiliation corresponds to Spain and Mexico

this survey shows in our community is related to classical machine learning techniques that use representations based on handcrafted features. In this problem, techniques based on graph neural networks and architectures based on attention mechanisms, such as the Transformer (Devlin et al. 2019), show a slow adoption.

4.4 What has the English-spoken community done to address this problem? Concerning what they have done, What still appears unexplored to our community?

The adoption of techniques based on deep learning has been fast in the English-spoken community. Many of the papers devoted to fake news and rumor verification and/or stance detection use deep learning models. The availability of English word embeddings trained using various language modeling approaches has facilitated these representations in misleading information detection and analysis. We can mention that the methods that are considered state of the art in fake news and rumor verification (Ma et al. 2020) are strongly based on deep learning. Its use in this problem in our community is incipient. Regarding stance classification, some methods (Bugueño and Mendoza 2020) started using the transformer architecture, suggesting that the study of these architectures in other languages could also offer advantages. Finally, in bot detection tasks, recent methods use representation learning based on social and interaction networks (Cresci 2020; Mendoza et al. 2020). The bot detection study in Spanish shows few approaches based on these techniques.

4.5 A short agenda for the upcoming years

What is unexplored and works well in the English-spoken community sets a roadmap. Deep learning is widely used and shows good results in fake news and rumor verification, stance classification, and bot detection. When the network of interactions is a source of relevant information for the task (e.g., bot detection), methods based on graph representations are crucial. These elements that still appear as unexplored in misleading detection in Spanish allow us to configure a line of work for the next years, which are consolidated in the following agenda:

- *Creation of datasets for automatic learning in misleading information detection in Spanish* One of the crucial factors driving the study of a problem is the availability of data. The study of misleading detection in Spanish will require creating datasets that allow training automatic learning machines, which can systematically validate these tasks' predictability in our language. The creation of datasets for fake news and rumor verification, stance classification, and bot detection in Spanish is crucial for these purposes. Because the discipline's dominant methods are based on deep learning architectures, these datasets must be of adequate volume. The interrelationship between these tasks drives the creation of multitasking datasets, which allow the study of several tasks simultaneously, exploiting the dependency that these tasks have on analyzing and detecting misleading information. The development of these datasets will depend on experts' collaboration that can curate the data and the community's participation through crowdsourcing platforms that allow large volumes of data to be tagged for these purposes.
- *Promote the study of deep learning methods for the detection of misleading information in Spanish* Our survey shows that the adoption of deep learning methods in our community is slow. To accelerate its adoption, it will be necessary to develop new resources to use these architectures. The creation of repositories with Spanish word embeddings points in the right direction to push this issue. Likewise, it will be necessary to improve the availability and access to Spanish lexical resources that facilitate sentiment analysis in Spanish, strongly related to stance classification. Finally, the design of bot detection methods in Spanish will require the use and extension of methods based on graph-representation learning. The use of high-volume networks of interactions between users, as well as the release of bot detection web services specifically trained to detect bots in Spanish, is a crucial task to push this issue in the upcoming years.
- *The need for better articulation of the Spanish-spoken community* This survey has shown that the Spanish-spoken community dedicated to studying misleading information in Spanish is located in a few countries. The need to establish more and better communication channels among Spanish-speaking researchers is crucial to address this and other research problems. The competitions and challenges in big conferences have managed to articulate efforts. We believe that this is an interesting path in which we must persevere. More and better skills, for example, in stance classification in Spanish, appear as immediate tasks for our community. Furthermore, building bridges with other disciplines is crucial since misleading information in Spanish will require sociologists, political scientists, and journalists. Interdisciplinary research obtained, for example, in collaboration with Social Sciences to understand the scope of the automatic methods designed by our community is a key aspect that will show the study's maturity.

5 Conclusion

In this survey, we have reviewed work detected using an SLR for misleading detection in Spanish. Our approach addresses the problem from three complementary perspectives: the detection of fake news and rumors, the analysis of the reactions of users in social networks, and the detection of bots. Our study allows us to affirm that these three relevant variants have been studied at different depth levels. Our study also reveals that our community's efforts are poorly articulated, that there are few resources to research this topic in Spanish, and that the adoption of deep learning in this problem has been slow.

We propose a short agenda to advance these points, reducing the gap with the advances that this topic shows in the English-spoken community. The key aspects that stand out are related to the creation of more and better datasets to study the problem, the promotion of deep learning in our community to study this problem, and the urgent need for a better articulation of researchers from the Spanish-spoken community to address these issues in a collaborative and interdisciplinary manner.

This survey can be extended in many ways. The growing interest in the topic covered in this survey suggests that the coming years will show a lot of activity in the subject. Also, new relationships between social media phenomena have recently begun to be explored (Giachanou et al. 2019). For example, misleading information and hate speech show a relation based on cyberbullying, i.e., a way to generating harmful online information (Giachanou and Rosso 2020). An exciting line to explore is determining how hate speech helps amplify false and detrimental statements based on sensitive attributes, such as race, gender, ethnicities, and religion (Basile et al. 2019). The propagation of stereotyped information will undoubtedly become a new way of spreading false information with unsuspected risks and healthy coexistence threats.

Appendix

Review planning

The review planning step starts by defining search keywords, which will retrieve the first body of literature. These search keywords were defined using logical AND and OR connectors to control the coverage of documents matched by the search system. The search strings used for this process include the three variants of the problem that are the object of this study: stance, rumors, and bots. To locate the Spanish-speaking community's results, we added the keyword Spanish to each of these terms. We also include Twitter as a keyword, the social media platform that concentrates the most cited studies in English (Zhang and Ghorbani 2020). To avoid restricting the search results to English publications, we also use these search strings in Spanish. The set of search strings used in the review is shown in Table 1.

