THE AFRICAN JOURNAL OF INFORMATION AND COMMUNICATION (AJIC)

ISSUE 31, 2023

RESEARCH ARTICLES

CRITICAL INTERVENTION

Published by the LINK Centre
University of the Witwatersrand (Wits)
Johannesburg, South Africa
https://www.wits.ac.za/linkcentre

ISSN 2077-7213 (online version)
ISSN 2077-7205 (print version)

The African Journal of Information and Communication (AJIC)

Issue 31, 2023

Published by the LINK Centre, School of Literature, Language and Media (SLLM),
Faculty of Humanities, University of the Witwatersrand (Wits), Johannesburg, South Africa
https://ajic.wits.ac.za

The African Journal of Information and Communication (AJIC) is a peer-reviewed, interdisciplinary, open access academic journal focused on the myriad dimensions of electronic and digital ecosystems that facilitate information, communication, innovation and transformation in African economies and in the broader Global South. Accredited by the South African Department of Higher Education and Training (DHET), *AJIC* publishes online, free to the user, under a Creative Commons licence, and does not impose article processing charges. *AJIC* is indexed in Scientific Electronic Library Online (SciELO) SA, the Directory of Open Access Journals (DOAJ), Sabinet African Journals and Wits University WIReDSpace, and is hosted on the Academy of Science of South Africa (ASSAf) Khulisa Journals platform.

Editors

Corresponding Editor: Lucienne Abrahams, Director, LINK Centre, University of the Witwatersrand, PO Box 601, Wits 2050, Johannesburg, South Africa, ajic.submissions@gmail.com
Publishing Editor: Chris Armstrong, Research Associate, LINK Centre, University of the Witwatersrand, Johannesburg, South Africa, chris.armstrong@wits.ac.za

Dedication

This issue of *AJIC* is dedicated to Wits University's "Prof Barrry" Dwolatzky, who passed away in Johannesburg on 16 May 2023. Prof. Dwolatzky was a pioneering figure in technolology innovation, education and development at Wits, in South Africa, and internationally. *AJIC* was privileged and honoured to have him serve on our Editorial Advisory Board. A written tribute to Prof. Dwolatzky, by *AJIC* Corresponding Editor and LINK Centre Director Lucienne Abrahams, is available at: https://www.wits.ac.za/linkcentre/news/tribute-to-prof-barry-dwolatzky

Peer-Reviewing

AJIC acknowledges with gratitude the following peer reviewers of submissions to this issue: Lucienne Abrahams, Aje-Ori Agbese, Ufuoma Akpojivi, Lateef Amusa, Andrew Barendse, Mark Burke, Sha'ista Goga, Trudi Hartzenberg, Eduan Kotze, Ramona Kunene-Nicolas, Ronny Mabokela, Laston Mukaro, Blessing Ogbuokiri, Elphas Okango, Christoph Stork, Scott Timcke, and Thando Vilakazi.

Production

Sub-editing: LINK Centre
Proofreading: Linda Van de Vijver
Desktop-publishing: LINK Centre

AJIC is published by the LINK Centre, School of Literature, Language and Media (SLLM), Faculty of Humanities, University of the Witwatersrand (Wits), PO Box 601, Wits 2050, Johannesburg, South Africa. The LINK Centre is based at the Wits Tshimologong Digital Innovation Precinct, 41 Juta Street, Braamfontein, Johannesburg, https://www.tshimologong.joburg

ISSN 2077-7213 (online version)
ISSN 2077-7205 (print version)

Past issues of *AJIC*, and its precursor *The Southern African Journal of Information and Communication (SAJIC)*, are available at https://ajic.wits.ac.za/issue/archive and https://www.wits.ac.za/linkcentre/sajic

CONTENTS

RESEARCH ARTICLES

CRITICAL INTERVENTION

RESEARCH ARTICLES

"If it is circulating widely on social media, then it is likely to be fake news": Reception of, and motivations for sharing, COVID-19-related fake news among university-educated Nigerians

Chikezie E. Uzuegbunam
Lecturer, School of Journalism and Media Studies, Rhodes University, Grahamstown, South Africa
https://orcid.org/0000-0002-3958-5494

Chinedu Richard Ononiwu
PhD candidate, Department of Mass Communication, Nnamdi Azikiwe University, Awka, Nigeria
https://orcid.org/0000-0003-2900-5561

Abstract
This study explores how university-educated Nigerians living in two urban centres engaged with, and made choices about whether to share or not share, "fake news" on COVID-19 in 2020. The research adopted a qualitative approach by conducting focus group interviews with participants, all university graduates aged 25 or older, sampled from Lagos and Umuahia—two major metropolitan cities in Nigeria. Participants' sense-making practices with regard to fake news on COVID-19 were varied. One core finding was that social media virality was typically seen as being synonymous with fake news due to the dramatic, exaggerated, and sometimes illogical nature of such information. Many participants demonstrated a high level of literacy in spotting fake news. Among those who said that they sometimes shared fake news on COVID-19, one motivation was to warn of the dangers of fake news by making it clear, while sharing, that the information was false. Other participants said that they shared news without being certain of its veracity, because of a general concern about the virus, and some participants shared news if it was at least partially true, provided that the news aimed to raise awareness of the dangers of COVID-19. However, some participants deliberately shared fake news on COVID-19 and did so because of a financial motivation. Those who sought to avoid sharing fake news on COVID-19 did so to avoid causing harm. The study provides insights into the reception of, and practices in engaging with, health-related fake news within a university-educated Nigerian demographic.

Keywords
fake news, misinformation, disinformation, COVID-19, social media, news media, reception, sense-making practices, Nigeria, Lagos, Umuahia

DOI: https://doi.org/10.23962/ajic.i31.14518

Recommended citation
Uzuegbunam, C. E. & Ononiwu, C. R. (2023). "If it is circulating widely on social media, then it is likely to be fake news": Reception of, and motivations for sharing, COVID-19-related fake news among university-educated Nigerians *The African Journal of Information and Communication (AJIC), 31*, 1-23.
https://doi.org/10.23962/ajic.i31.14518

Acknowledgements
The authors are thankful to research assistants Janet Chijiuba, Ebube Nwigwe, and Nneamaka Daniella Okereke for their assistance with the facilitation of the focus groups.

1. Introduction

Social media "fake news" has been a major source of concern in Nigeria ever since the 2015 general elections. During those elections, there was a growing population of internet users, with online and social media playing an active role in citizens' vigorous political participation—and, at the same time, providing fertile ground for politicians to disseminate amplified, partisan, and distorted messages (Ogwezzy-Ndisika et al., 2023). The 2015 general elections in Nigeria were conducted at a time when the use of social media in electioneering and political participation was becoming steadily more popular, following the inaugural use of social media in the preceding elections of 2011 (Uzuegbunam, 2020). Concerns about social media fake news have since escalated to the extent that most of the ensuing political, social, and economic problems present in the country have to some extent been attributed to it (Adegoke & BBC, 2018; Anderson, 2019).

In addition to the threat that fake news poses to Nigeria's peace, unity, security, and positive international reputation, it also poses particular challenges in relation to its healthcare system, which is hampered by problems of infrastructure, policy, and outdated health beliefs, all of which have contributed to disease outbreaks (Welcome, 2011). Health-related fake news can delay or prevent effective care and, in some cases, threaten the lives of individuals, i.e., people misled by fabricated messages. Stated differently, the problem of health-related fake news becomes consequential when news consumers do not recognise a particular news item as fabricated and respond to it as true (Lara-Navarra et al., 2020; Wang et al., 2019). In the Nigerian context, research has found that individuals who rely on social media platforms as their main source of news tend to be more likely to believe fake news on health issues such as COVID-19 (Uwalaka, 2022).

The term "fake news" is somewhat contested in communication scholarship. Allcott and Gentzkow (2017) conceptualise fake news as being intentionally and verifiably false, and, at the same time, potentially deceptive to audiences. Gelfert (2018) argues that fake news should be seen as but one category of "false news". Given that even reliable news sources can occasionally make mistakes, Gelfert (2018) believes that the definition of fake news as news that contains inaccurate information is inadequate, because a justifiable mistake about an irrelevant or incorrect detail does not necessarily render the entire report fake news. For Gelfert (2018), fake news is a deliberate attempt by the originators to deceive an audience, to manipulate public opinion, and to increase the circulation of false information. Accordingly, Gelfert (2018) defines fake news as the "deliberate presentation of […] false or misleading claims as news, where the claims are misleading *by design*" (pp. 85-86, italics in original). Lilleker (2018) is also of the view that fake news is news that is deliberately misleading. These definitions seem to equate fake news with "disinformation" (false information created and spread *with the intention* to mislead), but, in our analysis, fake news is more usefully understood as comprising both disinformation and "misinformation" (false information created and spread *without an intention* to mislead, or without concern as to whether the information is false or not).

Both terms, disinformation and misinformation, are widespread in the academic literature on manifestations of false information, or what is also sometimes referred to as "information disorder" (see Wardle, 2019). In our research and this article, we use the term "fake news" to refer to any kind of false information circulating in the various media channels, including social media, regardless of whether the false information's creation, or distribution, is performed with the intent to mislead, i.e., all false information, including the intentionally misleading subset that is disinformation, can constitute fake news if disseminated via some sort of media platform. Our definition of fake news aligns with the definition used in the Madrid-Morales et al. (2021) article on sharing of misinformation, in which it is stated that that the terms "misinformation" and "fake news" are used "interchangeably to refer to all the expressions and formats in which made-up and inaccurate information has been found to be common" (p. 1201, footnote).

For Choy and Chong (2018), fake news has lexical features that are different from those of factual reports. For example, as they explain, biased information has been associated with specific linguistic cues, including active verbs, implicative verbs (verbs such as "manage to" and "bother to", which suggest that the "truth value" of a clause is conditional), and subjective intensifiers (such as "extremely" and "utterly"). Choy and Chong point to "clickbait"—a type of deceptive online content that uses special lexical elements like emotional language, action words, suspenseful language, and the overuse of numerals. The authors argue that since fake news has unique lexical features, these features can be used to detect its presence in online news content. The authors also argue that deceivers (creators of fake news) tend to tell less complex

stories. When compared to the stories of truth tellers, deceivers' messages show lower cognitive complexity. Thus, these deceptive messages have lower average sentence length and lower average word length (Choy & Chong, 2018). The messages also tend to make frequent use of motion verbs such as "walk", "move", and "go", as these provide simpler and more concrete descriptions than words that focus on evaluations and judgments (such as "think" and "believe"). With respect to the affective dimension of fake news, Choy and Chong (2018) contend that those who fabricate fake news do so with a view to appealing to the audience's emotions.

The outbreak of the COVID-19 pandemic has reignited global interest in the discourse around health-related fake news, further deepening investigation into its implications for public health. Commenting on how online fake news has worsened the spread of COVID-19, the Director-General of the World Health Organisation (WHO), Tedros Adhanom Ghebreyesus, noted that the world is faced with both a pandemic and an "infodemic"—a deluge of all manner of information, including both factual and inaccurate information, in offline and online spaces, during a disease outbreak or health crisis. Online fake news on COVID-19 has had serious impacts in parts of the world. For instance, it was responsible for deaths in Iran in March 2020, when 2,100 Iranians ingested methanol (a toxic industrial form of alcohol) after exposure to social media messages that suggested alcohol consumption could prevent infection by the virus (Soltaninejad, 2020).

In this study, we specifically explored the reception of, and motivations for sharing (or not sharing) fake news on COVID-19 in Nigeria. Here, "reception" refers to how information is interpreted and made sense of by its recipients. We draw from reception studies and thereby view "reception" as the participants' reactions to fake news (Sandikci, 1998; Wagner & Boczkowski, 2019). "Motivations", in this study, refer to the reasons or determinant factors behind recipients' decisions to share or not to share fake news.

Research aims and questions

This study sought to make a scholarly contribution in the areas of both social media reception and health communication. The aim was to empirically assess how selected cohorts of university-educated Nigerians in urban locations responded to COVID-related social media fake news, and the extent to which they shared such news. The study was grounded in two core questions:

- How do university-educated Nigerians receive and interpret fake news on COVID-19 issues?
- What are their motivations for sharing (or not sharing) fake news on COVID-19 issues?

2. Reception of, and motivations for sharing, fake news

The reception of, and motivations for sharing, fake news have both drawn significant research attention. In their Nigerian research on the role of misinformation in undermining the containment of Ebola, Allgaier and Svalastog (2015) highlight instances where the audience received false information on the virus as true. The authors cite accounts of people dying and being admitted to hospitals in Nigeria as a result of audience adoption of incorrect information about dangerous methods of combating Ebola. Also looking at factors influencing the sharing of health-related misinformation, Aquino et al. (2017) identify anti-vaxxers as major sources or propagators of misinformation. These authors find that discussions among this category of audience tend to revolve around rhetorical as well as personal arguments that induce negative emotions such as anger, fear, and sadness.

