Dr Burgert Senekal

Department of Computer Science and Informatics, University of the Free State, Bloemfontein, South Africa
Email: burgertsenekal@ yahoo.co.uk
ORCID: https://orcid. org/0000-0002-1385- 9258

THE DEPICTION OF ORANIA IN PRINT MEDIA (2013-2022): A QUANTITATIVE ANALYSIS USING NATURAL LANGUAGE PROCESSING (NLP)

ABSTRACT

The current article investigates the depiction of the town of Orania in print media. Being an exclusive Afrikaner town, this town is controversial and is often seen as a remnant of apartheid, leading residents of the town to form the perception that the media treats them unfairly. Using Natural Language Processing (NLP) techniques, namely a lexicon-based sentiment analysis classification and a machine-learning political bias classification, it is shown that the majority of news reports and opinion pieces on this town exhibit minimal political bias, and publications on this town are evenly distributed between left and right political bias. In addition, while the majority of news reports and opinion pieces published on this town are neutral, more publications are positive than negative. However, differences in the depiction of this town based on the language of publications are also discussed, with English publications more negative and Afrikaans publications more positive, and the majority of publications on this town are in Afrikaans. Overall, the study finds that while some individual publications present Orania in a negative light, in general, the media reports on this town in a balanced way.

Keywords: media studies, media fairness, journalism, Orania, machine learning, political bias classification, sentiment analysis, South African media

INTRODUCTION

Given South Africa's post-1994 attempt to integrate different ethnicities, and government pressure to promote English-only spaces (such as in education), Orania's emphasis on preserving the traditional Afrikaner identity is inevitably controversial. Although other exclusively Afrikaner communities such as Kleinfontein have also come under criticism, it is Orania in particular that is often presented in the media as the bastion of racism. Orania is seen from this point of view as a remnant of apartheid, where apartheid is synonymous

with racism. Seldon (2014) notes that Orania has become "a by-word for racism and the nostalgic recreation of a time when they [Afrikaners] held power over others", and Davis (2020) remarks, "[Orania] has become a byword for unchecked racism; a metonym for all unreconstructed white South Africans and problematic race relations".

As Orania is associated with racism, the town receives a great deal of attention from the media, and Roets (2019), then acting Chief Executive Officer of the Orania Movement, stated that as many as three international news teams visit Orania per week. The Orania Movement mostly oversees media liaison, taking journalists on tours and arranging interviews. Some journalists, such as Davis (2020), attempt to portray what they experience as an unusual town in a nuanced way but, as she writes, there is a perception among many residents of Orania that journalists treat them unfairly. According to Roets (2019), the town has received negative publicity for the past 30 years, while Delvecki and Greiner (2014) aver, "the town has endured negative press and misunderstandings from its beginning".

The present article investigates the nature and extent of print media coverage of Orania using Natural Language Processing (NLP) techniques, and more specifically, sentiment analysis and political bias classification. The purpose is to determine whether this controversial community is treated fairly by the media, or if news reports show bias and a tendency to frame this community in a negative light. Note that the current article focuses on the depiction of Orania in the media and does not argue for or against this community.

The article is structured as follows. First, some key terms used in the current article are discussed. Thereafter, background is provided on current trends in terms of people's trust in the media, as well as text analysis approaches that attempt to make news reports more transparent. This is followed by a discussion of Orania's troubled history with the media. Then, data gathering and analysis are discussed, followed by a presentation and discussion of the findings. The article concludes with summary remarks and suggestions for further research.

A NOTE ON TERMINOLOGY

Some terms used in the current study warrant explanation, and therefore the current section briefly discusses the terms "Afrikaner" and "political leaning".

Afrikaner

In the simplest definition, the term "Afrikaner" refers to Afrikaans-speaking people of predominantly European descent. Hollfelder *et al*. (2020) for example found that on average 95.3% of the Afrikaner's ancestors are of European origin, with the remaining 4.7% people brought to the Cape as slaves in colonial times (3.4%) and from local Khoe-San people (1.3%). This European genetic heritage means that the Afrikaner is generally considered to be white, and Steyn (1984) for instance argues that the definition of the Afrikaner offered by Abel Coetzee is the most accurate: "'n Blanke Afrikaanssprekende inwoner van Suid-Afrika" [*A white Afrikaans-speaking resident of South Africa*].

In addition to language and genetic heritage, traditional Afrikaners are often Christian and conservative. Finlayson (2019) found that the typical characteristics attributed to Afrikaners can be summarised as follows: a common genetic ancestry, adherence to Protestant Christianity, use of the Afrikaans language, and the adoption of a distinct set of social values often referred to as conservatism (see also Vestergaard (2001)). These facets of Afrikaner identity are also found in the comprehensive definition offered by Orania's founder, Prof. Carel Boshoff (2012), from the constitution of the Afrikaner Vryheidstigting [*Afrikaner Freedom Foundation*, or Avstig]:

> 'n Kultuur-politieke gemeenskap van Afrikaanssprekende Suid-Afrikaners van oorwegend Europese afkoms wat aan die hand van 'n gedeelde verlede en gedeelde verwagtings van die toekoms 'n selfbewuste bestaan voer, wat as 'n volk erken word en oor soewereiniteit in 'n eie staat beskik het, wat steeds aan taal en kultuur herken word en wat die reg het om sy identiteit te handhaaf en om nie aan gedwonge inlywing uitgelewer te word nie. [*A cultural-political community of Afrikaans-speaking South Africans of predominantly European descent who lead a self-conscious existence on the basis of a shared past and shared expectations of the future, who are recognised as a people and who possessed sovereignty in their own state, who are still recognised by language and culture and have the right to maintain their identity and not to be relinquished to forced incorporation*.]

Orania aims to preserve traditional Afrikaner culture, and therefore generally follows the definition of the Afrikaner used by Boshoff (2012).

Left and right political leanings

"Political leaning" as used in the current study refers to the traditional left/right divide, where left-wing ideals encompass liberal concepts such as individual freedom, equality, fraternity, human rights, progress, reform, and internationalism, whereas right-wing ideals encompass conservative concepts such as authority, hierarchy, order, duty, tradition, reaction, and nationalism (Heywood 2015). In this sense, Orania is generally regarded as a right-wing community because of its emphasis on traditional values and ethnicity. Residents also mostly support the conservative Freedom Front Plus (FF+), and leaders in the community, including Boshoff and his sons, Carel (IV) and Wynand, have worked for the FF+, while there are no formal ties with the more liberal Democratic Alliance (DA).