The search in Scopus was restricted to works published since 2009, ruling out works of rumors not related to this phenomenon's explosion in social media. The works were restricted to two specific areas of knowledge: Computer Science and Engineering. In this way, the retrieved papers will include works on automatic detection methods, which is the focus of this study.

The review planning process also considers the definition of inclusion/exclusion criteria. These criteria are subsequently used in the literature screening stage, during which the content of the works retrieved during the search phase is reviewed. The works that meet the inclusion criteria and do not match any exclusion criteria are included within the literature's definitive body. We first define a list of exclusion criteria with four items:

- Exclusion criteria 1 (ExCr1): When an article appears in more than one search, it will be considered only once. Accordingly, the articles repeated in the search results are eliminated, as well as versions of the same work published in different media (duplication by media).
- Exclusion criteria 2 (ExCr2): Articles written in a language other than Spanish or English are not considered.
- Exclusion criteria 3: Reviews (ExCr3-a), editorials (ExCr3-b), notes and erratum (ExCr3-c), and conference reviews ((ExCr3-d)) are not considered.

Table 1 Search strings used in the SLR, A total of 4506 results were retrieved using these search strings

ID	Search string	Results
B1	(fake news OR fake-news) AND Spanish	82
B2	bots AND Spanish	830
B3	stance AND Spanish	382
B4	(rumour OR rumor) AND Spanish	129
B5	Twitter AND Spanish	2417
B6	microblogging AND Spanish	340
B7	noticias falsas AND Español	3
B8	bots AND Español	39
B9	rumor AND Español	6
B10	redes sociales AND Español	75
B11	Twitter AND Español	175
B12	fake news AND spreaders	27
B13	difusores AND noticias falsas	1

- Exclusion criteria 4 (ExCr4): Articles whose title or abstract do not refer to the study (semantic mismatch) are discarded.

The inclusion criteria consider two items:

- Inclusion criteria 1 (InCr1): Three sections of the work are reviewed. These are abstract, introduction, and conclusion. We verify if the work focuses on solving any of the tasks object of this study in Spanish.
- Inclusion criteria 2 (InCr2): If there is no conclusive evidence identified when applying inclusion criteria 1, the full article is read. If the work does not address any of the tasks in the Spanish language, the paper is discarded.

Literature search and screening

The search for papers was carried out during 2020. By applying the search strings to Scopus, we retrieved a total of 4506 documents. This first body of literature was examined, applying exclusion and inclusion criteria defined in this study. Figure 7 shows how many documents were deleted after applying the criteria. The reduction of the initial set is notorious. A total of 4360 documents were eliminated using the exclusion criteria. The remaining 146 documents were analyzed using inclusion criteria. The first inclusion criterion was validated in 102 documents, of which 67 also match the second inclusion criterion. As a result, the first body of literature records 67 documents.

The second body of literature was created by analyzing the works that cite the first body of literature. The citations include related work relevant to these articles, which provides an important source of papers connected to the survey subject that was not detected using search strings. A total of 194 documents were identified in this process, which was reduced to 69 after applying the exclusion criteria, and 27 after applying the inclusion criteria.

Both stages of the systematic review made it possible to identify a total of 94 documents. We conducted an exhaustive review of their references for these documents, looking for works related to this survey subject that had not been detected in the previous two stages. In this last process, three more papers were identified, which passed the exclusion criteria matching both inclusion criteria. In total, the