Chua and Banerjee (2017) examine the role played by epistemic belief in affecting people's decisions regarding whether or not to spread health-related rumours online, and the authors discover that people who are "epistemologically naïve" are more likely than "epistemologically robust" people to share health-related misinformation online. For Chua and Banerjee (2017), epistemologically naïve people are those who believe that knowledge is relatively rigid and easily attainable. Conversely, as noted by Chua and Banerjee (2017), people who believe that knowledge is largely fuzzy and requires significant effort to obtain are epistemologically robust. The foregoing would suggest that epistemologically naïve people are more likely to receive (and share) fake news uncritically, while the epistemologically robust will more easily identify (and not share) fake news.

Chakrabarti et al.'s (2018) study in five cities in Kenya and Nigeria identified two main categories of motivations for sharing fake news. First, individuals share fake news due to their desire to be seen as "in the know" socially, with such sharing viewed as a path to gaining social currency. Second, a sense of civic duty leads some users of social media to share warnings of, and to inform others of news received on, an imminent danger—irrespective of whether the sharer thinks the news is reliable or not. Chakrabarti et al. (2018) found an assumption among sharers that it is preferable to tell people widely just in case the information could benefit them, on the grounds that if the information about a supposed imminent danger turns out to be false, no substantial harm will be caused. However, if the information turns out to be accurate, it can have significant practical advantages for many.

The aforementioned study by Madrid-Morales et al. (2021) looked at motivations for sharing misinformation online among university students in six African countries. The study found two key motivations among this demographic. First, participants share out of a sense of civic duty, where they feel they have to warn others of inherent dangers. Second, they share misinformation for the fun of it, in order to elicit laughter or humour. Another study, by Wasserman and Madrid-Morales (2019), establishes a

link between lack of trust in the news media and the sharing of fake news. In a survey of 1,847 Kenyans, Nigerians, and South Africans, a significant relationship was found between high levels of perceived exposure to misinformation and low levels of media trust. This corresponds with similar findings elsewhere (see Chadwick and Vaccari (2019) in the British context) that suggest that the widespread sharing of false news may signify a growing cynicism towards the accuracy of news in general. In their study conducted in Chicago, Philadelphia, and Miami, Wagner and Boczkowski (2019) found that a general mistrust and scepticism regarding the veracity of the news ecosystem as a whole is linked to the consumption and sharing of fake news.

In parts of Africa, cultural influences, including the long-standing importance of informal sources of information such as gossip, rumour, and satire (Madrid-Morales et al., 2021; Nyamnjoh, 2005) can often play a role in the tendency for social media users sharing fake news. In addition to these cultural influences in the African context, the long history of untrustworthy news media on the continent, and of muzzled media environments controlled by the state or socioeconomic elites, has given rise to strong alternative channels of information on which fake news can thrive (Wasserman & Madrid-Morales, 2019). A study by Tully (2021) of how Kenyans experience misinformation found that the participants' consumption and sharing of misinformation is determined by their personal interest in a particular topic, the extent to which it trends within their social networks, and the perceived importance of the information.

The motivations for sharing information, including false information on social media, can also be psychological or emotional. Drawing from data from cross-sectional surveys in the US, Petersen et al. (2023) argue that psychological motivations underpin the sharing of hostile political rumours, a form of false news, as participants in their study felt a personal burden to challenge the political system as a whole and to mobilise receivers of such messaging against a particular political setting. Dafonte-Gómez's (2018) study finds that some sharing practices are motivated by affective connections and emotions because of the heightened emotion they feel while interacting with viral news on social media.

Furthermore, other scholars have found that the social identity of the audience is a factor that can influence their sharing of information on social media. Bigman et al. (2019), in their online survey of 150 college students of black, white, and "other" races/ethnicities in the US, found that race influences how young social media users selectively expose themselves to news on social media. Black students, more so than the students of other races/ethnicities, reported seeing and posting race-related content on social media. Bigman et al. (2019) also found an orientation towards civic participation or civic purpose as a motivation for sharing information on social media.

Other studies have found that the main motivation for fake news production and dissemination is commercial (Hirst, 2017; Marwick & Lewis, 2017), with creation and dissemination of misinformation used to boost traffic to an online site and increase advertising revenue (Allcott & Gentzkow, 2017). This phenomenon was witnessed during the 2016 US elections when a group of teenagers in Macedonia created fake news for economic gain. Sharing pro-Trump content, even when it contained falsehoods, generated online traffic that helped them to make money through advertisements (Marwick & Lewis, 2017).

While the above review shows that there have been numerous studies of audiences' reception of, and motivations for sharing, fake news, few studies in African settings have focused on how people in certain narrow demographics receive, and decide whether to share, such information. This study fills this research gap, in the Nigerian context, with a focus on the consumption and sharing behaviours of university-educated individuals in two urban settings.

3. Research design

This study adopted a qualitative approach, using focus groups with 60 individuals (all of whom were university graduates) aged 25 and above living in Umuahia and Lagos—two metropolitan cities in the south-eastern and south-western parts of Nigeria, respectively. The decision to focus on these two cities was based on their socioeconomic significance. Umuahia is one of the major commercial hubs of Nigeria's South East Region, and Lagos is the largest hub for the corporate and entrepreneurial sector in Nigeria. Both cities are thus strategically positioned and have residents drawn from various parts of the country.

Focus group discussion was chosen as an appropriate data collection method because it suits research that seeks to explore complex issues and to collect in-depth data at minimal cost (Brennen, 2012; Carey, 1994). A total of ten focus groups were conducted—five in each city—between August and October 2020. Each consisted of six participants. The study participants were gathered through a snowball technique where initial reliable contacts generated further contacts. Seven of the 10 focus groups had a 50:50 gender ratio (three males and three females) and the gender ratios of the other three groups were 40:60 for the males and females, respectively.

Each focus group lasted for approximately 90 minutes, and explored participants' reception of, and motivations for sharing (or not sharing), fake news on COVID-19 issues (see Appendix for the focus group discussion protocol). The focus group discussions explored questions such as: how media users would define fake news on COVID-19; how they identified false stories on COVID-19 when they saw them; the forms of media they saw as most likely to carry false or accurate information on COVID-19; and their views on how the media reported the pandemic.

Participants were also asked what their reactions would be if they saw certain instances of fake news on their social media timelines; whether they would share such stories; and their motivations for sharing or not sharing. To elicit detailed responses, participants in each focus group were presented with four stimuli. The stimuli were screenshots of online news headlines and lead paragraphs on COVID-19 identified as fake by Africa Check, an independent fact-checking organisation, headquartered in Johannesburg (Cox, 2013). The participants were not informed beforehand that the stimuli had been confirmed as fake news by Africa Check. This was to allow them to make sense of the stimuli on their own. The four stimuli are presented in the Appendix. Stimulus 1 showed a fake news item where former Nigerian president, Olusegun Obasanjo was quoted as saying that there was no COVID-19 in Nigeria (Africa Check, 2020a). Stimulus 2 showed a fake news item where the Sultan of Sokoto argued that the first case of COVID-19 in Nigeria (an Italian national) was faked by an actor (Africa Check, 2020b). Stimulus 3 showed a fake news item where garlic was presented as a cure for COVID-19 (Africa Check, 2020c). In Stimulus 4, a false record of COVID-19 cases in Nigeria was presented (Africa Check, 2020d).

The researcher who organised the focus groups (Ononiwu), or a research assistant trained for this purpose, mediated and audio-recorded each discussion session. Participants were given consent forms and information sheets before each of the focus group meetings commenced, and they were given the opportunity to ask questions about their involvement in the study. Before participants agreed to participate, and signed the consent forms, they were informed about the study's objective and significance, as well as the methodology and how the qualitative data would be used.

The qualitative data from the 10 focus groups was transcribed word-for-word and analysed thematically. Before the thematic coding, we listened to each audio recording and double-checked the transcripts for accuracy. We also took note of the responses' frequency, context, and specificity. Within each focus group, this method allowed for the identification of patterns, themes, and contradictions. In addition, quotations that best reflected the primary topics were chosen as part of the data analysis below. Two academic colleagues read the data independently and recognised themes that were similar to the ones we identified, ensuring the study's validity. We limited access to the data and provided secure data storage to guarantee confidentiality.

In addition to taking part in a focus group discussion, each participant was requested to indicate their gender, age, highest educational qualification, and occupation (as summarised in Table 1).

Table 1: Participants' demographic information

Variables		Frequency	%
Gender	Female	37	61.7%
	Male	23	38.3%
	Total	60	100%
Age bracket	25-30 years	21	35%
	31-39 years	18	30%
	40-49 years	11	18.3%
	50-55 years	10	16.7%
	Total	60	100%
Highest educational qualification	First degree	48	80%
	Postgraduate qualification	12	20%
	Total	60	100%
Occupation	Civil servants	12	20%
	Businesspeople	11	18.3%
	Teachers	10	16%
	Others	8	13.3%
	Journalists	8	13.3%
	Bankers	6	10%
	Postgraduate students	5	8.3%
	Total	60	100%

4. Findings and analysis

In this section, the results of the analysis of the qualitative data are discussed under two broad themes:

- reception of fake news on COVID-19; and
- motivations for sharing, or not sharing, fake news on COVID-19.

Reception of fake news on COVID-19

Participants in eight of the 10 focus groups claimed that they knew news was likely to be fake when it was going viral on social media and it was dramatic, exaggerated, and illogical when juxtaposed with well-known events. In the words of a participant in one of the Lagos focus groups:

> If it is circulating widely on social media, then it is likely to be fake news. Fake news on COVID-19 has elements that make it appealing and therefore likely to spread quickly. Fake news is sensational. It is also exaggerated. How can someone tell you that 1,000 persons are infected with the virus but when you look around, you cannot find anyone you know who has the virus? (male businessperson, age 44, Lagos).

According to another Lagos participant:

> We all know that the number of persons peddled online is far more than the actual number infected by the virus. COVID-19 does not affect us here as much as it affects people from other countries, especially in the West. The figures we see online are from fake news. Where is this 472 confirmed cases coming from [referring to Stimulus 4]? (male civil servant, age 41, Lagos).

Participants in six of the 10 focus groups (two in Lagos, four in Umuahia) saw social media as a natural home for fake news on COVID-19, saying that most COVID-19 information that was disseminated virally on WhatsApp and Facebook was fake. In the words of one of the Umuahia participants:

> News {items] on WhatsApp and Facebook are the major culprits. You can hardly see information on these platforms that is not fake. Credible news channels seldom use these platforms effectively. What people do is pick information from somewhere, modify it and spread [it] mostly on Facebook to attract more engagement for their social media pages (female teacher, age 33, Umuahia).

According to one of the Lagos participants:

> Fake news on COVID-19, as spread on social media, is designed to draw enormous public attention. This is why it spreads every quickly. It is like gossip, it is exciting. For instance, when it was alleged that cow urine could cure COVID-19, you can tell it is fake. Cow urine? Does it make common sense? Then look at this one that says garlic cures COVID-19 [referring to Stimulus 3]. It is unsubstantiated. People who post these things just want to cause a stir (male banker, age 30, Lagos).

The foregoing responses point to the widely held view among the respondents that there was a great deal of deliberateness in the development and posting of COVID-related fake news—deliberateness that has also been identified by Gelfert (2018).

Participants in six of the focus groups (four in Lagos, two in Umuahia) stated that fake news on COVID-19 was often identifiable through its use of information that was contrary to what had been presented by media entities that the participants described as credible, such as BBC and CNN, or contrary to information provided by health bodies such as the WHO:

> Fake news on COVID-19 always presents what is different from what reputable organisations like WHO says (female businessperson, age 25, Lagos).

> It presents information contrary to what the reliable news outlets are presenting. When I talk about reliable news outlets, I am talking about BBC and CNN (male civil servant, age 30, Umuahia).

> If a particular news [item] is showing something different from what you saw somewhere, then there is something suspicious about it. This is quite different from when a news source tries to get another angle of the same story. What I mean here is that fake news is outrageously different from what other [more credible] sources are saying (male civil servant, age 38, Umuahia).

> Health organisations such as the World Health Organisation and the NCDC [Nigeria Centre for Disease Control and Prevention] are giving us the factual figures. Credible news organisations are relying on NCDC for factual information, especially figures on the number of new cases and discharged people. However, fake news does not present these facts. You see all sorts of figures [from non-credible sources] (male postgraduate student, age 31, Lagos).