RELIABILITY AND OBJECTIVITY OF MAINSTREAM MEDIA

In the contemporary world, mainstream media and science have been increasingly questioned by the general public. Kavanagh and Rich (2018) refer to "truth decay" in the contemporary world, which they argue is described by four characteristics:

♦ increasing disagreement about facts and analytical interpretations of facts and data;

♦ a blurring of the line between opinion and fact;

♦ the increasing relative volume, and resulting influence, of opinion and personal experience over fact; and

♦ declining trust in formerly respected sources of factual information.

Mainstream media have been primary targets of this "declining trust in formerly respected sources of factual information", with both television news and print media (nowadays available online) affected (Kavanagh & Rich 2018). Mainstream media is often accused of furthering political agendas, biased reporting, manipulating the public, obscuring the truth, and even fabricating stories. While the decline in trust in science is less marked than for the media, trust in science has also eroded (Kavanagh & Rich 2018). Kavanagh and Rich (2018) state that the decline in trust in these institutions is partly due to unethical practices of some of these institutions, such as falsifying data or manipulating the peer review process in terms of scientific publications, or biased and inaccurate reporting by some media outlets. The resulting lack of trust in established providers of information "leaves people searching for new sources of credible and objective information and increases uncertainty about basic facts, data, and analysis as people turn to new entities, not all of them trustworthy, to fill this vacuum" (Kavanagh & Rich 2018). These alternative sources of information are often not comparably reliable. Toepfl *et al*. (2022) for instance found that mainstream media usually refrain from spreading conspiracy theories, but niche and special interest media play a key role in spreading conspiracy theories, along with the Russian state-sponsored news channel, Sputnik. Xu *et al*. (2022) in turn found that mainstream media outlets such as CNN and BBC were the fifth and third most reliable news sources respectively out of the 13 news agencies in their study, compared to Russia's RT (formerly Russia Today) being the third most unreliable news source. Nevertheless, as people lose trust in Western mainstream media, they turn to alternative media such as Sputnik, RT, or other alternative news sources. This turn to alternative and unreliable sources of information contributes to the spread of misinformation and conspiracy theories, for instance in reference to the COVID-19 pandemic (Johnson & Marcellino 2021; Van Mulukom *et al*. 2022).

One way of regaining trust in mainstream media is by reducing bias, or at least making bias explicit. Bias can be defined as "prejudice in favor of or against a particular thing, person, or group in comparison to another, usually in an unjust manner" (Raza *et al*. 2022). Bias can be based on various factors, including gender, race, ethnicity, age, disability, mental health, religion, and political ideology (Raza *et al*. 2022). Political bias is a particularly important manifestation of bias, as it may influence voter behaviour and perceptions of both domestic and foreign policy, and it also perpetuates and even exacerbates existing divisions in society.

Given that political bias in the media is such a crucial factor in determining the credibility of the source, various websites focus on highlighting the bias inherent in different news sources. One of the most well-known of these is Media Bias/Fact Check (https://mediabiasfactcheck.com/), a website that provides information about the bias and credibility of news sources by country, including South Africa. Similarly, Biasly (www.biasly.com) provides bias rankings of news sources around the world, including some South African sources. In addition, Biasly provides a tool (www.biasly. com) created using Artificial Intelligence (AI) that allows a user to determine the bias of a particular online publication, similar to the tool provided by The Bipartisan Press (www.thebipartisanpress.com).

As Iyyer *et al*. (2014) argue, it is impractical and expensive to detect ideological bias in political texts manually, especially in the era of big data. Therefore, various studies have investigated the feasibility of using machine learning to classify political bias. Iyyer *et al*. (2014) for instance used Recursive Neural Networks (RNNs) combined with Word2Vec (Le & Mikolov 2014) to train a model to classify news reports in terms of political bias. Chun *et al*. (2019) used Support Vector Machines (SVM) (Cortes & Vapnik 1995), Convolutional Neural Networks (CNN) (LeCun & Bengio 1998), Word2Vec (Le & Mikolov 2014) and BERT (Devlin *et al*. 2018) to train a model with up to 89% accuracy. Wang (2019) describes the tool used by The Bipartisan Press, which was trained using BERT (Devlin *et al*. 2018), and later upgraded using the RoBERTa model (Liu *et al*. 2019). Raza *et al*. (2022) developed Dbias, a Python library to identify bias in news reports, also using BERT (Devlin *et al*. 2018), RoBERTa (Liu *et al*. 2019) and DestilBERT (Sanh *et al*. 2019). The bias classification tool by The Bipartisan Press (Wang 2019) is used in the current study.

Apart from political bias, another way of evaluating the way the news depicts topics is through sentiment analysis. The media tends to focus on negative stories (Soroka *et al*. 2015), and Haselmayer and Jenny (2017) showed how sentiment analysis could be used to study how the media reports on German parliamentary elections, while Van Atteveld *et al*. (2008) studied the news coverage of Dutch elections. Garvey and Maskal (2020) used the Google Cloud Natural Language API Sentiment Analysis tool to study how the media depicts AI and found that the media frames AI more positively than commonly thought. Taking a lexicon-based approach, Suryadi (2021) studied how news about COVID-19 was framed in the Indonesian media. Given the controversial nature of Orania, and perceptions that Orania receives negative media coverage, investigating whether this community is depicted in a positive or negative way will shed valuable light on whether the media depicts Orania in a balanced way.

BACKGROUND TO ORANIA'S RELATIONSHIP WITH THE MEDIA

Even before Orania was purchased in 1991, the media portrayed Afrikaners who campaigned for a *volkstaat* (Afrikaner homeland) negatively. Messina (1989) for example refers to the Oranjewerkers (Orange Workers, one of the pro-*volkstaat* institutions during the 1980s) as "die lemoenmannetjies en -vrouetjies van Morgenzon" [*the little orange men and women of Morgenzon*] whose "dorp-staat gedagtes en die beweegrede daarvoor [...] [is] 'n duidelike teken van onchristelikheid" [*town-state thoughts and the motivation for them [...] [are] a clear sign of unchristian attitudes*]. Boshoff (2012), a leading figure in the Oranjewerkers and later a founder of Avstig and Orania, also notes negative media coverage of *volkstaat* movements, "Vir die media en politieke kommentators was dit onmoontlik om die gedagtegang en groei van hierdie Afrikanervryheidsbeweging anders te interpreteer as om dit as verregs, konserwatief en die handhawing van apartheid te stigmatiseer" [*For the media and political commentators it was impossible to interpret the thinking and growth of this Afrikaner freedom movement other than to stigmatise it as far right, conservative and for the maintenance of apartheid*].