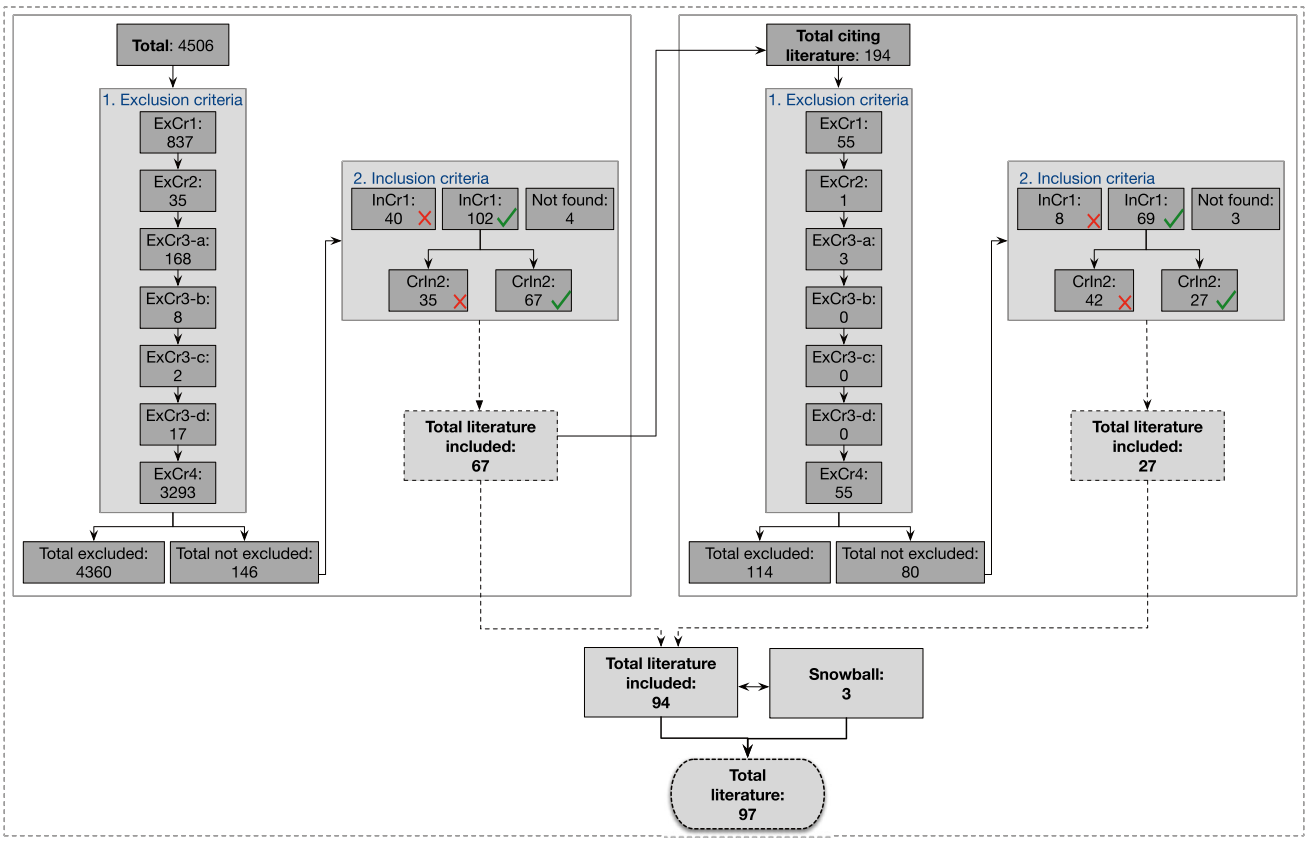

Fig. 7 Exclusion/inclusion criteria applied to the documents detected in this SLR. The study considered two stages, the first based on the documents identified using search strings and the second based on the articles that cited the first body of literature. A total of 94 documents met the exclusion/inclusion criteria. Finally, after reviewing the selected documents' references, 3 more papers were added to the survey validating the exclusion/inclusion criteria

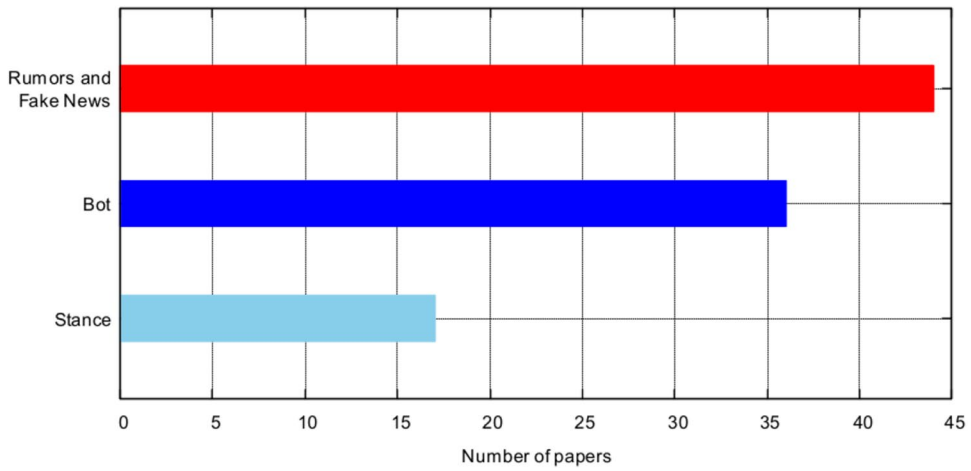

Fig. 8 Papers per task

SLR allowed the identification of 97 works related to the subject of this survey. The total number of papers per task is shown in Fig. 8.

Acronyms

- Systematic literature review: SLR
- Bag-of-Words: BOW
- Part-of-Speech: POS
- Term Frequency Inverted Document Frequency: TF-IDF
- Latent Semantic Analysis: LSA
- Universal Language Model Fine-Tuning: ULMFiT
- Singular Value Decomposition: SVD
- Recurrent Neural Network: RNN
- Bidirectional Encoder Representations based on the Transformer: BERT
- Supervised Autoencoder: SAE
- Pointwise Mutual Information: PMI
- Affective Norms for English Words: AFINN
- Linguistic Inquiry and Word Count: LIWC
- Named Entity Recognition: NER
- Global Vectors for word representation: GloVe
- Support Vector Machines: SVM
- Random Forests: RF
- Logistic Regression: LR
- Convolutional Neural Networks: CNN
- Long Short-Term Memory: LSTM
- Adaptive Boosting: ADABOOST
- Feed-Forward Neural Networks: FFNN
- Multinomial Naive Bayes: MNB
- Document frequency selection: DF
- Frequently co-occurring entropy: FCE
- Information Gain: IG
- Whale optimization: WO
- Genetic algorithms: GA
- Particle swarm optimization: PSO
- Hierarchical Attention Networks: HAN
- Bidirectional LSTM: Bi-LSTM
- Gated Recurrent Unit: GRU
- Spanish Billion Word Corpus and Embeddings: SBWCE
- The Catalonia Independence Corpus: CIC

Acknowledgements Mr. Mendoza acknowledge funding from the Millennium Institute for Foundational Research on Data. Mr. Mendoza was also funded by ANID PIA/APOYO AFB180002 and ANID FONDECYT 1200211.