Two of the focus group participants in Lagos identified the following additional elements that helped them to distinguish between fake and credible COVID-related news:

> There are a lot of spelling and grammatical errors in fake news about the virus. Those that write this fake news are not professional journalists. Their intention is to create things that can spread quickly on social media, so they come up with all sorts of things […]. The news is usually scattered. […] There is little or no editing. This is quite different from what you would expect from an established media outlet (male postgraduate student, age 29, Lagos).

> It is easy to spot. Fake news on COVID-19 is sometimes written haphazardly. You cannot trace it to any source. The writers are usually unknown. You know when you read the newspaper, you can see the name and contact address of the person that wrote a particular story. But this is not the case in fake news. You just see the story with no name, no by-line. You cannot do that in ethical journalism. It is not our standard. Fake news also lacks accuracy. The content is really questionable. Look at this [referring to Stimulus 2]. I do not think the Sultan can say something like that. He cannot use words like "let me say the truth and die" (female journalist, age 54, Lagos).

At the same time, some participants in one of the focus groups in Umuahia were of the view that it was not possible to draw a line between fake news and factual news on COVID-19:

> Every news basically looks the same to me. There is no way to know which one is fake, unless there is some other thing you know behind what is presented in the news or on a personal level. It is a difficult one. News na news ["all news is the same"] (female businessperson, age 50, Umuahia).

> I do not even know. I cannot tell which one is fake and which one is not. All sorts of things are just flying around (male civil servant, age 31, Umuahia).

> You cannot tell the real one from the fake one. For me, everything is fake (female civil servant, age 40, Umuahia).

Distrust of the Nigerian news ecosystem

Participants in two focus groups (one in Lagos, one in Umuahia) spoke of a lack of trust in the Nigerian news ecosystem, whether on social media platforms or on the platforms of Nigerian media outlets. When asked which type of media was more likely to carry factual news about COVID, a participant in Lagos said:

> I cannot really say as far as Nigeria is concerned because everything here is fake. If you really want something factual, you can rely on international media (female civil servant, age 40, Lagos).

In the words of another participant:

> There are a lot of manipulations everywhere. The Nigerian media is filled with manipulated stories with different motives and interests (female banker, age 30, Umuahia).

One participant pointed particularly at the government-owned media as the major source of fake news on COVID-19:

> I think it is government-owned media. I do not really trust them. They are working for the government and come up with a lot of things to suit the government's agenda (male banker, age 30, Umuahia).

According to another participant:

> For me, I really don't think any of this news [is] true. There is something manipulated in all of them. This is how I feel. I think even the ones coming from the government are manipulated for one purpose or the other. Some things are not said exactly the way they are by the government and other stakeholders (female banker, age 30, Lagos).

These sentiments resonate with the findings of other studies on general distrust of, and cynicism towards, all forms of news media (Chadwick & Vaccari, 2019; Madrid-Morales et al., 2021; Wagner & Boczkowski, 2019; Wasserman & Madrid-Morales, 2019), with such cynicism found, in some cases, to make people more likely to spread misinformation.

COVID-19 denialism

The participants in one of the focus groups in Umuahia expressed doubts about the seriousness—and in some cases even the existence—of the virus:

> Most of the news [items] in this country on COVID-19 are fake because they seem to present the virus as very serious. Personally, I do not feel the virus is very serious in this part of the world. We all know that it does not kill an African man. I think the government is just giving figures about cases of infection and death just to get money from international organisations (male civil servant, age 31, Umuahia).

> I think most news on COVID-19 is fake because I know that the virus does not exist. In the first years of HIV, I saw HIV patients with my eyes. It is real when you see someone that has it. But for COVID-19, as far as Nigeria is concerned, it is all fake (female civil servant, age 40, Umuahia).

Motivations for sharing, or not sharing, fake news on COVID-19

Desire to warn of the dangers of fake news

Participants in three of the focus groups (two in Umuahia, one in Lagos) spoke of sharing fake news but at the same time making people aware that it was fake, thus seeking to perform a public service. In the words of a Lagos participant:

> Yes I would share with a caption explaining the dangers of fake news on COVID-19 and air my views about the situation from the point of knowledge about the facts (female dentist, age 50, Lagos).

An Umuahia participant who identified as a journalist explained his actions in this way:

> I share fake news as background to my own factual information on the dangers of COVID-19 and how people can take precaution. Since I am really interested in making people aware of the possible dangers of the virus, then I have to attach something factual just like I said earlier. Sharing the fake news as it is cannot create awareness on the possible dangers of the virus without attaching some reliable information before sharing. For instance, when there was a fake news that chloroquine could cure the virus, I shared the news and added some facts as caption. I then concluded by encouraging people to present themselves to health facilities each time they noticed corona symptoms (male journalist, age 41, Umuahia).

These findings appear to add nuance to the findings of Chakrabarti et al. (2018) on sharing information based on a sense of civic duty. Chakrabarti et al. (2018) found that a sense of civic duty led some users of social media to share warnings of imminent danger regardless of their sense of the veracity (or not) of the warnings. However, in the present study, the participants who engaged in "fact addition" to make an item appropriate for sharing were practising a much more active, sophisticated, and ethical approach to civic duty.

Concern, or desire to generate awareness

Another stated motivation for sharing was general concern about the virus. For example, one Umuahia participant stated that she shared or retweeted news on COVID-19, without being concerned whether the news was fake or not, because of her concern about the virus:

> I think I share some of these items because the situation worries me. The daily increase in the number of cases and deaths really makes me worried. Generally, I am interested in what is going on, so I just share. Besides, like I said before, every news basically looks the same to me, so I share whatever that is of interest to me (female businessperson, age 50, Umuahia).

For another participant, in one of the Lagos focus groups, it was acceptable to share any COVID-19 news that seemed to be somewhat accurate, and not necessarily fully accurate, because, he felt, even partially accurate information could be helpful in generating awareness of the need to take precautions:

> I may share if I feel it is true to an extent. At least for people to know what is going on and know how to take care of themselves. It is better people just take precaution so if the news is about taking precaution generally, even if it is not 100%, I may consider sharing (male model, age 36, Lagos).

This quotation and the one preceding it both resonate quite directly with the findings of Chakrabarti et al. (2018) on many individuals' tendency to share danger warnings with no, or limited, concern about the veracity of the warnings.

Commercial gain

Two participants in one of the Lagos focus groups, both bloggers, said that they intentionally shared fake news when they felt it could result in commercial gain. This echoes the findings on the commercial motivation for disseminating fake news as outlined in Hirst (2017) and Marwick and Lewis (2017). Referring to the four stimulus fake news items (see Appendix) presented during the focus group discussion, one of the two participants who shared fake news, and who identified the four stimulus items as all being fake, said:

> I aggressively share such stories to get readership for my blog. People will find it interesting so I share. I would share to get traffic for my blog and social media account. That's one of the things I do for a living, so I have to sustain my blogs (female entrepreneur/blogger, age 25, Lagos).

While referring to Stimulus 1, which he identified as fake news, the other Lagos participant who mentioned a commercial motivation stated as follows:

> I am blogger. I focus on anything that can bring engagement. I want more, views, comments, shares, and likes. I want more subscriptions, and you know that people are attracted to things that look odd. The more engagement I have, the more money I get. People want things that are catchy and interesting (male blogger, age 25, Lagos).

Unwilling to share, due to health risks

Several participants were determined not to share fake news on COVID-19. This sentiment was reflected in these words from one of the Lagos participants who identified as a journalist:

> I won't share once I feel it is fake. Anything that is not true is not likely to be of health benefit to anybody. COVID-19 is a serious health issue that we cannot afford to joke with (male journalist, age 41, Lagos).

5. Conclusions

This study explored how university-educated Nigerians in two urban centres responded to COVID-related social media fake news, and the extent to which they shared such news. Participants' sense-making practices with regard to fake news on COVID-19 were relatively varied. However, the majority of the participants shared similar impressions and perspectives on the features and spread of false information about the viral disease. Social media virality was widely viewed as synonymous with fake news, due to the dramatic, exaggerated, and sometimes illogical nature of such information, especially when placed alongside other well-known factual events. Perhaps due to their level of education, many participants demonstrated a notable level of literacy in identifying fake news. This seems to suggest that university-educated Nigerians, much like the university students in six African countries in the study conducted by Madrid-Morales et al. (2021), demonstrate some behaviours that point to their competence in dealing with false information or questionable content.

Many participants identified social media as being a natural and frequent carrier of fake news, while at the same time many also had low levels of trust in the information provided by the legacy Nigerian news media, especially government-owned platforms. This finding resonates with other studies that have found widespread distrust of the entire media ecosystem.

The motivations for sharing or not sharing fake news on COVID-19 were wide-ranging and quite nuanced. Motivations for sharing news known to be potentially fake, or at least partially fake, included a sense of concern about the pandemic and/or a desire to create awareness of the dangers of the virus. Two respondents, both bloggers, said that they shared news that they knew was fake when they felt that the sharing would result in commercial gain. Other respondents spoke of sharing news known to be fake but adding information pointing to the fact that it was fake, so as to warn people about the dangers of fake news. It may be concluded that the individuals who felt compelled to share fake news about COVID-19 (either for commercial gain or in order to warn people of the dangers) were acting with a level of intentionality similar to that shown by people who create the fake news in the first place. Finally, some participants refused to share COVID-19 news under any circumstances. For this group of individuals, the belief that spreading false information about COVID-19 would have actual or potential negative effects on the public prevailed over any potential motivation for sharing.

The findings from this study add nuance to understanding the actions of epistemologically robust people, as examined in the research by Chua and Banerjee (2017). As detailed earlier in this article, Chua and Banerjee (2017) studied the impact of epistemic belief on sharing, or not sharing, health rumours (i.e., potential misinformation), finding that epistemologically robust people were less likely than epistemologically naïve people to share online health rumours. However, findings

from the present study suggest that epistemologically robust people can display a wide range of behaviours in respect of fake news or potential fake news. For example, some of the participants who showed signs of being epistemologically robust—i.e., they understood that verifiable knowledge can be difficult to acquire—were willing to share potentially fake (or partially fake) news, because of their concern about the virus or because of a belief that sharing even unreliable information could help to make people more vigilant.

This study was not without limitations. We focused on only two Nigerian cities, on only one news topic (COVID-19), and only on participants with university degrees, which means that the sense-making and sharing behaviour with regard to fake news of a significant proportion of Nigerians was not explored. Future research into the reception of, and motivations for sharing or not sharing, fake news in Nigeria could focus on different demographics, in different locales, and on different news topics.

References

Adegoke, Y., & BBC Africa Eye. (2018). *Like. Share. Kill: Nigerian police say false information on Facebook is killing people*. https://www.bbc.co.uk/news/resources/idt-sh/nigeria_fake_news

Africa Check. (2020a). No evidence Obasanjo said there was no coronavirus in Nigeria. https://africacheck.org/fact-checks/meta-programme-fact-checks/no-evidence-obasanjo-said-there-was-no-coronavirus-nigeria

Africa Check. (2020b). No, Sultan of Sokoto didn't say Nigeria's first Covid-19 case was faked.https://africacheck.org/fact-checks/meta-programme-fact-checks/no-sultan-sokoto-didnt-say-nigerias-first-covid-19-case-was-faked

Africa Check. (2020c). No, garlic doesn't cure coronavirus. Get Covid-19 facts only from experts.https://africacheck.org/fact-checks/meta-programme-fact-checks/no-garlic-doesnt-cure-coronavirus-get-covid-19-facts-only

Africa Check. (2020d). Screenshot of Lassa fever cases, not Covid-19 numbers in Nigeria. https://africacheck.org/fact-checks/meta-programme-fact-checks/no-screenshot-lassa-fever-cases-not-covid-19-numbers-nigeria

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Allgaier, A., & Svalastog, A. N. (2015). The communication aspects of the Ebola virus disease outbreak in Western Africa – do we need to counter one, two, or many epidemics? *Croatian Medical Journal*, *56*(5), 496–499. https://doi.org/10.3325/cmj.2015.56.496

Anderson, P. (2019, May 21). Tackling fake news: The case of Nigeria. Italian Institute for International Political Studies. https://www.ispionline.it/it/pubblicazione/tackling-fake-news-case-nigeria-23151

Aquino, F., Donzelli, G., De Franco, E., Privitera, G., Lopalco, P. L., & Carducci, A. (2017). The web and public confidence in MMR vaccination in Italy. *Vaccine*, *35*(35, Pt B), 4494–4498. https://doi.org/10.1016/j.vaccine.2017.07.029

Bigman, C. A., Smith, M. A, Williamson, L. D., Planey, A. M., & Smith, S. M. (2019). Selective sharing on social media: Examining the effects of disparate racial impact frames on intentions to retransmit news stories among US college students. *New Media & Society*, *21*(11/12), 2691–2709. https://doi.org/10.1177/1461444819856574

Brennen, B. S. (2012). *Qualitative research methods for media studies*. Routledge. https://doi.org/10.4324/9780203086490

Carey, M. A. (1994). The group effect in focus groups: Planning, implementing and interpreting focus group research. In J. Morse (Ed.), *Critical issues in qualitative research* (pp. 225–241). SAGE.