With the founding of Orania in 1991, its difficult relationship with the media began (Senekal 2021). The Afrikaans newspaper *Die Burger* called Orania a "rassistiese lugkasteel" [*a racist castle in the sky*] (Anonymous 1991b) in 1991. Even before Orania's official opening on 13 April 1991, the first settlers had already developed an aversion to journalists, and one of them, Johan Moolman, said, "Hulle had nie eers die pit om met ons te kom praat nie, maar hardloop na die Kleurlinge toe om dan agterna hier in die dorp te kom rondsluip" [*They didn't even have the guts to come and talk to us, but ran to the Coloureds and then sneak around here in town*] (Anonymous 1991b). Moolman was referring to the eviction of Coloured inhabitants prior to the opening of the town, as is described in detail in Cavanagh (2013) and Senekal (2021).

After former President Nelson Mandela's visit in 1995, Sello (1995) wrote about the "racists" of Orania in *City Press*, while Mkhondo (2001) called Orania an "ethnic and racist enclave" and suggested Oranians "[can] go to hell". McNally (2010) states his view equally clearly: "I found myself splitting people into two categories: true racist and racist from horrific crime incidents."s According to McNally (2010), Afrikaners moved to Orania "to be racist in peace", Orania is a "bio-dome of apartheid" and the "last blemish of the old South Africa". He also refers to Boshoff's son and then president of the Orania Movement, Carel Boshoff (IV), as an "intellectual racist".

Alleged racism remains a recurring theme in some reporting on Orania, and more than a decade after McNally (2010), Thamm (2021) refers to Orania as a cult, and states,

> Single men live in single quarters. Single mothers who arrive with children are immediately set to work and are 'rehabilitated' by the Kaalvoet Vrouens – the barefoot women. It is these matriarchs who will set you on the path to Christianity, the restoration of your dignity and your Afrikaner pioneer identity. Arbeit macht frei. Look it up.

Her use of German terms from the Nazi era in her report, such as *Lebensraum* and *Heimat*, along with the above motto from Auschwitz, suggests that Orania has a Nazi connection, although this suggestion is never explicitly stated.

False reporting on Orania by major news organisations is usually followed by formal complaints by the Orania Movement. In May 2019, a news team from eNCA visited Orania, and after the report, the newsreader, Vuyo Mvoko, said that black people are only welcome in Orania as domestic workers and gardeners – which is something that was not said by the reporters (Roets 2019). The Orania Movement lodged a complaint with the Broadcasting Complaints Commission of South Africa (BCCSA) and the case concluded in Orania's favour (Roets 2019; Strydom 2019).

While false reporting by major news outlets is usually addressed by the Orania Movement through official complaints, other false claims – often on social media – are usually left to supporters of Orania or the general public to address. In 2022, a user shared a "news report" on Twitter (now the social media platform X) which claimed that a 95-year-old woman had given birth in Orania after a "delayed pregnancy" of more than four decades, but users quickly spotted that the report was fake (Chloerii 2022).

In response to negative reporting, especially from foreign journalists, Roets (2019) posted a notice at Stokkiesdraai Shopping Centre in Orania that reads, "Attention all white journalists from Europe. Please leave your prejudice at the entrance." The notice was translated into Dutch, German and French.

The negative view of Orania has also been accompanied throughout its history with threats that the town will be demolished and destroyed. Even before Orania's opening in 1991, *Die Transvaler* (Anonymous 1991a) reported that Oranians received threats, for example that they would be "eliminated". Opperman (2016) also states that in the early 2000s, Orania received threats that people wanted to destroy the town.

In today's milieu, such threats are found, among other places, on social media, and in particular on Twitter (X), as discussed in Kotzé and Senekal (2018). Kotzé and Senekal (2018) for instance cite tweets such as, "How about we accidentally get our hands on some grenades and start genocide? Orania is a good place to start" and "That Orania shit place must be burnt down we want them back in Europe or Australia". In this study it was found that Orania is mostly depicted in a negative light on X, and that Orania is mostly mentioned on this social media platform in the context of racism.

Despite the perception of negative publicity, and the examples discussed above, news outlets often report on Orania's successes and development. Criticism of Orania is to be expected, but a cursory glance at news reports on Orania shows that the severely negative examples discussed above do not constitute the totality of media reports on the town. For instance, Wyngaard (2018) and Snyman (2013) write about Orania in positive terms. To determine whether this controversial community is treated fairly by the media, one must also take the numerous positive reports into account, as is done in the current study.

METHODS

Data gathering

Orania's archive of newspaper publications was used as data in the current study. This collection includes news reports, letters to news outlets and opinion pieces that mention Orania and constitute the most extensive news collection on Orania that is currently available. The collection covers a wide range of publication outlets, although more than 50% are from Netwerk24. This collection was previously digitised (Senekal & Kotzé 2018). During the digitisation process, efforts were made to include the best available copy, and therefore only publications that could not be found in a digital format were scanned, while others were stored in Portable Document Format (PDF) straight from web publications (Senekal & Kotzé 2018). This means that for the latter publications, no correction of Optical Character Recognition (OCR) errors is necessary, while the former would require extensive efforts to correct OCR errors and produce a perfect text for processing purposes. For this reason, only clippings published since 2013 are used in the current study, as this is the date when publications became available in a digital format. The author has permanent access to the archive for research purposes.

Because the NLP techniques applied below were designed for analysing text written in English, Afrikaans publications were first translated to English using Google Translate. While Google Translate's outputs were not expected to be perfect, Aiken (2019) and Benjamin (2019) have shown that translations between Afrikaans and English are some of the most accurate translations produced by this service.