References

Abonizio HQ, de Morais JI, Tavares GM, Barbon Junior S (2020) Language-independent fake news detection: English, Portuguese, and spanish mutual features. Future Internet 12(5):87

Agirrezabal M (2020) KU-CST at the profiling fake news spreaders shared task. In: Working Notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Al-Zoubi A, Faris H, Alqatawna J, Hassonah M (2018) Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts. Knowl-Based Syst 153:91–104

Almendros Cuquerella C, Cervantes Rodríguez C (2018) CriCa Team: MultiModal Stance detection in tweets on Catalan 1Oct Referendum (MultiStanceCat). In: Proceedings of the third workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval) colocated with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN), Sevilla, Spain, volume 2150 of CEUR Workshop Proceedings, pp 167–172

Ambrosini L, Nicolò G (2017) Neural models for StanceCat shared task at IberEval 2017. In: CEUR-WS, Conference of 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017, vol 1881, pp 210–216

Aragón ME, Jarquín-Vásquez HJ, Montes-y-Gómez M, Escalante HJ, Pineda LV, Gómez-Adorno H, Posadas-Durán JP, Bel-Enguix G (2020) Overview of MEX-A3T at iberlef 2020: fake news and aggressiveness analysis in Mexican Spanish. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)

Table 2 Labels used to deploy the citation network (see Fig. 5) of the work surveyed in this study

Task	References	Task	References
R1	Posadas-Durán et al. (2019)	S6	González et al. (2017)
R2	Boididou et al. (2018)	S7	Gómez et al. (2013)
R3	Congosto et al. (2017)	S8	Lai et al. (2020)
R4	Sánchez-Casado et al. (2015)	S9	García and Larriba Flor (2017)
R5	Montañés et al. (2018)	S10	Barbieri (2017)
R6	Barrón-Cedeño et al. (2020)	S11	Wojatzki and Zesch (2017)
R7	Valarezo-Cambizaca and Rodríguez-Hidalgo (2019)	S12	Ambrosini and Nicolò (2017)
R8	Caled and Silva (2019)	S13	Cuquerella and Rodríguez (2018)
R9	Rosa (2019)	S14	Segura-Bedmar (2018)
R10	Pimentel and Portugal (2020)	S15	González et al. (2018)
R11	Cegarra-Navarro and Martelo-Landroguez (2020)	S16	Graells-Garrido et al. (2020)
R12	Vogel and Meghana (2020)	S17	Zotova et al. (2020)
R13	Abonizio et al. (2020)	B1	Gallagher et al. (2019)
R14	Aragón et al. (2020)	B2	Suárez-Serrato et al. (2018)
R15	Villatoro-Tello et al. (2020)	B3	Khaund et al. (2018)
R16	Zaizar-Gutiérrez (2020)$^{*}_{MEX-A3T}$	B4	Volkova and Bell (2017)
R17	Arce-Cardenas et al. (2020)	B5	Al-Zoubi et al. (2018)
R18	Vogel and Meghana (2020)	B6	Srinivasarao and Manu (2019)
R19	Ikae and Savoy (2020)	B7	Ashraf et al. (2019)
R20	Cardaioli et al. (2020)	B8	Giachanou and Ghanem (2019)
R21	Majumder and Das (2020)	B9	Gishamer (2019)
R22	Agirrezabal (2020)	B10	Rangel and Rosso (2019)
R23	Giglou et al. (2020)	B11	Onose et al. (2019)
R24	Shashirekha and Balouchzahi (2020)	B12	Pizarro (2019)$^{*}_{PAN2019}$
R25	Lichouri et al. (2020)	B13	Valencia et al. (2019)
R26	Espinosa et al. (2020b)	B14	Goubin et al. (2019)
R27	Miranda et al. (2020)	B15	Petrik and Chuda (2019)
R28	Shashirekha et al. (2020)	B16	Oliveira et al. (2019)
R29	Fernández and Ramírez (2020)	B17	Bounaama and Abderrahim (2019)
R30	Koloski et al. (2020)	B18	Polignano et al. (2019)
R31	Pinnaparaju et al. (2020)	B19	Vogel and Jiang (2019)
R32	Bakhteev et al. (2020)	B20	Halvani and Marquardt (2019)
R33	Espinosa et al. (2020a)	B21	Bolonyai et al. (2019)
R34	López and Martí (2020)	B22	Jimenez-Villar et al. (2019)
R35	Manna et al. (2020)	B23	Castillo et al. (2019)
R36	Pizarro (2020)$^{*}_{PAN2020}$	B24	Bacciu et al. (2019)
R37	Buda and Bolonyai (2020)	B25	Fernquist (2019)
R38	Hashemi et al. (2020)	B26	Espinosa et al. (2019)
R39	Das et al. (2020)	B27	Przybyła (2019)
R40	Bello et al. (2020)	B28	Gamallo and Almatarneh (2019)
R41	Shrestha et al. (2020)	B29	Johansson (2019)
R42	Labadie et al. (2020)	B30	HaCohen-Kerner et al. (2019)
R43	Russo (2020)	b31	Fagni and Tesconi (2019)
R44	Pardo et al. (2020)	B32	Richards et al. (2019)
S1	Taulé et al. (2018)	B33	Van Halteren (2019)
S2	Taulé et al. (2017)	B34	López-Santillán et al. (2019)
S3	Vinayakumar et al. (2017)	B35	Pastor-Galindo et al. (2020b)
S4	Swami et al. (2017)	B36	Pastor-Galindo et al. (2020a)
S5	Lai et al. (2017)$^{*}_{IberEval2017}$		

Each label indicates if the paper is devoted to bot detection (B), stance classification (S) or rumor and fake news detection (R)

co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, 23 September 2020, volume 2664 of CEUR Workshop Proceedings, pp 222–235. CEUR-WS.org

Arce-Cardenas S, Fajardo-Delgado D, Carmona MÁÁ (2020) Tecnm at MEX-A3T 2020: fake news and aggressiveness analysis in Mexican Spanish. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, 23 September 2020, volume 2664 of CEUR Workshop Proceedings, pp 265–272. CEUR-WS.org