Chakrabarti, S., Rooney, W. C., & Kweon, M. (2018). *Verification, duty, credibility: Fake news and ordinary citizens in Kenya and Nigeria.* BBC. https://downloads.bbc.co.uk/mediacentre/bbc-fake-news-research-paper-nigeria-kenya.pdf

Chadwick, A., & Vaccari, C. (2019). *News sharing on UK social media: Misinformation, disinformation, and correction.* Online Civic Culture Centre, Loughborough University.

Choy, M., & Chong, M. (2018). Seeing through misinformation: A framework for identifying online fake news. https://doi.org/10.48550/arXiv.1804.03508

Chua, A. Y. K., & Banerjee, S. (2017). To share or not to share: The role of epistemic belief in online health rumors. *International Journal of Medical Informatics*, *108*, 36–41. https://doi.org/10.1016/j.ijmedinf.2017.08.010

Cox, P. (2013). Fledgling website brings fact checking to South Africa. *VOA News*. https://www.voanews.com/a/fledgling-webiste-brings-fact-checking-to-south-africa/1725136.html

Dafonte-Gómez, A. (2018). The key elements of viral advertising: From motivation to emotion in the most shared videos. *Comunicar Journal*, *43*. https://doi.org/10.3916/C43-2014-20

Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, *38*(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

Hirst, M. (2017). Towards a political economy of fake news. *The Political Economy of Communication*, *5*(2), 82–94.

Lara-Navarra, P., Falciani, H., Sánchez-Pérez, E. A., & Ferrer-Sapena, A. (2020). Information management in healthcare and environment: Towards an automatic system for fake news detection. *International Journal of Environmental Research and Public Health*, *17*(3), 1066, 1–12. https://doi.org/10.3390/ijerph17031066

Lilleker, D. (2018). Politics in a post-truth era. *International Journal of Media and Cultural Politics*, *14*(3), 277—282.

Madrid-Morales, D., Wasserman, H., Gondwe, G., Ndlovu, K., Sikanku, E., Tully, M., Umejei, E., & Uzuegbunam, C. (2021). Motivations for sharing misinformation: A comparative study in six Sub-Saharan African countries. *International Journal of Communication*, *15*, 1200–1219.

Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society. https://datasociety.net/library/media-manipulation-and-disinfo-online/

Nyamnjoh, F. B. (2005). *Africa's media, democracy, and the politics of belonging*. Zed Books.

Ogwezzy-Ndisika, A. O., Amakoh, K. O., Ajibade, O., Lawal, T. O., & Faustino, B. A. (2023). Fake news and general elections in Nigeria: Fighting the canary in a digital coal mine. In U. S. Akpan (Ed.), *Nigerian media industries in the era of globalization* (pp. 61–72). Rowman & Littlefield.

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (2023). The "need for chaos" and motivations to share hostile political rumors. *American Political Science Review, FirstView*, pp. 1–20. https://doi.org/10.1017/S0003055422001447

Sandikci, O. (1998). Images of women in advertising: A critical-cultural perspective. In B. G. Englis & A. Olofsson (Eds.), *E – European Advances in Consumer Research, Volume 3* (pp. 76–81). Association for Consumer Research.

Soltaninejad, K. (2020). Methanol mass poisoning outbreak: A consequence of COVID-19 pandemic and misleading messages on social media. *International Journal of Occupational and Environmental Medicine*, *11*(3), 148–150. https://doi.org/10.34172/ijoem.2020.1983

Tully, M. (2021). Everyday news use and misinformation in Kenya. *Digital Journalism*, *10*(1), 109–127. https://doi.org/10.1080/21670811.2021.1912625

Uwalaka, T. (2022). "Abba Kyari did not die of coronavirus": Social media and fake news during a global pandemic in Nigeria. *Media International Australia*, online pre-print. https://doi.org/10.1177/1329878X221101216

Uzuegbunam, C. E. (2020). A critical analysis of transgressive user-generated images and memes and their portrayal of dominant political discourses during Nigeria's 2015 general elections. In M. Ndlela & W. Mano (Eds.), *Social media and elections in Africa*, *Volume 2* (pp. 223–243). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-32682-1_12

Wagner, M. C., & Boczkowski, P. J. (2019). The reception of fake news: The interpretations and practices that shape the consumption of perceived misinformation. *Digital Journalism*, *7*(3), 1–16. https://doi.org/10.1080/21670811.2019.1653208

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, *240*, 112552, 1–12. https://doi.org/10.1016/j.socscimed.2019.112552

Wardle, C. (2019). *Understanding information disorder*. First Draft. https://firstdraftnews.org/long-form-article/understanding-information-disorder

Wasserman, H., & Madrid-Morales, D. (2019). An exploratory study of "fake news" and media trust in Kenya, Nigeria and South Africa. *African Journalism Studies*, *40*(1), 107–123. https://doi.org/10.1080/23743670.2019.1627230

Welcome, M. O. (2011). The Nigerian health care system: Need for integrating adequate medical intelligence and surveillance systems. *Journal of Pharmacy and Bioallied Sciences*, *3*(4), 470–478. https://doi.org/10.4103/0975-7406.90100

Appendix: Focus group discussion protocol

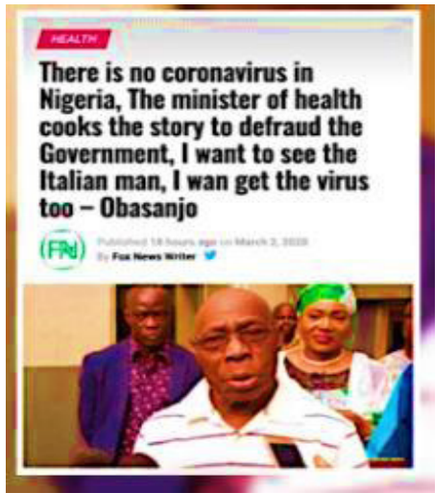
Date……..
Location……..
Team member……..
Note taker……
Observer…….
Duration of discussion……..
Start……
End…….

INTRODUCTIONS
First, I'd like to ask each of you to introduce yourself. A basic introduction should be enough. What is your name? Where are you from? What are you studying or doing for a living?

RECEPTION OF FAKE NEWS ON COVID-19
How do you know news on COVID-19 is fake when you see it?
How do you know news on COVID-19 is factual or true when you see it?
Which type of media is more likely to carry fake news on COVID-19?
Which type of media is more likely to carry factual news on COVID-19?
What can you say about how Nigerian media reported news on COVID-19 during COVID-19?
Which news media do you really on for news on COVID-19?

STIMULI #1-4: Motivations for sharing fake news on COVID-19
Next, I'd like you to have a look at these social media posts [show each stimulus on a large screen or mobile device].

Stimulus 1

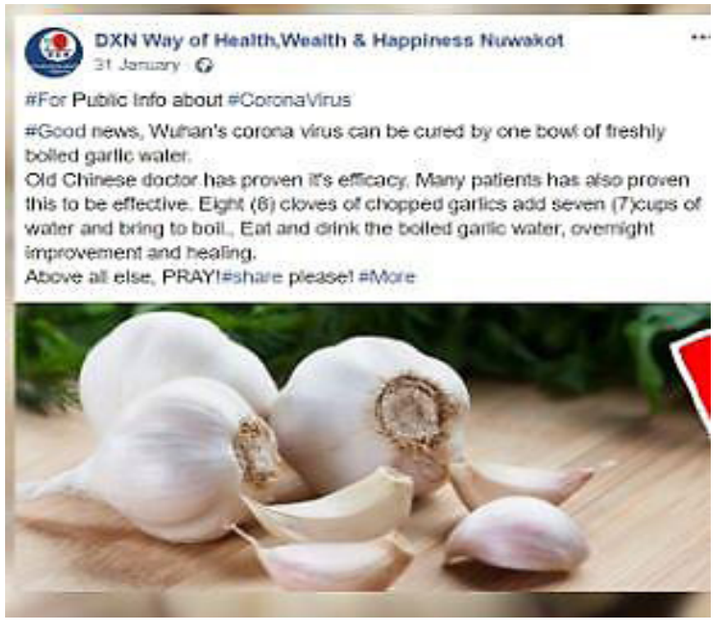


Source: Africa Check (2020a)

Stimulus 2

Source: Africa Check (2020b)

Stimulus 3



Source: Africa Check (2020c)

Stimulus 4

Source: Africa Check (2020d)

Questions (for each stimulus)

What would be your first reaction if you saw this on your Twitter or Facebook timeline?
Has anybody you know, maybe a friend or a relative, ever shared with you content similar to this? What did you do?

Would you consider sharing this post? Why or why not?

Prompts for additional reasons [*use if these reasons haven't been mentioned*]

Do you feel that putting this up on Facebook or Twitter would get you more likes or retweets? Why do you think this would be the case? [motivation: social currency / social media recognition]

Would you share these to make people aware of possible dangers? Even if you thought it wasn't true? [motivation: civic duty]

Does anybody here feel that they have a moral obligation or that it is their right to share this kind of information? [motivation: obligation, right]

How many of you would check with a more established news source before sharing? How often do you do this?
Do your friends or relatives share news on coronavirus? Do you ask them not to do it?
Can you recall sharing a story on COVID-19 that you later found out was not fully accurate?

Why did you share it?

Did something happen after you shared it? For example, did you eventually take it down or did you leave it? Why did you take it down / why did you leave it?

Did someone correct you, or ask you to take it down?

Has anyone deliberately shared news that you knew was completely made up? If yes, why did you share it?

OPTIONAL

How much of a problem do you think misinformation and fake news on COVID-19 are?
What do you usually do when somebody shares news on COVID-19 that you know is made up? What is your reaction?
How often do you use fact-checking websites? Do you know any that are reliable?

WRAP-UP

[This last section is meant to provide a quick summary of the discussion, to make sure participants agree with the key takeaways identified by the discussion leader. These questions should help people raise additional points they were unable to make during the discussion.]

Possible questions:
Today you have covered the following topics [provide a 3-to-5-point summary of the discussion] Do you feel this is an adequate summary?
Have we missed anything? Would you like to add one last thing?

Thank you all for your time. As explained before we started, today's discussion was aimed at gathering qualitative data for a study which explores the reception, and motivations for sharing, fake news on COVID-19. If somebody wants to know about the findings, please let me know and we'll be happy to share them with you.

With this we have come to the end of this focus group discussion. Thank you all for participating.

Competition regulation for digital markets: The South African experience

Simphiwe Gumede
Legal Counsel, Competition Commission South Africa, Pretoria
https://orcid.org/0009-0002-0158-2455

Phathutshedzo Manenzhe
Analyst, Competition Commission South Africa, Pretoria
https://orcid.org/0009-0007-8781-112X

Abstract
The study examines the recent experiences of South Africa's competition authorities in engaging with competition matters in the country's digital markets. Specifically, the authors examine engagements by the Competition Commission South Africa (CCSA), the Competition Tribunal of South Africa, and the Competition Appeal Court (CAC) with three regulatory elements: (1) mergers, examined through the *MIH and WeBuyCars* and *Google and Fitbit* cases; (2) abuse of dominance, examined through the *GovChat v Facebook* case; and (3) cartel conduct, examined through the *Competition Commission v Bank of America Merrill Lynch International Limited & Others* case. In reviewing the decisions made in these cases, the authors highlight regulatory considerations that are coming to the fore in response to competition matters in digital markets.

Keywords
digital markets, competition, regulation, mergers, abuse of dominance, cartel conduct, South Africa

Acknowledgements
This article draws on content from the authors' paper presented to the 7th Annual Competition and Economic Regulation (ACER) Week conference, 15-16 September 2022 in Senga Bay, Malawi. ACER Week 2022 was convened by the COMESA Competition Commission, the Competition and Fair Trading Commission of Malawi, and the University of Johannesburg's Centre for Competition, Regulation and Economic Development (CCRED). *AJIC* is grateful for the reviewing support provided for this article by CCRED Research Fellow Sha'ista Goga, who is Director of Acacia Economics, Johannesburg.