Note that the current study focuses on print media, and television news reports are not included. This means that many foreign reports will not be included in the analysis below. In addition, some news outlets such as *The New York Times* did not publish anything on Orania in the period under investigation, and while 82 publication platforms are included in the total collection, only 59 published on Orania between 2013 and 2022. While the collection includes publications in all languages, for the period under investigation, only Afrikaans and English publications are available.

Sentiment analysis

Sentiment analysis is an NLP technique that involves classifying a text as positive, neutral, or negative. Two popular approaches to sentiment classification are lexicon-based and machine learning approaches. In the case of Orania, Kotzé and Senekal (2018) used a lexicon-based approach. In the current study, the lexicon-based sentiment analysis method developed by Levallois (2013) was used, which has been demonstrated to be one of the most effective techniques for sentiment analysis (Ribeiro *et al*. 2016). The output is simply a classification between positive, negative, and neutral, without providing an indication of the level of sentiment that occurs in the text. This method was applied to Orania's news collection from 2013 to 2022 using the online set of tools available at https://nocodefunctions.com/, which include the necessary text preprocessing steps such as lemmatisation, the removal of stopwords, and the removal of special characters.

As Orania is generally associated with the Afrikaner and Afrikaans, it was expected that Afrikaans news outlets would more often depict Orania in a positive light than English news outlets, although as the background section showed, negative publications on Orania occur in Afrikaans as well as English news sources. In addition, as Orania is considered to be conservative and right wing, it was expected that right biased reports would present Orania in a positive light, while left biased reports were expected to depict Orania in a negative light.

Political bias classification

Political bias classification using machine learning involves employing a machine learning model to classify whether a given text exhibits left or right political bias. For political bias classification, the tool developed by The Bipartisan Press (www.thebipartisanpress.com), as described by Wang (2019), was used. This tool produces an output stating if the text is left or right biased and offers the level of bias (minimal, moderate, strong, or extreme), but it does not classify texts as neutral. A minimum of 100 words is necessary for an accurate classification using this model.

While this model's accuracy has not been formally tested in a South African context, the concept of right and left bias is not specific to the political climate in the United

States where the tool was developed, and therefore the model was expected to also classify texts accurately in a South African setting. Nevertheless, spot checks were carried out, and the model for instance classified the election manifesto of the FF+, as published by News24 (Anonymous 2014), as moderate right, and Afrikaans singer Steve Hofmeyr's views on the Afrikaner (Hofmeyr 2013) as strong right, while the criticism of Orania by Thamm (2021) was classified as moderate left, which in all cases agree with the author's own assessment.

As Orania is considered to be conservative, traditional and right wing, it was expected that right biased reports would depict Orania in more favourable terms, as opposed to left biased publications. The four examples above illustrate what was expected: Hofmeyr has performed in Orania, and Boshoff, as well as his sons, Wynand and Carel (IV), have worked for the FF+. If Hofmeyr or the FF+ refer to Orania, they are generally expected to do so in positive terms. On the other hand, Thamm is negative about Orania, and from her publication and the classification using the tool by The Bipartisan Press, it is clear that she writes from a left perspective. In other words, bias and sentiment were expected to show some agreement. In addition, as the Afrikaner is generally considered to be a more conservative ethnic group (Vestergaard 2001; Finlayson 2019), it was expected that Afrikaans publications would more often be biased right in order to cater to an Afrikaans audience.
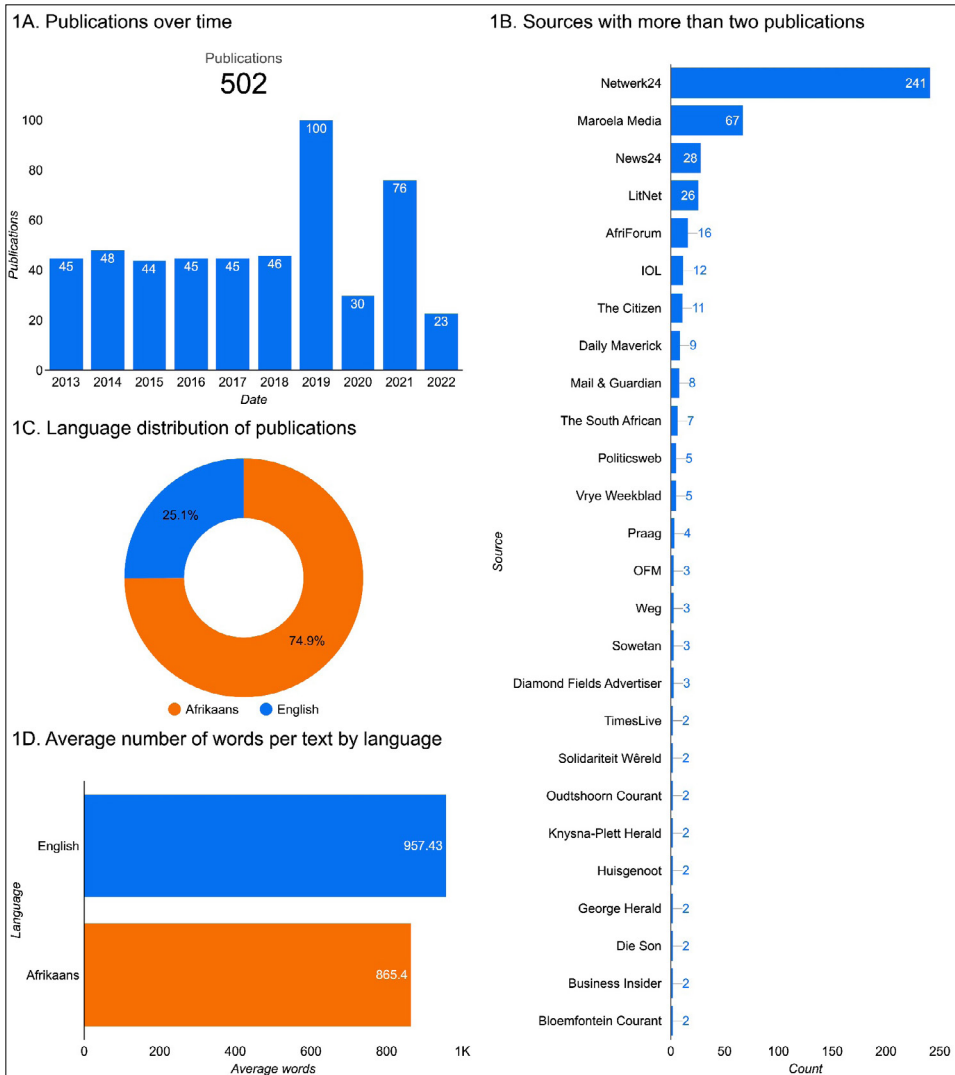
RESULTS

Overview of the dataset

Figure 1 provides an overview of the dataset. Figure 1A shows the publication dates of texts, and as the graph shows, there is no discernible pattern that would suggest either a general increase or a decrease in the attention that Orania receives in the media. The high number of publications in 2019 is related to controversy in this year around Boshoff (IV) (at the time the president of the Orania Movement), and the amount of attention Orania received around that year's general election.