Ashraf S, Javed O, Adeel M, Rao H, Nawab M (2019) Bots and gender prediction using language independent stylometry-based approach notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th Working Notes of Conference and Labs of the Evaluation Forum. CLEF, vol 2380

Bacciu A, Morgia M, Mei A, Nemmi E, Neri V, Stefa J (2019) Bot and gender detection of twitter accounts using distortion and LSA notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th Working Notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Bakhteev O, Ogaltsov A, Ostroukhov P (2020) Fake news spreader detection using neural tweet aggregation. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Barbieri F (2017) Shared task on stance and gender detection in tweets on catalan independence—LaSTUS system Description. In: CEUR-WS Conference of 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 217–221

Barrón-Cedeño A, Elsayed T, Nakov P, Da San Martino G, Hasanain M, Suwaileh R, Haouari F (2020) CheckThat! at CLEF 2020: enabling the automatic identification and verification of claims in social media. In: Conference of 42nd European Conference on IR Research, ECIR, in Lecture Notes in Computer Science, 12036 LNCS. Springer, pp 499–507

Basile V, Bosco C, Fersini E, Nozza D, Patti V, Pardo FMR, Rosso P, Sanguinetti M (2019) Semeval-2019 task 5: multilingual detection of hate speech against immigrants and women in twitter. In: Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2019, Minneapolis, MN, USA, 6–7 June 2019. Association for Computational Linguistics, pp 54–63

Bello HRM, Heilmann L, Ronan E (2020) Detecting fake news spreaders with behavioural, lexical and psycholinguistic features. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Boididou C, Papadopoulos S, Zampoglou M, Apostolidis L, Papadopoulou O, Kompatsiaris Y (2018) Detection and visualization of misleading content on Twitter. Int J Multimed Inf Retrieval 7(1):71–86

Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. Trans Assoc Comput Linguist (TACL) 5:135–146

Bolonyai F, Buda J, Katona E (2019) Bot or not: a two-level approach in author profiling notebook for PAN at CLEF 2019. In: CEUR-WS Conference of 20th Working Notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Bounaama R, Abderrahim M (2019) Tlemcen university: bots and gender profiling task notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th Working Notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Brereton P, Kitchenham BA, Budgen D, Turner M, Khalil M (2007) Lessons from applying the systematic literature review process within the software engineering domain. J Syst Softw 80(4):571–583

Buda J, Bolonyai F (2020) An ensemble model using n-grams and statistical features to identify fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Bugueño M, Mendoza M (2020) Learning to combine classifiers outputs with the transformer for text classification. Intell Data Anal 24(S1):15–41

Caled D, Silva M (2019) FTR-18: collecting rumours on football transfer news. In: CEUR-WS, Conference on Information and Knowledge Management Workshops, CIKM, vol 2482

Cardaioli M, Cecconello S, Conti M, Pajola L, Turrin F (2020) Fake news spreaders profiling through behavioural analysis. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World Wide Web, WWW 2011, Hyderabad, India, 28 March–1 April 2011, pp 675–684

Castillo S, Allende-Cid H, Palma W, Alfaro R, Ramos H, Gonzalez C, Elortegui C, Santander P (2019) Detection of bots and cyborgs in Twitter: a study on the Chilean Presidential Election in 2017. In: Conference of 11th international conference on Social Computing and Social Media, SCSM 2019, held as part of the 21st International Conference on Human–Computer Interaction, HCI, in Lecture Notes in Computer Science, LNCS, vol 11578. Springer, pp 311–323

Cegarra-Navarro J-G, Martelo-Landroguez S (2020) The effect of organizational memory on organizational agility: testing the role of counter-knowledge and knowledge application. J Intellect Capital 21(3):459–479

Cer D, Yang Y, Kong S, Hua N, Limtiaco N, John RS, Constant N, Guajardo-Cespedes M, Yuan S, Tar C, Strope B, Kurzweil R (2018) Universal sentence encoder for English. In: Proceedings of the 2018 conference on Empirical Methods in Natural Language Processing, EMNLP 2018: System Demonstrations, Brussels, Belgium, 31 October–4 November 2018. Association for Computational Linguistics, pp 169–174

Chung CK, Pennebaker JW (2012) Linguistic inquiry and word count (liwc): pronounced luke . . . and other useful facts

Clark K, Luong M, Le QV, Manning CD (2020) ELECTRA: pre-training text encoders as discriminators rather than generators. In: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, 26–30 April 2020

Congosto M, Basanta-Val P, Sanchez-Fernandez L (2017) T-Hoarder: a framework to process Twitter data streams. J Netw Comput Appl 83:28–39

Cresci S (2020) A decade of social bot detection. Commun ACM 63(10):72–83

Cruz FL, Troyano JA, Pontes B, Ortega FJ (2014) Building layered, multilingual sentiment lexicons at synset and lemma levels. Expert Syst Appl 41(13):5984–5994

Cücük D, Can F (2020) Stance detection: a survey. ACM Comput Surv 53(1):1–37

Das KA, Baruah A, Barbhuiya FA, Dey K (2020) Ensemble of ELECTRA for profiling fake news spreaders. In: Cappellato L, Eickhoff C, Ferro N, Névéol A (eds) Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Davis CA, Varol O, Ferrara E, Flammini A, Menczer F (2016) Botornot: a system to evaluate social bots. In: Proceedings of the 25th international conference on World Wide Web, WWW

2016, Montreal, Canada, 11–15 April 2016, Companion volume, pp 273–274

Devlin J, Chang M, Lee K, Toutanova K (2019) 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, NAACL-HLT 2019, Minneapolis, MN, USA, 2–7 June 2019, volume 1 (Long and Short Papers), pp 4171–4186