DOI: https://doi.org/10.23962/ajic.i31.16013

Recommended citation
Gumede, S. & Manenzhe, P. (2023). Competition regulation for digital markets: The South African experience. *The African Journal of Information and Communication (AJIC), 31*, 1-19. https://doi.org/10.23962/ajic.i31.16013

1. Introduction

Since the introduction of internet connectivity, the digital economy has been growing rapidly throughout the world. It provides businesses in digital markets with an unprecedentedly large geographic reach with which to access customers and intermediaries. This digital economy has brought with it challenges that have necessitated appropriate competition regulation. Traditional competition tools, which have been successfully applied to prevent elevated levels of concentration and the establishment of monopolies within the traditional economy, have come under great scrutiny given the potential difference in the dynamics of digital markets (Maihaniemi, 2020). These tools appear to have fallen short in some respects, given the high concentration levels in digital markets, leading to calls for differences in approach when regulating certain elements of competition in the digital economy (Maihaniemi, 2020).

The economics that drive digital markets provide significant challenges for competition law and regulation (Ademuyiwa & Adeniran, 2020). These challenges include defining markets through the use of the SSNIP (small but significant non-transitory increase in price) test, as the price arrived at must, in most digital markets, also take into account the multi-sided nature of the market and thus the need to balance the effect that a price on one side of the market has on the other side of the market (OECD, 2018). For example, in the *Google LLC and Alphabet Inc v European Commission* case, the General Court acknowledged that competition can take place in terms of quality and innovation, and found that the small but significant non-transitory deterioration in quality (SSNDQ) test, rather than an SSNIP test, constituted relevant evidence for the purpose of defining a market (*Google LLC and Alphabet Inc v European Commission*, 2018). Furthermore, the traditional tools used

to establish market power (such as elasticities, diversion ratios, and market shares) now need to also take into account the effect on consumer behaviour on one side of the market, contrasted with the effect on the other side of the market (OECD, 2018). At the same time, evaluation of the effect of high market share on one side of the market is unlikely to assist in understanding why a digital firm is also able to command dominance on the other side of the market where it does not have high market share (OECD, 2018).

The provisions of the South African Competition Act No. 89 of 1998, as amended (the Act), and the decisions of the Competition Commission South Africa (CCSA), the Competition Tribunal of South Africa, and the Competition Appeal Court (CAC), have come into sharp focus with respect to their ability to effectively regulate competition in South Africa's digital markets. This article examines recent decisions by the CCSA, the Tribunal, and the CAC with respect to three areas of competition regulation that have taken on new complexities in the digital dispensation: (1) merger control, which we examine through consideration of the *MIH and WeBuyCars* and *Google and Fitbit* cases; (2) preventing abuse of dominance, which we examine through consideration of the *GovChat v Facebook* case; and (3) preventing cartel conduct, which we examine through consideration of the *Competition Commission v Bank of America Merrill Lynch International Limited & Others* case. In examining these cases, we seek to highlight some of the regulatory elements that have gained increased prominence in the digital dispensation.

2. Background: Competition in digital markets

Global context

The proliferation of digital platforms has caused significant disruptions to the traditional economy, through prioritisation of the growth of platforms, in the short and medium term, over profit-making (UNCTAD, 2019). Digital platforms undertake this growth by focusing on the collection of large amounts of user data and using this data to improve their algorithms for more effective attraction of new users to the platform.

A comparison (see Table 1) of the industries of the top 10 companies in the world (by market capitalisation) in 2009 with the industries of the top 10 companies in 2023 reveals just how far the digital economy has grown and increased its impact on our lives (PwC, 2023; UNCTAD, 2019).

Table 1: World's top 10 companies in terms of market capitalisation (2009 and 2023)

2009			2023		
Company	**Market capitalisation (USD)**	**Industry**	**Company**	**Market capitalisation (USD)**	**Industry**
Exxon Mobil	$337 billion	Oil and gas	**Apple**	**$2,609 billion**	**Technology**
Petro China	$287 billion	Oil and gas	**Microsoft**	**$2,146 billion**	**Technology**
Walmart	$204 billion	Consumer services	Aramco	$1,983 billion	Oil and gas
Industrial and Commercial Bank of China	$188 billion	Financial services	**Alphabet (Google)**	**$1,330 billion**	**Technology**
China Mobile	**$175 billion**	**Telecoms**	**Amazon.com**	**$1,058 billion**	**Technology**
Microsoft	**$163 billion**	**Technology**	**Nvidia**	**$685 billion**	**Technology**
AT&T	**$149 billion**	**Telecoms**	Bershire Hathaway	$676 billion	Financial services
Johnson and Johnson	$145 billion	Health care	Tesla	$659 billion	Consumer services
Royal Dutch Shell	$139 billion	Oil and gas	**Meta (Facebook)**	**$550 billion**	**Technology**
Procter and Gamble	$138 billion	Consumer goods	Visa	$464 billion	Financial services

Sources: PricewaterhouseCoopers (PwC) figures reported in UNCTAD (2019, pp. 2–3), and PwC (2023, p. 17)

As depicted in Table 1, in 2009 the top 10 companies included one technology company and two telecommunications companies. By 2023, the majority of the top 10 companies were technology-driven entities. The economic power of the "GAFAM" companies—Google (Alphabet), Apple, Facebook (Meta), Amazon, and Microsoft—achieved through intensive investment in their digital platforms and the deployment of user data collected via these platforms, is evident in their presence in the top 10 companies of 2023. These firms together provide digital infrastructure and software for, inter alia, applications, search engines, social networking sites, social media, e-commerce, marketplaces, personal computers, and cloud computing. These firms control the largest digital platforms—a position that enables them to dominate both sides of a market and leverage their dominance for their own benefit (US House of Representatives, 2022).

One characteristic of the digital economy is the two-sided or multi-sided nature of the digital platforms that operate in various markets (UNCTAD, 2019). In such markets, firms simultaneously interact with two or more separate customer groupings. These can include users, content providers, and advertisers (OECD, 2018). The

effects of what is done on one side of the market can also influence what happens on the other side(s). For example, a price increase for users may have a negative impact not just on the user side but also on the advertising side and the content creation side (OECD, 2018). The underlying business model of these firms is predicated on the collection of large amounts of data (Maihaniemi, 2020). On the one hand, digital platforms offer free services to consumers while gaining their attention and data, and, on the other, they monetise this attention and data by selling advertising opportunities to advertisers and allowing business users to sell goods and services to consumers through the platforms (ACCC, 2020; Maihaniemi, 2020). The consumer data that digital platforms collect is often seen as a commodity, valuable enough to influence big data companies' ability to "make decisions with economic impacts, environmental impacts or effects on health, education or society in general" (Coyle et al., 2020, as cited in UNCTAD, 2021, p. 5). The role of data is paramount for digital platforms because it often becomes the tool through which market power is established (UNCTAD, 2019).

The digital economy has brought with it some unique features that have facilitated high levels of market concentration. The control of big data, and use of machine-learning algorithms and analytics, generates enormous digital intelligence for the data controller. The high and increasing returns available through the use of big data entrenches the dominance of the data controllers. For instance, digital platform companies collect data on their users, and on their users' online behaviour, and sell the data to third parties who in turn use this data to develop their online businesses (Maihaniemi, 2020). Digital platform companies also use the data they collect for targeted/personalised marketing and to improve their consumer offerings, thus attracting more use on, and more users of, their platforms (Maihaniemi, 2020). The economic power of the collected data is further intensified by the digital platforms' low distribution costs, which facilitate global reach.

Today's most powerful digital platform companies have links to pioneering players in the digital economy, who continue to benefit from first-mover advantages. The pioneers were able to amass the first large collections of user data and to deploy this data (in the ways described above) in order to grow their digital platforms and market power, thus providing access to ever-larger pools of user data (UNCTAD, 2021). The largest digital platform companies have also been able to implement deliberate strategies to retain their early leadership, such as acquiring start-up entities that could potentially pose a competitive threat (UNCTAD, 2019) and imposing restrictive terms and conditions on business users that are not imposed on the businesses or business units within the platform company's conglomerate. Most of the core competition matters present in digital markets include elements of abuse of market power by large digital platforms—i.e., exploitation of their ability to behave, to an appreciable extent, in a manner independent from the behaviour of their competitors, customers, and/or consumers, usually resulting in anti-competitive effects.

Smaller digital platforms struggle to compete effectively with the large platforms. In addition to up-front investment costs, new entrants are faced with incumbent digital platforms that enjoy substantial economies of scale wherein the marginal cost of acquiring large numbers of additional users is close to zero. The incumbent platforms also benefit from substantial network effects. Business users seek strategies to grow their consumer base, and this rewards digital platforms which have strong network effects. The platforms' network effects allow businesses users to rapidly increase their user base, thus building business user dependence on the platforms. Such network effects can result in virtuous cycles where growing users on the business user side drives the growth of consumers on the other side, which then repeats itself.

A number of theories of harm are relevant to competition in digital markets. These include forced free-riding, which arises from digital platforms' use of business users' competitively sensitive information, concerning both sellers and consumers. Forced free-riding is defined by Shelanski (2013, p. 1699) as occurring "when a platform appropriates innovation by other firms that depend on the platform for access to consumers". In the e-commerce context, Khan (2017) explains that forced free-riding can, for instance, arise when a digital platform operator copies the design of popular goods sold by third-party retailers on its platform. Also, digital platform operators can easily identify which goods are bestsellers on their platforms and, potentially, favour their own products in advertising efforts and search rankings (OECD, 2018). The threat of free-riding on digital platforms is afflictive as it cannot be easily identified or proven.

Another practice identified and characterised as abusive or discriminatory leveraging is when a dominant digital platform that is active in numerous related markets leverages its dominance in one market for the benefit of its products in a related market (OECD, 2020). While the leveraging of dominance by a firm in one market to an adjacent market is not unique to digital markets, this type of conduct can be particularly egregious in digital markets where the dominant firm is a digital platform owner. The dominant firm can, through self-preferencing, treat business users and third-party users in a discriminatory manner and favour its own products or services over competing goods or services. The prevalence of this type of conduct in the digital economy is of concern because large platforms provide an important gateway into the broader market.

A more established issue identified in digital markets is tying or bundling, which refers to sales practices whereby customers (business users) are either required or incentivised to buy two or more distinct products as a combined sales package. Tying or bundling can harm competition through the exertion of market power from one market segment to another, thereby foreclosing the latter. For instance, where dominant platform operators offer multiple services—such as online marketplaces providing both retail listing and delivery services or price comparison—the platform

can engage in tying or bundling to foreclose rival sellers (OECD, 2018). However, it has also been found that tying or bundling practices may generate significant welfare-enhancing efficiencies, and therefore an outright prohibition may not be the most appropriate response (OECD, 2018).

Envelopment strategies have also been identified as potential causes of harm to competition in digital markets. Such strategies concern the imposition of unfair data collection terms by dominant firms. These terms allow dominant firms to collect the data of consumers and use it beyond the limited instances prescribed in the firms' privacy policies. These firms subsequently use this data to enter a new but related market with an overlapping user base, while entrenching their position in their original market (OECD, 2020). This form of strategy has also been referred to as privacy-tying (OECD, 2020).

Potential competitive harms relating to commissions, access conditions, and trading terms have also caused great concern in respect of digital markets. Several jurisdictions, including the US, the EU, Germany, and India, have recorded concerns regarding exploitative practices by large platforms in the form of high commission fees, unfair terms and conditions of access to platforms, and the treatment of smaller suppliers on digital platforms.

A clear understanding, by competition authorities, of how digital markets operate is imperative if such authorities are to be effective in achieving their fundamental purpose of ensuring competition. South Africa's approach to some of the challenges highlighted above will be considered next. Specifically, consideration will be given to specific cases—involving proposed mergers, potential abuse of dominance, and potential cartel conduct—in order to examine the functioning of South African competition regulation on, and the Act's ability to deal with, digital markets.

South African context

In September 2020, the CCSA published its strategic review, *Competition in the Digital Economy* (CCSA, 2020a). The report identifies four features of the digital economy in South Africa, namely: (1) the dynamic and innovative nature of digital markets; (2) concentration in digital markets, often created by first-mover advantages, data accumulation, and network effects; (3) the ease of entry in some secondary and tertiary levels of digital markets; and (4) the rapid pace of change (CCSA, 2020a). The report concludes that market inquiries are among the most effective mechanisms for addressing practices that may limit or prevent competition, and reduce barriers to entry, or expansion, in digital markets.

Based on the recommendation of the 2020 strategic review, the CCSA launched, in May 2021, an Online Intermediation Platforms Market Inquiry (OIPMI). The OIPMI Terms of Reference stated that the Inquiry stemmed from a "general

recognition that normal enforcement tools may be inadequate on their own to prevent initial market leaders from durably entrenching their position […]" (Department of Economic Development, 2021). Of specific concern were: the lack of competition enforcement tools to regulate data and data's ability to lead to market power; and the inability to define conduct which distorts and/or impedes competition and thus prosecute big data firms.