Figure 1B shows the publication platforms where more than two documents in this dataset were published for the period 2013 to 2022, and it can be seen here that while a wide variety of publication platforms are included, the vast majority of documents were published by Netwerk24 and Maroela Media. This distribution across platforms already suggests that the amount of attention paid to Orania is not equally distributed between Afrikaans and English news outlets.

Figure 1C shows the language distribution of documents, with 75% Afrikaans and 25% English. This is important as Orania receives vastly more attention in Afrikaans news publications than in English, which could affect the way that Orania is depicted.

Figure 1D shows the average number of words per text for Afrikaans and English documents, with English texts being slightly longer. The highest number of words per text in this dataset is 7 628 words and the lowest number 95 words, and together with the averages shown in Figure 1D, most texts exceed the minimum requirement of 100 words for political bias classification using the model by Wang (2019).

FIGURE 1: A SUMMARY OF THE DATASET

Sentiment analysis

Figure 2 shows the results of the sentiment analysis. Figure 2A shows that the largest segment of documents (38%) was classified as neutral, followed by positive (36%) and negative (27%). Despite some very negative outliers as noted in the background section, publications in the media about Orania are therefore generally neutral or positive, with negative publications forming the minority of publications and only constituting slightly more than a quarter of publications.

Figure 2B shows sentiment over time, and it can be seen here that there is a general positive trend from 2015 onwards, with positive publications in 2021 exceeding those from the other categories. Negative sentiment however fluctuates between years, without any discernible pattern.

Figures 2C and 2D show the sentiment distribution by language, and it can be seen here that Afrikaans publications are more often positive (41%), followed by neutral (35%), with the minority of publications classified as negative (25%). This is in sharp contrast with English publications, which tend to more often be neutral (45%), followed by negative (34%), with the smallest category being positive (21%). As noted earlier, this is to be expected, given Orania's emphasis on Afrikaans, which may lead to Afrikaans media outlets being more positively disposed towards this community.

Figure 2E visualises the relationship between language and sentiment for both languages, with the thickness of ties indicating higher percentages. Neutral publications are split almost evenly between Afrikaans and English, but negative publications are more often produced in English, and positive publications are more often produced in Afrikaans. These three figures (2C, 2D and 2E) show that publications in Afrikaans media take a more positive stance towards Orania, while English media report more negatively on this community.
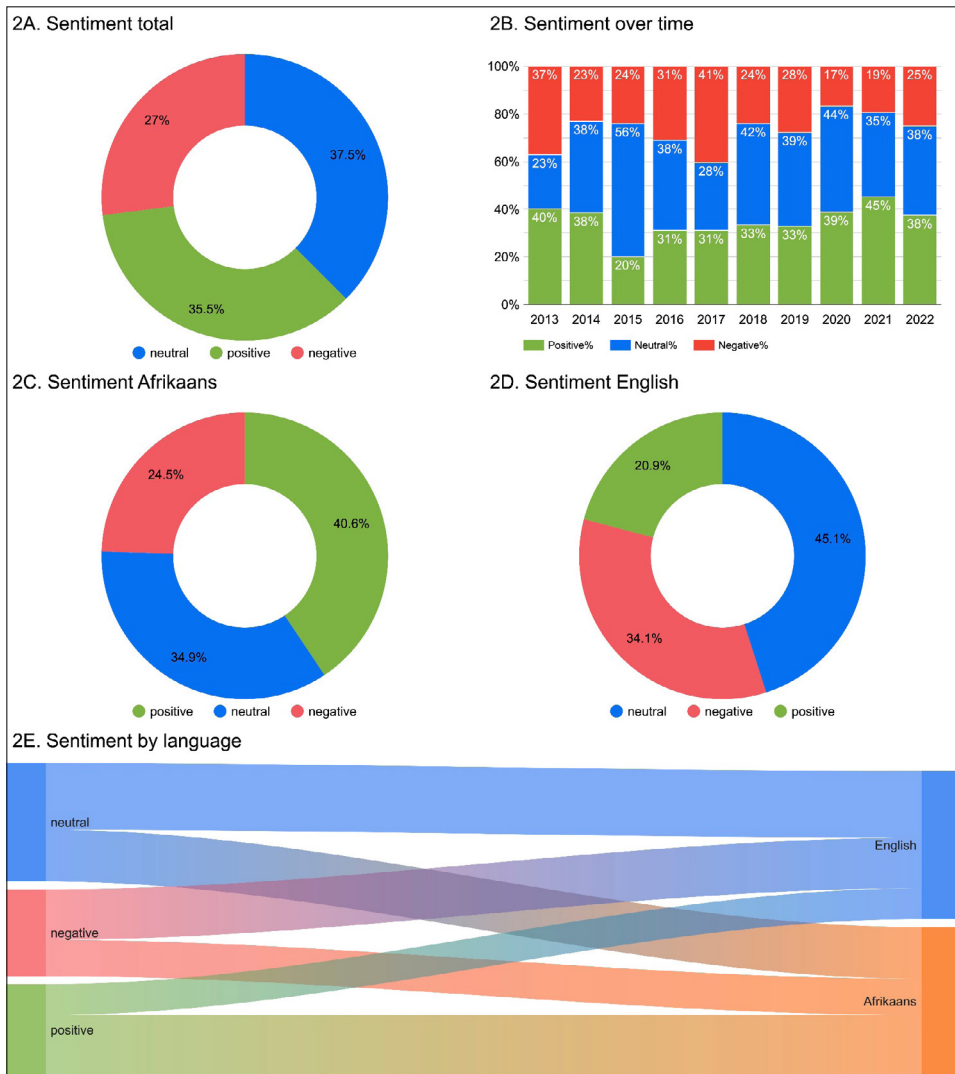
FIGURE 2: SENTIMENT ANALYSIS RESULTS

Political bias classification

Figure 3 provides the results of the political bias classification. Figure 3A shows that political bias is almost evenly distributed through this dataset, but with a slight preference for right biased publications.