Espinosa D, Gómez-Adorno H, Sidorov G (2019) Bots and gender profiling using character bigrams notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Espinosa DY, Gómez-Adorno H, Sidorov G (2020a) Profiling fake news spreaders using character and words n-grams. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Espinosa MS, Centeno R, Rodrigo Á (2020b) Analyzing user profiles for detection of fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Fagni T, Tesconi M (2019) Profiling twitter users using autogenerated features invariant to data distribution notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Fernández JL, Ramírez JAL (2020) Approaches to the profiling fake news spreaders on twitter task in English and Spanish. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Fernquist J (2019) A four feature types approach for detecting bot and gender of twitter users notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Gallagher E, Suárez-Serrato P, Velazquez Richards E (2019) Socialbots whitewashing contested elections; a case study from Honduras. Adv Intell Syst Comput 797:547–552

Gamallo P, Almatarneh S (2019) Naive-Bayesian classification for bot detection in twitter notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

García D, Larriba Flor A (2017) Stance detection at IberEval 2017: a biased representation for a biased problem. In: CEUR-WS, Conference of 2nd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 204–209

Germani F, Biller-Adorno N (2020) The anti-vaccination infodemic on social media: a behavioral analysis. Lancet Digit Health 2(10):504–505

Giachanou A, Ghanem B (2019) Bot and gender detection using textual and stylistic information notebook for pan at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Giachanou A, Rosso P (2020) The battle against online harmful information: the cases of fake news and hate speech. In: CIKM '20: the 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, 19–23 October 2020. ACM, pp 3503–3504

Giachanou A, Rosso P, Crestani F (2019) Leveraging emotional signals for credibility detection. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, 21–25 July 2019. ACM, pp 877–880

Giglou HB, Razmara J, Rahgouy M, Sanaei M (2020) Lsaconet: a combination of lexical and conceptual features for analysis of fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Gishamer F (2019) Using hashtags and pos-tags for author profiling notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

González J-A, Pla F, Hurtado L (2017) ELiRF-UPV at IberEval 2017: stance and gender detection in tweets. In: CEUR-WS, Conference of 2nd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 193–198

González J, Hurtado L, Pla F (2018) ELiRF-UPV at MultiStanceCat 2018. In: Proceedings of the third workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval) colocated with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN), Sevilla, Spain, volume 2150 of CEUR Workshop Proceedings, pp 173–179

Goubin R, Lefeuvre D, Alhamzeh A, Mitrović J, Egyed-Zsigmond E, Ghemmogne Fossi L (2019) Bots and gender profiling using a multi-layer architecture notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Graells-Garrido E, Baeza-Yates R, Lalmas M (2020) Every colour you are: stance prediction and turnaround in controversial issues. In: 12th ACM Conference on Web Science, pp 174–183

Gómez V, Kappen H, Litvak N, Kaltenbrunner A (2013) A likelihood-based framework for the analysis of discussion threads. World Wide Web 16(5–6):645–675

HaCohen-Kerner Y, Manor N, Goldmeier M (2019) Bots and gender profiling of tweets using word and character N-grams notebook for PAN at CLEF

Halvani O, Marquardt P (2019) An unsophisticated neural bots and gender profiling system notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Hashemi A, Zarei MR, Moosavi MR, Taheri M (2020) Fake news spreader identification in twitter using ensemble modeling. notebook for PAN at CLEF 2020. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. In: Proceedings of the 56th annual meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, 15–20 July 2018, volume 1: Long Papers. Association for Computational Linguistics, pp 328–339

Ikae C, Savoy J (2020) Unine at PAN-CLEF 2020: profiling fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Jimenez-Villar V, Sánchez-Junquera J, Montes-Y-Gómez M, Villaseñor-Pineda L, Ponzetto S (2019) Bots and gender profiling using masking techniques notebook for pan at clef 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Johansson F (2019) Supervised classification of twitter accounts based on textual content of tweets notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Khaund T, Al-Khateeb S, Tokdemir S, Agarwal N (2018) Analyzing social bots and their coordination during natural disasters. In: Conference of 11th International Conference on Social

Computing, Behavioral-Cultural Modeling, and Prediction conference and Behavior Representation in Modeling and Simulation, SBP-BRiMS, in Lecture Notes in Computer Science, LNCS, vol 10899. Springer, pp 207–212

Kollanyi B, Howard PN, Woolley SC (2016) Bots and automation over twitter during the first U.S. election. Data Memo 2016.4. Oxford, UK: Project on Computational Propaganda

Koloski B, Pollak S, Skrlj B (2020) Multilingual detection of fake news spreaders via sparse matrix factorization. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Labadie R, Castro-Castro D, Bueno RO (2020) Fusing stylistic features with deep-learning methods for profiling fake news spreader. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Lai M, Cignarella A, Farías D (2017) ITACOS at IberEval2017: detecting stance in Catalan and Spanish tweets. In: CEUR-WS, Conference of 2nd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 185–192

Lai M, Cignarella A, Hernández Farías D, Bosco C, Patti V, Rosso P (2020) Multilingual stance detection in social media political debates. Comput Speech Lang 63:101075

Lichouri M, Abbas M, Benaziz B (2020) Profiling fake news spreaders on twitter based on TFIDF features and morphological process. Notebook for PAN at CLEF 2020. In: Working Notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Liu H, Singh P (2004) Conceptnet—a practical commonsense reasoning tool-kit. BT Technol J 22:211–226

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) Roberta: a robustly optimized BERT pretraining approach. CoRR arXiv:1907.11692