The initial scope of the OIPMI, which is nearing its conclusion as this article is being finalised in May 2023, included e-commerce marketplaces, online classifieds (in property and automotive), food delivery services, software application stores, as well as travel and accommodation services (CCSA, 2021). The OIPMI subsequently identified a need to include price comparator services in its scope, including insurance price comparator sites. In its July 2022 OIPMI *Provisional Summary Report*, the CCSA identified (see Table 2) the leading online intermediation platforms operating in eight South African platform categories that may require remedial action.

Table 2: Leading online intermediation platforms operating in South Africa

Platform category	Leading online intermediation platforms
Software app stores	Apple App store, Google Play store
eCommerce	Takealot
Property classifieds	Property24, Private Property
Automotive classifieds	Autotrader, Cars.co.za
Food delivery	Mr. Delivery, UberEats
Travel and accommodation	Booking.com
General search (as an input to platform competition)	Google Search
Insurance comparison sites	Hippo.co.za

Source: CCSA (2022b)

The identification of these platforms as leading was not premised on their dominance (as it is not necessary to establish dominance in a market inquiry), but rather on their exhibition of characteristics of a dominant firm in a particular market. The OIPMI is focused on market features that: inhibit platform competition and business user competition; lead to the exploitation of business users; and/or hinder the ability of small businesses and historically disadvantaged firms to participate in the markets (CCSA, 2022b).

We now turn to an examination of recent decisions taken by South African competition authorities in respect of three types of market behaviour—mergers, abuse of dominance, and cartel conduct—that have taken on significant new dimensions in the digital dispensation.

3. Mergers

Of the 87 proposed mergers in South African digital markets between 2011 and 2018, 82 were approved without conditions and the remaining five were approved with public interest conditions (CCSA, 2020a). As a consequence of the dynamic nature of digital markets and the difficulty in regulating them because of unmet merger thresholds, the CCSA was not able to impose any substantive conditions or prohibit any of these mergers (CCSA, 2020a). It is important to note that this data does not indicate the number of mergers that did not meet the merger notification thresholds. Section 13(3) of the Act allows the CCSA to require parties to a "small merger" to file their merger at the request of the CCSA—provided that the CCSA is of the view that the merger may result in a lessening of competition in a market or that the merger cannot be justified on public interest grounds. However, in some cases the small-merger threshold fails to capture high-potential-value "killer acquisitions", which occur when a big data firm acquires a small firm still at an early stage of its operations and prior to the small firm acquiring assets or generating turnover that is sufficient to meet the merger notification thresholds.

In December 2022, in response to concerns regarding such "killer acquisitions", the CCSA published revised Guidelines on Small Merger Notification. The Guidelines require that the CCSA be notified before the implementation of a small merger where: (1) the acquiring firm's turnover or asset value (irrespective of the turnover or asset value of the target firm) exceeds the large-merger threshold (currently set at ZAR6.6 billion); and one of two additional conditions is met: (a) the acquisition or investment in the target firm exceeds the intermediate-merger threshold (currently set at ZAR190 million); or (b) the effective valuation of the target firm exceeds the intermediate-merger threshold (CCSA, 2022c).

The 2018 amendments to the Act have increased the CCSA's powers of scrutiny when considering mergers, by allowing for the examination of previous mergers that either of the parties may have engaged in over a period determined by the CCSA. This affords the CCSA the ability to consider a firm's previous merger activity and the potential impact it may have in a proposed merger. This is important for the CCSA when assessing potential mergers in the digital economy, given the ease with which firms can make new acquisitions, and enter new markets, within a short space of time.

MIH eCommerce Holdings (Pty) Ltd (t/a OLX) and WeBuyCars

This proposed merger involved the acquisition, by MIH eCommerce (Pty) Ltd (MIH) t/a OLX South Africa (OLX), of 60% of the shares in WeBuyCars (Pty) Ltd (WeBuyCars). OLX is an online classified advertising platform that "carries advertisements for a broad range of goods and services for purchase and sale, including used cars" (*MIH and WeBuyCars*, 2020). OLX is a wholly owned subsidiary of South African technology and media holding company Naspers, which operates The Car Trader (Pty) Ltd, t/a AutoTrader, an online specialised classified advertising platform for the purchase and sale of used and new vehicles. WeBuyCars, on the other hand, operates as a guaranteed buyer and seller of used cars.

Following the CCSA's investigation of the proposed merger, the CCSA was required, in terms of section 14A of the Act, to refer the merger (because of its size) to the Tribunal, along with a recommendation as to whether it should be approved or prohibited. In this case, the CCSA recommended to the Tribunal that the proposed transaction be prohibited. The theories of harm examined by the Tribunal were: (1) unilateral effects, specifically the removal of a potential entrant (Frontier Care Group (FCG)) into the market; and (2) conglomerate or portfolio effects, specifically the likelihood that the merger would entrench Naspers' dominance in the market.

During the investigation, it emerged that Naspers had already acquired a controlling interest in FCG, a German-based online car-buying and -selling company that was understood be on the verge of entering the South African market to compete with WeBuyCars. FCG buys used vehicles from private individuals and from car rental companies, and then sells the used vehicles to dealers (it does not sell to individual consumers).

The Tribunal identified a narrow South African market for car-buying services, and pointed out that the merging parties' own strategic documents stated that "traditional dealers do not exert any meaningful constraint upon car-buying services such as WeBuyCars" (*MIH and WeBuyCars*, 2020). Evidence presented during the hearing, including a survey done by AutoTrader, confirmed that traditional car dealers were customers of WeBuyCars rather than competitors. The Tribunal pointed out that online car-buying services are different from traditional dealers because an online car-buying service (such as WeBuyCars) focuses on: (1) the provision of immediate cash for a high volume of used cars; (2) a strong reputation coupled with a strong marketing strategy; (3) an online platform that eliminates geographic constraints, allowing private sellers to approach dealers on the platform without having to travel from one dealer to another; (4) a more rapid and convenient method of selling a car than that provided by traditional offline dealers; and (5) a greater scale for buying and storing cars than traditional car dealers are able to achieve.

In assessing the likelihood of the removal of a potential entrant from the market, the Tribunal pointed out that in terms of the merging parties' strategic documents and those of FCG,[1] FCG was "poised to enter the market in South Africa but for the proposed merger" (*MIH and WeBuyCars*, 2020). The Tribunal further pointed out that FCG would have been a formidable entrant in South Africa, in competition with WeBuyCars, because it had been able to achieve success in other developing-world markets.

The Tribunal then examined the second theory of harm: conglomerate or portfolio effects. This theory of harm analysed whether, post-merger, Naspers' complementary businesses, including OLX and AutoTrader (as well as its broader portfolio of businesses, including its media and technology assets) could be leveraged to entrench WeBuyCars' dominance in the car-buying services market. The Tribunal found that WeBuyCars would be able to leverage OLX's market position in private listing offerings as customers would be able to receive a quote from WeBuyCars automatically when listing their vehicle on the OLX website.

With no competitor in the market to constrain WeBuyCars, the Tribunal concluded that the competition harm identified was likely to arise. Furthermore, the Tribunal highlighted that AutoTrader would be provided with "over 30 years of WeBuyCars' intricate data on both the cars and consumer behaviour" (*MIH and WeBuyCars*, 2020), and that this would enable it to set purchase and sale prices of vehicles coupled with an established dealership network to on-sell vehicles to.

The Tribunal prohibited the merger in 2020, and its decision was grounded in the identification of two crucial elements of digital markets: (1) the centrality of an online platform in the market; and (2) the increasing importance of data access in this market. The Tribunal's assessment of portfolio or conglomerate effects focused on the data that WeBuyCars would have access to post-merger and how this would effectively entrench WeBuyCars' dominant position in the market. Here we see a fundamental change from a traditional portfolio effects assessment, which would consider the products that the merged entity would have access to post-merger. When it comes to digital markets, the wealth of data that the merged entity would have access to has become a central assessment metric.

Google LLC and Fitbit Inc

This was classified as a small merger, due to the low turnover and asset value of Fitbit, which was filed at the request of the CCSA. Section 13(1)(a) of the South African Act states that "a party to a small merger is not required to notify the Competition Commission of that merger unless the Commission requires it to do so". The merger

1 These included FCG's public statements and email exchanges between WeBuyCars and OLX FCG.

was also notified in the EU, the UK, and Australia. The theories of harm considered by the CCSA were: (1) the removal of potential competition in the production and supply of wearable devices (fitness trackers and/or smart watches); (2) input foreclosure (leveraging of Google's dominance in the provision of operating systems (OS) for Android mobile devices into the market for the production and supply of wrist-worn wearable devices); (3) using Fitbit data or data collected from wrist-worn wearable devices to enter the market for the provision of digital health; and (4) preventing future competition in the provision of digital health.

In assessing these theories, the CCSA did not conclude on a relevant market. However, it did consider the national upstream market for: (1) the production and supply of OS for wrist-worn wearable devices; (2) the production and supply of OS for smart mobile devices; and (3) the national downstream market for the production and supply of wrist-worn wearable devices. The CCSA found that Google's entry into the market for the supply of wearable devices through this transaction would lead to the removal of Fitbit as a non-vertically-integrated competitor in the market. The Commission also found that this transaction would raise the barriers to entry in the market and would enhance Google's already existing data, which would allow Google to enhance its dominance in the advertising market. However, the CCSA found that it was unlikely that Google would be able to foreclose Fitbit's competitors, because Google's Wear OS was not a significant input in the production and supply of wrist-worn wearable devices.

The CCSA also found that ownership of Fitbit's existing health data, when combined with Google's individualised non-health data, could result in Google entering markets (and tipping them in its favour) for the provision of health and health insurance. Thus, in its own words, the CCSA "found that the proposed transaction is likely to result in a substantial prevention or lessening of competition" (CCSA, 2020b).

Accordingly, though the Commission approved the merger in 2020, it made the approval conditional on Google committing to several behavioural conditions for a period of 10 years, with the conditions to be monitored "by an independent Trustee who will have the necessary skills, competencies, and technical abilities to monitor these conditions" (CCSA, 2020b). The conditions comprised Google's commitments to:

- make access to the Android OS available for free, without discrimination, and with unchanged licence conditions, to competing makers of wrist-worn-wearables;
- keep Fitbit data separate from existing Google data, to not automatically use Fitbit data in any Google services, and to allow South African users to decide whether or not to allow storage of their "measured body data" in their Google or Fitbit accounts; and

- to allow third parties currently accessing users' Fitbit data to continue to have that access, without a charge from Google and provided the user gives consent (CCSA, 2020b).

This case demonstrates the utility of imposing conditions as a regulatory tool for regulating the conduct of firms operating in digital markets. The case also shows that, as with some traditional markets, mergers in digital markets can be assessed without a concluded market definition. The identification of the perimeters of competition were sufficient for the assessment of the transaction. It is also significant that this merger revealed that the effects of a merger in digital markets can have an impact on a number of other markets. In addition to its identification of the three aforementioned broad markets, the CCSA also identified the potential impact of the merger in adjacent markets, namely the markets for the provision of health, and health insurance. The CCSA's imposition of a 10-year monitoring period provides it with the opportunity to examine and consider the impact of the merger on these adjacent markets. Setting behavioural conditions and monitoring behavioural compliance are important mechanisms for competition authorities seeking to control and assess the impact of data collection and use in digital markets. This is due to such markets being dynamic, rapidly-evolving, and double- or multi-sided in nature.

4. Abuse of dominance

In its *Competition in the Digital Economy* report (CCSA, 2020a), the Commission highlights several challenges for abuse-of-dominance prevention in digital markets, including: (1) jurisdiction—due to the globalised nature of digital markets, firms that operate in them often have a limited presence in South Africa, and holding them to account may therefore be difficult; (2) meeting the evidentiary burden for effects (often challenging in digital markets); and (3) assessing market power with an increased focus on competitive relationships and strategies as opposed to market shares, and assessing new types of market power, such as "gatekeeper" market power. The challenges faced by competition authorities around the world in seeking to prevent abuse of dominance in digital markets have prompted specific focus in international fora (see, for example, OECD, 2020).

GovChat v Facebook

This was an application for interim relief brought, before the Competition Tribunal, by GovChat, a South African government online services platform. GovChat requested that the Tribunal prevent Facebook (the owner of WhatsApp) from "off-boarding" GovChat from WhatsApp Business (WhatsApp's paid business messaging platform). GovChat argued that WhatsApp is dominant and that its conduct amounted to a prohibited practice, namely a contravention of section 8(1)(d)(ii) and section 8(1)(c) of the Act. Section 8(1)(d)(ii) prohibits a dominant firm from "refusing to supply scarce goods or services to a competitor or customer when supplying those goods or services is economically feasible". Section 8(1)(c), in the

alternative, prohibits a dominant firm from engaging "in an exclusionary act, other than an act listed in paragraph (d), if the anti-competitive effect of that act outweighs its technological, efficiency or other pro-competitive gain".