Figure 3B shows that the majority of publications were classified as exhibiting minimal bias, both for left and right bias. Note that this model does not classify texts as neutral, but if it did, many of the minimal bias classifications could have been neutral classifications. When it comes to moderate bias, more left leaning publications

exhibited moderate bias than is the case for right bias, with 7.37% of publications showing moderate left bias, compared with 5.78% of right leaning publications showing moderate bias. Strong bias is rare in this dataset, but even when it does occur, it is more often left leaning (1.58% of publications) than right leaning (1.2% of publications). No publication was classified as exhibiting extreme bias.

Figure 3C shows that the even distribution in Figure 3A applies to time segments as well, with some years showing more right bias and others showing more left bias. There is no discernible pattern with regards to publication bias over time: publications that mention Orania are not becoming either more or less left or right biased.

The same holds for language (Figure 3D): Afrikaans and English publications are more or less evenly divided between right and left bias, with English publications showing a slightly greater right bias (58.73% of English publications) than Afrikaans publications (51.06% of Afrikaans publications). While this is a surprising finding, since Afrikaans publications were expected to be more right biased than English publications, the difference between the languages remains small.

Figure 3E shows political bias by sentiment. Contrary to expectations, right biased publications did not overwhelmingly depict Orania in a positive light, nor did left biased publications present Orania in a negative light. Rather, sentiment is fairly evenly spread between left and right bias. Indeed, negative sentiment is more often expressed in right biased publications than in left biased publications, with 41.05% of negative reports classified as left biased, and 58.95% of negative publications classified as right biased.
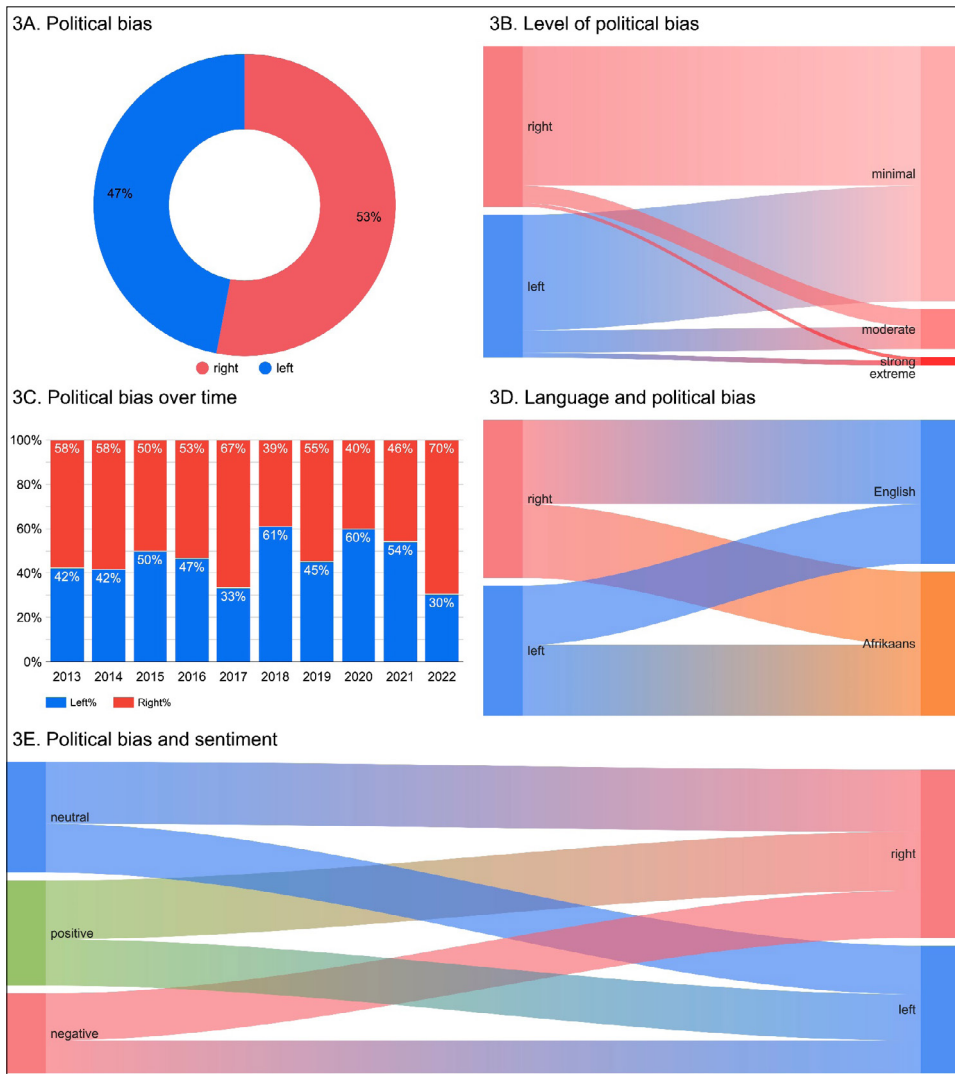
FIGURE 3: THE RESULTS OF THE POLITICAL BIAS CLASSIFICATION

DISCUSSION

From the above, it is clear that Orania is not generally represented in a negative light in the media. While the largest segment of publications was classified as neutral, positive publications constitute the second largest segment. However, as expected, Afrikaans publications on Orania tend to be more positive, while English publications are often negative.

It is also clear that strongly biased depictions of Orania are the exception rather than the rule. The majority of documents in the current study showed only minimal bias, which could be considered close to neutral. In addition, the distribution of bias is more or less equal, with right biased reports occurring only slightly more frequently (53%) than left biased reports (47%). Contrary to expectations, bias is also fairly evenly distributed across Afrikaans and English publications, with English publications even showing a slightly greater tendency to be right biased than Afrikaans publications. Also contrary to expectations, right biased publications did not overwhelmingly report on Orania in positive terms, but even showed a slight tendency to report negatively on this community. Left biased reports, on the other hand, showed a surprising tendency to report positively on Orania.