López Á, Martí P (2020) Profiling fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

López-Santillán R, González-Gurrola L, Montes-Y-Gómez M, Ramírez-Alonso G, Prieto-Ordaz O (2019) An evolutionary approach to build user representations for profiling of bots and humans in twitter notebook for PaN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Ma J, Gao W, Joty SR, Wong K (2020) An attention-based rumor detection model with tree-structured recursive neural networks. ACM Trans Intell Syst Technol (ACM-TIST) 11(4):42:1-42:28

Magallón Rosa R (2019) Verificado Mexico 2018. Disinformation and fact-checking on electoral campaign [Verificado México (2018) Desinformación y fact-checking en campaña electoral]. Revista de Comunicacion 18(1):234–258

Majumder S, Das D (2020) Detecting fake news spreaders on twitter using universal sentence encoder. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Manna R, Pascucci A, Monti J (2020) Profiling fake news spreaders through stylometry and lexical features. unior NLP @pan2020. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Mendoza M, Poblete B, Castillo C (2010) Twitter under crisis: can we trust what we rt? In: Proceedings of the 1st workshop on Social Media Analytics, SOMA 2010, Washington, USA, 28 June 2010, pp 71–79

Mendoza M, Tesconi M, Cresci S (2020) Bots in social and interaction networks: detection and impact estimation. ACM Trans Inf Syst (TOIS) 39(1):1–32

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J, (2013) Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems, 2013 Proceedings of a meeting held December 5–8, 2013. Lake Tahoe, Nevada, United States, pp 3111–3119

Mohammad S, Turney PD (2013) Crowdsourcing a word-emotion association lexicon. Comput Intell 29(3):436–465

Molina-González MD, Martínez-Cámara E, Martín-Valdivia MT, Perea-Ortega JM (2013) Semantic orientation for polarity classification in Spanish reviews. Expert Syst Appl 40(18):7250–7257

Montañés R, Aznar R, Nogueras S, Segura P, Langarita R, Meléndez E, Peña P, Del Hoyo R (2018) Social media monitoring [Monitorizacion de Social Media]. Procesamiento de Lenguaje Natural 61:177–180

Oliveira R, De Andrade C, Figuerêdo J, Rocha-Junior J, Calumby R, Da Conceição Silva I, Da Silva Neto A (2019) Bot and gender identification: textual analysis of tweets notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Onose C, Nedelcu C-M, Cercel D-C, Trausan-Matu S (2019) A hierarchical attention network for bots and gender profiling notebook for PaN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Pardo FMR, Giachanou A, Ghanem B, Rosso P (2020) Overview of the 8th author profiling task at PAN 2020: profiling fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings

Pastor-Galindo J, Zago M, Nespoli P, Bernal SL, Celdrán AH, Pérez MG, Valiente JAR, Pérez GM, Mármol FG (2020a) Spotting political social bots in twitter: a use case of the 2019 Spanish general election. IEEE Trans Netw Serv Manag 17(4):2156–2170

Pastor-Galindo J, Zago M, Nespoli P, Bernal SL, Celdrán AH, Pérez MG, Valiente JAR, Pérez GM, Mármol FG (2020b) Twitter social bots: the 2019 Spanish general election data. Data Brief 32:106047

Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pp 1532–1543

Petrik J, Chuda D (2019) Bots and gender profiling with convolutional hierarchical recurrent neural network notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Pimentel B, Portugal R (2020) Fake news in Spanish: towards the building of a corpus based on Twitter. Commun Comput Inf Sci (CCIS) 1070:333–339

Pinnaparaju N, Indurthi V, Varma V (2020) Identifying fake news spreaders in social media. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Pizarro J (2019) Using N-grams to detect Bots on Twitter Notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th Working Notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Pizarro J (2020) Using n-grams to detect fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September

2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Polignano M, De Pinto M, Lops P, Semeraro G (2019) Identification of bot accounts in Twitter using 2D CNNs on user-generated contents notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Posadas-Durán J-P, Gomez-Adorno H, Sidorov G, Escobar J (2019) Detection of fake news in a new corpus for the Spanish language. J Intell Fuzzy Syst 36(5):4868–4876

Przybyła P (2019) Detecting bot accounts on twitter by measuring message predictability notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Rangel F, Rosso P (2019) Overview of the 7th author profiling task at Pan 2019: bots and gender profiling in twitter. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Russo I (2020) Sadness and fear: classification of fake news spreaders content on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Salazar ME, Tenorio AG, Naranjo ZL (2020) Evaluation of the precision of the binary classification models for the identification of true or false news in Costa Rica. Revista Iberica de Sistemas e Tecnologias de Informacao (RISTI) 2020(E38):156–170

Saralegi X, Vicente IS (2013) Elhuyar at tweet-norm 2013. In: Proceedings of the tweet normalization workshop co-located with 29th conference of the Spanish Society for Natural Language Processing (SEPLN 2013), Madrid, Spain, 20 September 2013, pp 64–68

Segura-Bedmar I (2018) LABDA's early steps toward multimodal stance detection. In: Proceedings of the third workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval) colocated with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN), Sevilla, Spain, volume 2150 of CEUR Workshop Proceedings, pp 180–186

Shashirekha HL, Balouchzahi F (2020) Ulmfit for twitter fake news spreader profiling. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Shashirekha HL, Anusha MD, Prakash NS (2020) Ensemble model for profiling fake news spreaders on twitter. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Shrestha A, Spezzano F, Joy A (2020) Detecting fake news spreaders in social networks via linguistic and personality features. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Speer R, Chin J, Havasi C (2016) Conceptnet 5.5: an open multilingual graph of general knowledge. CoRR, arXiv:1612.03975