Because this was an application for interim relief, the standard of proof for success was a *prima facie* standard. The Tribunal identified a market for over-the-top (OTT) messaging applications, in which WhatsApp was active, and distinguished the WhatsApp platform from other services (these included SMS, MMS, USSD) based on the fact that WhatsApp's users only required internet connection on a suitable phone and because the app was capable of sending and receiving a variety of media, including "photos, music, videos, voice memos, animated GIFs and even documents like MS Word or PDF files" (*GovChat v Facebook*, 2020, p. 10). Another important distinguishing feature that the Tribunal identified was WhatsApp's end-to-end encryption. Based on these distinguishing technological features, the Tribunal categorised WhatsApp in a narrowed OTT messaging apps market along with WeChat, Facebook Messenger, and Snapchat. The Tribunal established that WhatsApp was dominant in this market on the grounds that: (1) 89% of all internet users in South Africa between the ages of 16 and 64 reported having used WhatsApp; (2) at least 58% of all mobile phone users in South Africa had downloaded WhatsApp; (3) WhatsApp comes pre-loaded on almost all Android smartphones; and (4) mobile networks in South Africa offer WhatsApp data bundles (*GovChat v Facebook*, 2020).

In assessing the contravention (off-boarding) alleged by GovChat, the Tribunal observed that it is very difficult to duplicate OTT apps without extensive capital investment, and thus such apps fall within the meaning of "scarce goods or services". The Tribunal found, on a *prima facie* standard, that WhatsApp was selectively applying its terms and conditions in support of its own business service providers (BSPs), and that WhatsApp was rendering services directly to government departments and its direct approach to GovChat's government clients indicated that WhatsApp sought to foreclose GovChat from the market. The Tribunal further found that GovChat had established a *prima facie case* of anticompetitive conduct on the part of WhatsApp because of WhatsApp's threats to off-board GovChat from the WhatsApp Business platform in favour of its own BSPs.

After concluding that GovChat had *prima facie* met the requirements of section 8(1)(d)(ii), the Tribunal ordered interim relief. We are of the view that this case opened an important door in South Africa for the interpretation and application of abuse-of-dominance provisions to digital markets. By making the crucial finding that the WhatsApp platform was scarce and could not be easily duplicated, and by showing the effects of Facebook's conduct (through the selective application of its terms), this case broadened the application of section 8(1)(d)(ii) for application to digital markets, confirming that the abuse-of-dominance provisions of the Act are able to apply to more than just brick-and-mortar markets.

With respect to the Tribunal's application of the concept of scarcity, it can be argued that the reason why the Tribunal could apply this concept in this case was because of the popularity and the significant start-up costs involved in establishing a platform such as WhatsApp. It remains to be seen whether the same will be true in an instance where the start-up costs are not high.

5. Cartel conduct

The proliferation of algorithms as tools used by firms to participate in digital markets has provided firms with greater efficiencies in their efforts to compete. Algorithms have also facilitated advanced mechanisms for collusion (CCSA, 2020a). Globally, the use of big data by firms has been found to enable collusion where firms can share identical pricing algorithms and use real-time data analysis to monitor compliance with a collusive agreement. The multi-country or globalised nature of many digital platforms also means that the prosecution of cartels in a single country will often be met by jurisdictional challenges.

Competition Commission v Bank of America Merrill Lynch International Limited & Others

In February 2017, the CCSA referred to the Tribunal allegations of price-fixing and dividing markets, concerning the South African Rand, against 17 local and international banks operating in South Africa.[2] The CCSA alleged that the 17 banks had, in part, divided markets through the allocation of "customers in the USD/ZAR currency pair" (CCSA, 2020c, p. 1). The CCSA pointed out in its media statement that the investigation of this conduct stemmed from the identification of an agreement between the banks to collude, in 2007, on prices for "bids, offers and bid-offer spreads for the spot trades in relation to currency trading involving US Dollar / Rand currency pair" (CCSA, 2020c, p. 1). In investigating this alleged cartel, the CCSA had relied on traditional and familiar tools of investigation such as the identification of the agreement between the banks and the examination of online chats between bank employees held in chatrooms and on trading platforms.

This potential cartel had been detected in 2007 but only referred to the Tribunal in 2017, after a CCSA investigation that started in 2015. Following the CCSA's referral of this case, many of the internationally based banks raised jurisdictional challenges before the Tribunal. The Tribunal concluded that it did not have personal

2 The initial group of 17 banks comprised Bank of America Merrill Lynch International Limited, BNP Paribas, JP Morgan Chase & Co, JP Morgan Chase Bank NA, Investec Ltd, Standard New York Securities Inc, HSBC Bank Plc, Standard Chartered Bank, Credit Suisse Group, Standard Bank of South Africa Ltd, Commerzbank AG, Australia and New Zealand Banking Group Limited, Nomura International Plc, Macquarie Bank Limited, ABSA Bank Limited (ABSA), Barclays Capital Inc, and Barclays Bank plc. The number later increased to 28 when the CCSA, in June 2020, filed a new referral.

jurisdiction over the foreign-based banks. The Tribunal determined that both personal jurisdiction and subject-matter jurisdiction were necessary for it to have jurisdiction over the foreign-based banks (*Competition Commission v Bank of America Merrill Lynch International Limited & Others*, 2019). The Tribunal reasoned that section 3(1) of the Act established only subject-matter jurisdiction, with the lack of physical presence in South Africa meaning that no personal jurisdiction could be established.

The matter was ultimately appealed by the CCSA to the CAC, whose 2020 ruling in this case provided the CCSA with the opportunity to set out a clearer case against the foreign-based banks in respect of their conduct and its impact on South Africa. The CAC held that this opportunity would ensure that the CCSA was able to show that the Act and the Tribunal had jurisdiction over the foreign-based banks. The CAC, crucially, held that cartel conduct involving a foreign firm and impacting South Africa can be subject to the jurisdiction of the Act. The CAC further asserted that "courts should examine whether the forum which is sought to be employed has a real and substantial connection with the action; and whether the relevant connecting factors tie the action to the forum in question" (*Competition Commission v Bank of America Merrill Lynch International Limited & Others*, 2020, p. 22).

This assertion by the CAC was crucial in buttressing the point that anticompetitive conduct that has an effect on the South African economy would be in danger of falling outside of the scope of domestic enforcement if courts did not carefully examine and consider the merits of the case and jurisdiction (*Competition Commission v Bank of America Merrill Lynch International Limited & Others*, 2020). The CAC thus adopted a broader interpretation of section 3(1) of the Act than that which had been adopted by the Tribunal, which afforded the CCSA the opportunity to set out a clearer case against the foreign-based banks, particularly on the question of jurisdiction.

In March 2023, the Tribunal ruled that the CCSA had now satisfactorily set out the evidence of cartel conduct, and that the Tribunal had determined that it did indeed have jurisdiction to hear the case, on the grounds that "[t]he Respondents are accused of engaging in conduct considered the most egregious in competition law. Furthermore, the alleged conduct relates to fixing and manipulating the rand/dollar exchange rate, which has a central and crucial role in the South African economy" (Competition Tribunal, 2023). This ruling by the Tribunal, and the earlier ruling by the CAC, provide an important gateway for the CCSA to prosecute, where appropriate, conduct which involves firms that are based outside of South Africa. Given the global nature of digital markets, this is an important gateway.

We submit that this case reveals that the current legal framework is capable of prosecuting cartel activity even when digital firms located globally use quite novel means to collude. The CAC's wide interpretation of section 3(1) of the Act means that

digital platforms that provide services or goods in South Africa need not necessarily have a physical presence in South Africa for the Act's jurisdiction to apply. While this discussion relating to jurisdiction has taken place within the context of cartel conduct, it is important to point out that the principles expressed by the CAC can find equal resonance within an abuse-of-dominance matter.

6. Conclusion

The cases examined in this study illustrate the ways in which South Africa's competition authorities are grappling with competition issues in the context of digital markets. In the cases outlined above, we have seen examples of: identifying the importance of the accumulation of data post-merger; preventing a merged digital platform firm from limiting the ability of competitors and new entrants to compete with it; requiring several behavioural conditions, with a 10-year monitoring period, in order to approve a merger of two global digital platform firms; and broadening the interpretation of the Act's provisions on scarce goods and services in order to account for some of the unique dynamics of scarcity in digital markets and in the strategies of digital firms.

We have also seen confirmation that the Act's jurisdiction clause can be interpreted such that the Act applies to the conduct of digital firms based outside of South Africa when the firms' actions have clear negative impacts on the South African economy.

The actions of the South African competition authorities in addressing merger, abuse-of-dominance, and cartel cases in digital markets have shown that, to date, the Act—and the bodies interpreting and conducting enforcement in terms of the Act—are capable of adapting it to digital markets. The final OIPMI report will, it is assumed, bolster the existing capabilities through its focused recommendations on how to address the competition dynamics at play in and among online intermediation platforms. It will also be important to track, in the years to come, how the public interest considerations in the Act (e.g., the impact on small and medium enterprises, and the impact on ownership by historically disadvantaged individuals) are interpreted by the South African competition authorities as they relate to digital markets.

References

Australian Competition and Consumer Commission (ACCC). (2019). *Digital Platforms Inquiry: Final report.* https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf

Ademuyiwa, I. & Adeniran, A. (2020). *Assessing digitalisation and data governance issues in Africa*. Centre for International Governance Innovation (CIGI). https://www.cigionline.org/documents/1767/no244_0.pdf

Competition Commission South Africa (CCSA). (2017, February 15). Competition Commission prosecutes banks (currency traders) for collusion. [Press release.] http://www.compcom.co.za/wp-content/uploads/2017/01/Competition-Commission-prosecutes-banks-currency-traders-for-collusion-15-Feb-2016.pdf

CCSA. (2020a). *Competition in the digital economy: Version 2*. http://www.compcom.co.za/wp-content/uploads/2021/03/Digital-Markets-Paper-2021-002-1.pdf

CCSA. (2020b, December 22). Competition Commission conditionally approves the Google/Fitbit merger. [Media release.] http://www.compcom.co.za/wp-content/uploads/2020/12/Competition-Commission-conditionally-approves-the-Google-Fitbit-merger.pdf

CCSA. (2020c, June 2). Competition Commission refers its case against banks for Rand manipulation to the Tribunal for prosecution. [Media release.] https://www.compcom.co.za/wp-content/uploads/2020/06/Competition-commission-refers-its-case-against-banks-for-rand-manipulation-to-the-tribunal-for-prosecution.pdf

CCSA. (2021a). Online Intermediation Platforms Market Inquiry: Statement of issues. https://www.compcom.co.za/wp-content/uploads/2021/05/OIPMI-Statement-of-Issues_May-2021.pdf

CCSA. (2021b). Notification to approve with conditions the transaction involving: Google LLC (USA) and Fitbit Inc. (USA). Case Number: 2020sep0045. https://www.gov.za/sites/default/files/gcis_document/202110/45250gon968.pdf

CCSA. (2022a, March 14). Facebook prosecuted for abusing its dominance. [Media release.] https://www.compcom.co.za/wp-content/uploads/2022/03/FACEBOOK-PROSECUTED-FOR-ABUSING-ITS-DOMINANCE.pdf

CCSA. (2022b). *Online Intermediation Platforms Market Inquiry: Provisional summary report*. https://www.compcom.co.za/wp-content/uploads/2022/07/OIPMI-Provisional-Summary-Report.pdf

CCSA. (2022c). Final Guidelines on Small Merger Notification. https://www.compcom.co.za/wp-content/uploads/2022/09/Guidelines-on-small-merger-notification.pdf

Competition Commission of South Africa v Bank of America Merrill Lynch International Limited and Others (CR121Feb17) [2019] ZACT 50; [2020] 1 CPLR 205 (CT) (12 June 2019).

Competition Commission v Bank of America Merrill Lynch International Limited and Others (175/CAC/Jul19) [2020] ZACAC 1; 2020 (4) SA 105 (CAC); [2020] 1 CPLR 26 (CAC) (28 February 2020).