The above shows that media reporting on Orania is far more balanced than the perception would suggest. While positive reporting on Orania in Afrikaans media was expected, a large percentage of English publications also depicted Orania in a positive light (although negative publications were more common). Left or right bias is also not an accurate predictor of the sentiment of a publication, with both presenting Orania in a positive as well as a negative light. Overall, while some journalists have shown prejudice towards this community and there have been cases where facts were distorted, in general, standards of professionalism seem to apply to this controversial community.

An important question that neither sentiment analysis nor bias classification answers is why recent print media reports on Orania tend to be more positive. One possible explanation is that in light of the South African government's failures, for example in terms of electricity provision, service delivery and crime, Orania's successes are increasingly recognised. It may also be that Orania's self-help attitude is increasingly seen as a solution to government failures. To determine why Orania is increasingly depicted in a positive light, a thematic analysis of news reports will have to be conducted, which was not attempted here.

As mentioned before, the current study's focus on print media meant that television news reports were not included. This is an important limitation, since the mentioned distortion of facts by eNCA was a television news piece, and other television news reports could also depict Orania in more biased ways.

CONCLUSION

This study investigated the depiction of Orania using two NLP methods, namely sentiment analysis and bias classification. It was shown that the depiction of Orania is generally balanced, with positive and negative reports in similar percentages (with slightly more positive than negative reports), and only a few publications showing more than minimal bias, while bias itself is also roughly evenly distributed. In general, the media seems to approach this controversial community in a balanced way, although some extreme examples were also discussed. It should also be noted that interest in this community comes mostly from Afrikaans news outlets, although interest is not confined to Afrikaans news outlets.

If mainstream media is to retain its role as a provider of trustworthy information, eliminating bias and reporting on topics in a balanced way will be important. The current study has shown that, apart from a small number of exceptions, the media has generally reported on this controversial community in a balanced way.

While Orania is one controversial South African topic, similar techniques could be applied to study how the media depicts other controversial South African topics. Haselmayer and Jenny (2017) argue that these text analysis techniques open up a wide variety of research questions, among which the study of public opinion, media framing and political polarisation are areas where sentiment analysis might be useful at the nexus of communication science and political science.

REFERENCES

Aiken, M. 2019. An updated evaluation of Google Translate accuracy. *Studies in Linguistics and Literature* 3(3): 253. DOI: 10.22158/sll.v3n3p253.

Anonymous. 1991a. Blankes glo gedreig. *Die Transvaler*, 3 April: 2.

Anonymous. 1991b. Orania 'n rassistiese lugkasteel. *Die Burger*, 12 February: 14.

Anonymous. 2014. *Manifesto FF+*. Available at: https://www.news24.com/Elections/MANI FESTO-FF-20140311 [Accessed: 25 May 2018].

Benjamin, M. 2019. *Empirical evaluation of Google Translate across 107 languages*. Available at: https://www.teachyoubackwards.com/empirical-evaluation/ [Accessed: 25 June 2021].

Boshoff, C. 2012. *Dis nou ek*. Pretoria: LAPA Uitgewers.

Cavanagh, E. 2013. The history of dispossession at Orania and the politics of land restitution in South Africa. *Journal of Southern African Studies* 39(2): 391-407. DOI: 10.1080/03057070.2013.795811.

Chloerii. 2022. *95 year old lady allegedly gives birth after being impregnated 45 years ago*. Available at: https://opera.news/za/en/parenting/4d075565e76f1de1528bc9a 46f2fa9f6 [Accessed: 28 April 2022].

Chun, S., Holowczak, R., Dharan, K., Wang, R., Basu, S. & Geller, J. 2019. Detecting political bias trolls in Twitter data. *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, SCITEPRESS - Science and Technology Publications: 334-342. DOI: 10.5220/0008350303340342.

Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20(3): 273-297. DOI: 10.1007/BF00994018.

Davis, R. 2020. *'Everyone in Orania is woke': A journey to SA's most notorious town*. Available at: https://www.dailymaverick.co.za/article/2020-01-21-everyone-in-orania -is-woke-a-journey-to-sas-most-notorious-town/ [Accessed: 21 January 2020].

Delvecki, A. & Greiner, A. 2014. Circling of the wagons? A look at Orania, South Africa. *Focus on Geography* 57(4): 164-173. DOI: 10.1111/foge.12042.

Devlin, J., Chang, M.W., Lee, K. & Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Finlayson, K. 2019. 'ik ben een afrikaander': Redrawing Afrikaner ethnic boundaries in New Zealand. *Sites: A Journal of Social Anthropology and Cultural Studies* 16(2). DOI: 10.11157/sites-id436.

Garvey, C. & Maskal, C. 2020. Sentiment analysis of the news media on artificial intelligence does not support claims of negative bias against artificial intelligence. *Omics: A journal of Integrative Biology* 24(5): 286-299. DOI: 10.1089/omi.2019.0078.

Haselmayer, M. & Jenny, M. 2017. Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & Quantity* 51(6): 2623-2646. DOI: 10.1007/s11135-016-0412-4.

Heywood, A. 2015. *Key concepts in politics and international relations*. London: Springer. https://doi.org/10.1007/978-1-137-49477-1

Hofmeyr, S. 2013. *Wie is die Afrikaner? Steve lewer repliek*. Available at: https://maroelamedia.co.za/debat/meningsvormers/wie-is-die-afrikaner-steve-lewer-repliek/ [Accessed: 28 September 2020].

Hollfelder, N., Erasmus, J.C., Hammaren, R., Vicente, M., Jakobsson, M., Greeff, J.M. & Schlebusch, C.M. 2020. Patterns of African and Asian admixture in the Afrikaner population of South Africa. *BMC Biology* 18(1): 16. DOI: 10.1186/s12915-020-0746-1.

Iyyer, M., Enns, P., Boyd-Graber, J. & Resnik, P. 2014. Political ideology detection using recursive neural networks. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics: 1113-1122. DOI: 10.3115/v1/P14-1105.

Johnson, C. & Marcellino, W. 2021. *Bad actors in news reporting: Tracking news manipulation by state actors*. Santa Monica: RAND Corporation. DOI: 10.7249/RRA112-21.

Kavanagh, J. & Rich, M. 2018. *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Santa Monica: RAND Corporation. DOI: 10.7249/RR2314.