Srinivasarao M, Manu S (2019) Bots and gender profiling using character and word N-grams notebook for PAN at CLEF 2019. In: Conference of 20th working notes of CLEF Conference and Labs of the Evaluation Forum, vol 2380

Suárez-Serrato P, Richards E. Velázquez, Yazdani M (2018) Socialbots supporting human rights. In: AIES—Proceedings AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery, Inc, Conference of 1st AAAI/ACM—AI, Ethics, and Society, AIES, pp 290–296

Swami S, Khandelwal A, Shrivastava M, Akhtar S (2017) LTRC IIITH at IBEREVAL 2017: stance and gender detection in tweets on catalan independence. In: CEUR-WS, Conference of 2nd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 199–203

Swire-Thompson B, Lazer D (2020) Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 41(1):433–451

Sánchez-Casado N, Cegarra-Navarro J, Tomaseti-Solano E (2015) Linking social networks to utilitarian benefits through counterknowledge. Online Inf Rev 39(2):179–196

Taulé M, Martí M, Rangel F, Rosso P, Bosco C, Patti V (2017) Overview of the task on stance and gender detection in tweets on catalan independence at IberEval 2017. In: CEUR-WS, Conference of 2nd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 157–177

Taulé M, Rangel F, Martí M Antònia, Rosso P (2018) Overview of the task on multimodal stance detection in Tweets on catalan #1Oct referendum. In: CEUR-WS, Conference of 3rd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 2150, pp 149–166

Tiedemann J (2012) Parallel data, tools and interfaces in OPUS. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, 23–25 May 2012. European Language Resources Association (ELRA), pp 2214–2218

Valarezo-Cambizaca L-M, Rodríguez-Hidalgo C (2019) Innovation in journalism as an antidote to fake news [La innovación en el periodismo como antídoto ante las fake news]. RISTI Revista Iberica de Sistemas e Tecnologias de Informacao E20:24–35

Valencia A Valencia, Adorno H, Rhodes C, Pineda G (2019) Bots and gender identification based on stylometry of tweet minimal structure and n-grams model notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Van Halteren H (2019) Bot and gender recognition on tweets using feature count deviations Notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Varol O, Ferrara E, Davis CA, Menczer F, Flammini A (2017) Online human-bot interactions: detection, estimation, and characterization. In: Proceedings of the eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, 15–18 May 2017, pp 280–289

Velazquez Richards E, Gallagher E, Suárez-Serrato P (2019) Boostnet: bootstrapping detection of socialbots, and a case study from Guatemala. In: Conference of 33rd National Forum of Statistics, FNE 2018 and 13th Latin-American Congress of Statistical Societies, CLATSE, vol 301. Springer, pp 145–154

Villatoro-Tello E, Ramírez-de-la-Rosa G, Kumar S, Parida S, Motlícek P (2020) Idiap and UAM participation at MEX-A3T evaluation campaign. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, 23 September 2020, volume 2664 of CEUR Workshop Proceedings. CEUR-WS.org, pp 252–257

Vinayakumar R, Kumar S Sachin, Premjith B, Prabaharan P, Soman K (2017) Deep stance and gender detection in tweets on catalan independence@Ibereval 2017. In: CEUR-WS, Conference of 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 222–229

Vogel I, Jiang P (2019) Bot and gender identification in Twitter using word and character n-grams notebook for PAN at CLEF 2019. In: CEUR-WS, Conference of 20th working notes of Conference and Labs of the Evaluation Forum, CLEF, vol 2380

Vogel I, Meghana M (2020) Fake news spreader detection on twitter using character n-grams. In: Working notes of CLEF 2020—Conference and Labs of the Evaluation Forum, Thessaloniki,

Greece, 22–25 September 2020, volume 2696 of CEUR Workshop Proceedings. CEUR-WS.org

Volkova S, Bell E (2017) Identifying effective signals to predict deleted and suspended accounts on Twitter across languages. In: Proceedings of the 11th International Conference on Web and Social Media, ICWSM. AAAI Press, pp 290–298

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151

Wojatzki M, Zesch T (2017) Neural, non-neural and hybrid stance detection in tweets on catalan independence. In: CEUR-WS, Conference of 2nd workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval, vol 1881, pp 178–184

Yang Y, Cer D, Ahmad A, Guo M, Law J, Constant N, Ábrego GH, Yuan S, Tar C, Sung Y, Strope B, Kurzweil R (2020) Multilingual universal sentence encoder for semantic retrieval. In: Proceedings of the 58th annual meeting of the Association for Computational Linguistics: System Demonstrations, ACL 2020, Online, 5–10 July 2020. Association for Computational Linguistics, pp 87–94

Zaizar-Gutiérrez D, Fajardo-Delgado D, Carmona M Á Á (2020) Itcg's participation at MEX-A3T 2020: aggressive identification and fake news detection based on textual features for Mexican Spanish. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, 23 September 2020, volume 2664 of textitCEUR Workshop Proceedings. CEUR-WS.org, pp 258–264

Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization, detection, and discussion. Inf Process Manag 57(2):102025

Zhou X, Zafarani R (2020) A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput Surv (CSUR) 53(5):1–40

Zotova E, Agerri R, Nuñez M, Rigau G (2020) Multilingual stance detection: the catalonia independence corpus, 03

Zubiaga A, Kochkina E, Liakata M, Procter R, Lukasik M (2016) Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In: COLING 2016, 26th international conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 11–16 December 2016, Osaka, Japan, pp 2438–2448

Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: a survey. ACM Comput Surv 51(2):32:1-32:36