Kotzé, E. & Senekal, B.A. 2018. Employing sentiment analysis for gauging perceptions of minorities in multicultural societies: An analysis of Twitter feeds on the Afrikaner community of Orania in South Africa. *The Journal for Transdisciplinary Research in Southern Africa* 14(1): a564. DOI: 10.4102/td.v14i1.564.

LeCun, Y. & Bengio, Y. 1998. Convolutional Networks for images, speech, and time series. In: Arbib, M.A. (ed.). *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press.

Levallois, C. 2013. Umigon: Sentiment analysis for tweets based on terms lists and heuristics. *International Workshop on Semantic Evaluation*. *7th International Workshop on Semantic Evaluation*.

Le, Q. & Mikolov, T. 2014. Distributed representations of sentences and documents. *International conference on machine learning*: 1188.

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. *arXiv*. DOI: 10.48550/arxiv.1907.11692.

McNally, P. 2010. *Orania tourism: Come gawk at the racists*. Available at: http://thoughtleader.co.za/paulmcnally/2010/02/01/orania-tourism-come-gawk-at-the-racists/ [Accessed: 12 October 2017].

Messina, E.A. 1989. Ossewatrekke: Mites rondom Groot Trek-feeste. *Kronos: Journal of Cape History*: 30-41.

Mkhondo, R. 2001. It's a pipe dream for the bittereinders. *The Citizen*, 11 April: 12.

Opperman, M. 2016. Interview conducted by Sebastiaan Biehl. Orania Archive.

Raza, S., Reji, D.J. & Ding, C. 2022. Dbias: Detecting biases and ensuring fairness in news articles. *International Journal of Data Science and Analytics*: 1-21. DOI: 10.1007/s41060-022-00359-4.

Ribeiro, F.N., Araújo, M., Gonçalves, P., André Gonçalves, M. & Benevenuto, F. 2016. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science* 5(1): 23. DOI: 10.1140/epjds/s13688-016-0085-1.

Roets, S. 2019. *Orania – Die waarheid sal seëvier*. Available at: https://maroelamedia.co.za/debat/meningsvormers/orania-die-waarheid-sal-seevier/ [Accessed: 13 November 2019].

Sanh, V., Debut, L., Chaumond, J. & Wolf, T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv*. DOI: 10.48550/arxiv.1910.01108.

Seldon, S.R. 2014. *Orania and the reinvention of Afrikanerdom*. Unpublished doctoral thesis. University of Edinburgh, Scotland, United Kingdom.

Sello, S. 1995. Good manners from deluded folk in Orania. *City Press*, 20 August: 17.

Senekal, B.A. & Kotzé, E. 2018. Die ontwikkeling van 'n koste-effektiewe en byderwetse multimedia digitale argief by EPOG in Orania. *LitNet Akademies Geesteswetenskappe* 15(3): 239-275.

Senekal, B.A. 2021. Die eerste 40 jaar van Orania. *Tydskrif vir Geesteswetenskappe* 61(2): 526-550. DOI: 10.17159/2224-7912/2021/v61n2a8.

Snyman, D. 2013. *Van alle kante: Toe lees Madiba vir tant Betsie*. Available at: https://www.netwerk24.com/Stemme/Van-Alle-Kante-Toe-lees-Madiba-vir-tant-Betsie-20130627 [Accessed: 5 July 2021].

Soroka, S., Young, L. & Balmas, M. 2015. Bad news or mad news? Sentiment scoring of negativity, fear, and anger in news content. *The Annals of the American Academy of Political and Social Science* 659(1): 108-121. DOI: 10.1177/0002716215569217.

Steyn, J.C. 1984. Nuwe bedeling: Identiteitsverandering. *Aambeeld* 12(2): 20.

Strydom, N. 2019. *eNCA moet jammer sê vir Orania*. Available at: https://maroelamedia.co.za/nuus/sa-nuus/enca-moet-jammer-se-vir...fcba5-6a99042137-105219033&mc_cid=6a99042137&mc_eid=3d37c4e855 [Accessed: 5 July 2021].

Suryadi, D. 2021. Does it make you sad? A lexicon-based sentiment analysis on COVID-19 news tweets. *IOP Conference Series: Materials Science and Engineering* 1077(1): 012042. DOI: 10.1088/1757-899X/1077/1/012042.

Thamm, M. 2021. *Above us only sky: A road trip through South Africa's heartland*. Available at: https://www.dailymaverick.co.za/article/2021-10-12-above-us-only-sky-a-road-trip-through-south-africas-heartland/ [Accessed: 12 October 2021].

Toepfl, F., Kravets, D., Ryzhova, A. & Beseler, A. 2022. Who are the plotters behind the pandemic? Comparing Covid-19 conspiracy theories in Google search results across five key target countries of Russia's foreign communication. *Information, Communication & Society*: 1-19. DOI: 10.1080/1369118X.2022.2065213.

Van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N. & Schlobach, S. 2008. Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology & Politics* 5(1): 73-94. DOI: 10.1080/19331680802154145.

Van Mulukom, V., Pummerer, L.J., Alper, S., Bai, H., Čavojová, V., Farias, J., Kay, C.S., Lazarevic, L.B., Lobato, E.J.C., Marinthe, G., Pavela Banai, I., Šrol, J. & Žeželj, I. 2022. Antecedents and consequences of COVID-19 conspiracy beliefs: A systematic review. *Social Science & Medicine* 301: 114912. DOI: 10.1016/j.socscimed.2022.114912.

Vestergaard, M. 2001. Who's got the map? The negotiation of Afrikaner identities in post-apartheid South Africa. *Daedalus* 130(1): 19–44.

Wang, W. 2019. *Calculating political bias and fighting partisanship with AI*. Available at: https://www.thebipartisanpress.com/politics/calculating-political-bias-and-fighting-partisanship-with-ai/ [Accessed: 16 March 2023].

Wyngaard, B. 2018. *Orania, koeksisters en ek*. Available at: https://www.litnet.co.za/orania-koeksisters-en-ek/ [Accessed: 5 July 2021].

Xu, M., Luo, Z., Xu, H. & Wang, B. 2022. Media bias and factors affecting the impartiality of news agencies during COVID-19. *Behavioral Sciences (Basel, Switzerland)* 12(9). DOI: 10.3390/bs12090313.