Competition Tribunal of South Africa. (2021, March 25). Tribunal grants GovChat, #Letstalk interim relief against WhatsApp and Facebook. [Press release.] https://www.comptrib.co.za/info-library/case-press-releases/tribunal-grants-govchat-letstalk-interim-relief-against-whatsapp-and-facebook

Competition Tribunal of South Africa. (2023, March 30). Tribunal rules that it has jurisdiction to hear the "forex cartel case" against foreign and local banks. [Press release.] https://www.comptrib.co.za/info-library/case-press-releases/tribunal-rules-that-it-has-jurisdiction-to-hear-the-forex-cartel-case-against-foreign-and-local-banks

Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., & Tennison, J. (2020). *The value of data – Policy implications*. Bennett Institute for Public Policy, University of Cambridge; and Open Data Institute. https://www.bennettinstitute.cam.ac.uk/publications/value-data-policy-implications/

Da Silva, F., & Núñez, G. (2021). *Free competition in the post-pandemic digital era: The impact on SMEs*. Economic Commission for Latin America and the Caribbean (ECLAC), United Nations.

Department of Economic Development. (2021). Online Intermediation Platforms Market Inquiry: Terms of reference. *Government Gazette* No. 44432. https://www.gov.za/sites/default/files/gcis_document/202104/44432gon330.pdf

Google and Alphabet v Commission (Google Shopping) (2021). T-612/17.

Google LLC and Alphabet Inc v European Commission (2018). T-604/18.

Google LLC and Fitbit Inc (2020). 2020Sep0045.

GovChat (Pty) Ltd, Hashtag Letstalk (Pty) Ltd v Facebook Inc., WhatsApp Inc., Facebook SA (Pty) Ltd (2020). Order IR165Nov20. Competition Tribunal of South Africa.

Kelly, L., Unterhalter, D., Goodman, I., Smith, P., & Youens, P. (2016). *Principles of competition law in South Africa*. (1st ed.). Oxford University Press.

Khan, L. M. (2017). Amazon's antitrust paradox. *Yale Law Journal*, *126*(3), 710–805.

Maihaniemi, B. (2020). *Competition law and big data: Imposing access to information in digital markets*. Edward Elgar. https://doi.org/10.4337/9781788974264

MIH eCommerce Holdings Pty Ltd t/a OLX South Africa and WeBuyCars Pty Ltd (2020): Reasons for decision. Case No. LM183Sep18. Competition Tribunal of South Africa. https://www.comptrib.co.za/case-detail/8539

Organisation for Economic Co-operation and Development (OECD). (2018) *Implications of e-commerce for competition policy*. https://www.oecd.org/daf/competition/implications-of-e-commerce-for-competition-policy-2018.pdf

OECD. (2020). *Abuse of dominance in digital markets*. https://www.oecd.org/daf/competition/abuse-of-dominance-in-digital-markets-2020.pdf

PricewaterhouseCoopers (PwC). (2023). *Global top 100 companies – by market capitalisation*. https://www.pwc.com/gx/en/audit-services/publications/top100/pwc-global-top-100-companies-2023.pdf

Shelanski, H. A. (2013). Information, innovation and competition policy for the internet. *University of Pennsylvania Law Rev*iew, *161*(6), 1663–1705.

UN Conference on Trade and Development (UNCTAD). (2019). *Competition issues in the digital economy*. TD/B/C.I/CLP/54. https://unctad.org/system/files/official-document/ciclpd54_en.pdf

UNCTAD. (2021). *Digital economy report 2021: Cross-border data flows and development: For whom the data flow*. UNCTAD/DER/2021. https://unctad.org/publication/digital-economy-report-2021

US House of Representatives. (2022). Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary of the House of Representatives. Investigation of Competition in Digital Markets. Majority Staff Report and Recommendations (117 Congress 2d Session). U.S. Government Publishing Office.

COVID-19, *kovhidhi, dzihwamupengo*: Language use, language change, and pandemic perceptions among Shona-speakers in Gweru, Zimbabwe

Vincent Jenjekwa
Deputy Director and Research Fellow, MSU National Language Institute, Midlands State University, Gweru, Zimbabwe
https://orcid.org/0000-0003-4747-8442

Abstract
Through an examination of the linguistic practices encountered and used by Shona language-speakers in the Zimbabwean city of Gweru, this study explores intersections between language use, language change, and perceptions of the COVID pandemic—as caused by the virus referred to by Gweru's Shona-speakers as, variously, "COVID-19" in its English-language representation or *"kovhidhi"* or *"dzihwamupengo"* in its two most common Shona-language representations. The study is anchored in conceptions of the impacts that natural disasters and pandemics have on language and on communication needs, and in theories of semiotics and language change. The research finds that the predominant terms used by Gweru's Shona-speakers in relation to the pandemic carry connotations that, in the Zimbabwean socio-cultural context, potentially undermine optimal responses to the pandemic. The article concludes by emphasising the importance of careful management of language as a critical resource in the fight against natural disasters and pandemics.

Keywords
language, communication, semiotics, language change, pandemics, natural disasters, COVID-19, Shona, English, Gweru, Zimbabwe

DOI: https://doi.org/10.23962/ajic.i31.13995

Recommended citation
Jenjekwa, V. (2023). COVID-19, *kovhidhi, dzihwamupengo:* Language use, language change, and pandemic perceptions among Shona-speakers in Gweru, Zimbabwe. *The African Journal of Information and Communication (AJIC), 31*, 1-17. https://doi.org/10.23962/ajic.i31.13995

1. Introduction

The COVID-19 virus has, since its first detection in Zimbabwe in March 2020, generated the emergence of several new and altered terminologies in response to the pandemic. This evolving vocabulary is the result of universal and natural phenomena of language use and language change—dynamic processes through which language adapts and evolves to meet emergent communication realities and requirements. This study sought to establish the intersections between language use, language change, and perceptions of the COVID-19 pandemic among Shona-speakers in the city of Gweru, the capital of Zimbabwe's Midlands Province. Through qualitative analysis of data from traditional and social media content, from observation of public behaviours, and from semi-structured interviews, a picture emerged wherein language use and change were potentially militating against effective public responses to the pandemic.

Section 2 discusses the impact of natural disasters and pandemics on language change, and section 3 examines the role of communication in the reduction of risk and vulnerability in disaster and pandemic situations. Section 4 sets out the study's theoretical frameworks (semiotics and language change); section 5 describes the research design; section 6 provides findings and discussion; and section 7 sets out conclusions and recommendations.

2. Impact of natural disasters and pandemics on language change

In general terms, language is a system of communication. However, more technically, "[l]anguage is a system of arbitrary vocal symbols, which permit all people in a given culture, or other people who have learned the system of that culture to communicate or to interact" (Finocchiaro, 1964, p. 8). Language plays a critical role in the management of human activities, and one of its enduring features is its dynamism as it responds to changes in social circumstances. As stated by Paton (2020), "great social change brings great linguistic change". Such language change is a universal phenomenon that affects all languages at different times and to different degrees.

Language change is generally perceived as the evolution of a language's phonology, morphology, syntax, and semantics owing to various internal and external factors. As a system of signs and symbols generated culturally, language is dynamic and responsive to social changes. The internal factors relate to particular changes linked to the structure or articulation of certain vowels or consonants in a language. Internal changes can, for instance, occur as speakers exercise economy in communication, resulting in certain structures being omitted. In Zimbabwe's Shona language, for example, the word *mwanakomana* (a male child) is rather long and speakers would rather use *mukomana*. Similarly, the traditional Shona word *murandakadzi* (a woman) has been largely eclipsed by *mukadzi*, even though *murandakadzi* still exists in the language.

External language change can be occasioned by several factors, including, for instance, the arrival of speakers of another language in a community. Migration is, evidently, one of the key factors behind language change. Also, there is no doubt that in Africa, owing to colonisation, most Indigenous languages have been affected in varying degrees by exposure to European languages (English, in the case of Zimbabwe). As of today, Indigenous languages such as Shona have several words whose etymology is traceable to English. For example, *chipunu* is derived from the English word "spoon".

Also, Indigenous African languages have been seen to influence each other mutually, particularly in areas where these languages exist within the same (often colonially decided) national borders. Khumalo (2004), for example, identifies the contact between the Shona and Ndebele languages in Zimbabwe in Zimbabwe's south-western regions where the Ndebele people settled upon arrival from Zululand (in today's South Africa). The then-Ndebele King, Mzilikazi, conquered local Shona chiefs, known as *MadziMambo* (*Mambo* singular), and their subjects. These subjects were previously known as *VekwaMambo* in Shona, and the Ndebele borrowed the term and generated a new Ndebele word, *AbakaMambo* (Khumalo, 2004, p. 107). Other Ndebele words which emerged as a result of contact with Shona are *umuntu umile* (standing person), which came from the Shona phrase *munhu amire*, the Ndebele word *samukele* (welcome) derived from the Shona *tigamuchire/tigashire*, and the Ndebele word *hambisa* (hurry) derived from the Shona *fambisa* (Khumalo, 2004). As is discussed later in this article, the conceptual framework that has come to be applied to this kind of language change is known as "lexical borrowing".

Another set of key drivers of language change are natural disasters and pandemics. Natural disasters, according to IASC (2006, p. 8), are "the consequences of events triggered by natural hazards that overwhelm local response capacity and seriously affect the social and economic development of a region". A pandemic is a disease that affects many people in many different countries (WebMD, n.d.). As Popiolek (2020) states, "[h]istorically, major events like natural disasters and war have proven to have big impacts on language". Disasters and pandemics unleash unprecedented social activities that necessitate the emergence of new vocabulary and, sometimes, the reconfiguration of existing vocabulary (Chmutina & von Meding, 2019; Chmutina et al., 2019).

Natural disasters and pandemics can render certain habitable places uninhabitable and, therefore, encourage migration, resulting in the mingling and mixing of speakers of different languages. On another level, pandemics and natural disasters contribute to some of the most widely used vocabulary items globally. The Spanish influenza of 1918, for instance, still has its footprints in Shona as *furuwenza*. Also, poliomyelitis, a viral disease that causes paralysis of the legs and which emerged first in 1878 but worsened after World War II, is a permanent feature of both English and Shona. In English it is known as polio. In Shona, it is often translated as *mhetamakumbo*, which

is used to refer to both the disease and a victim of the disease. It is also transliterated as *poriyo*, drawing directly on the sounds of the English word. This word *poriyo* has ceased to be solely the name of a virus and is now used to refer to post-natal visits in general, with women who have babies colloquially saying "*tiri kuinda kuporiyo*" ("we are going to polio") when taking the baby to a medical facility.

Disasters and pandemics can also cause semantic shifts. Semantic shift is the change in the meaning of a word or phrase. For example, the Japanese word *tsunami*, once an expression used only to describe destructive tidal waves, has gathered new meanings as it has spread globally due to the global media coverage of the 2004 *tsunami* off the coast of Indonesia and the 2011 *tsunami* off the coast of Japan. In some cases, it is now used to refer to drastic organisational changes. And, in Zimbabwe, there is a herbal medicine available on the informal market that is called Tsunami (Moyo, 2017). This medicine is believed to be able to cure all ailments, akin to the manner in which a *tsunami* wave can wash away any item, big or small, in its path.

More recently, in the context of the coronavirus pandemic, the term "COVID-19" has gained global use (Paton, 2020). Also, the phrase "social distancing", which was previously seldom used and typically would have been understood as referring to a lack of interest in engaging with others, has, in the context of COVID, gained a new meaning, now referring to the avoidance of physical proximity. As new terminologies emerge, we need to examine the degree to which the new vocabulary mitigates or aggravates the impact of a natural disaster or pandemic. The following section deals with the importance of communication in such situations.

3. Role of communication in responding to natural disasters and pandemics

The way in which information is disseminated can have a profound effect on the success or failure of mitigatory interventions during natural disasters and pandemics. Writing about the role of communication in combating Ebola in the Democratic Republic of the Congo (DRC), Kemp (2020) argues that "effective communication with communities at risk is essential to containing disease outbreaks". Kemp (2020) further notes that the use of militaristic language is counterproductive as it engenders feelings of fear and might discourage affected people from seeking medical attention. Wise (2020) contends that military metaphors make a "desperate appeal to the necessity of chaos" and justify the abandonment of the rule of law, contrary to the standard practice of medicine which is underpinned by paying particular heed to procedure and practice. However, militaristic jargon in medicine has a long history and is used widely. The relationship between the development of medicine historically and the military is very intimate, and, accordingly, narratives of ill-health and recovery are often similar to narratives of vulnerability and defence in military discourse.

Chmutina et al. (2019) argue that the use of certain words and phrases has created erroneous perceptions about pandemics and diseases. These authors find that the very term "natural disaster" is a misnomer that blames nature and ignores the roles of people and entities that create disaster risk. According to Chmutina and Von Meding (2019, p. 284), "disasters result from the combination of natural hazards and social and human vulnerability, including development activities that are ignorant of local hazardous conditions". Yet the use of language can dissemble the truth by diverting attention from the issues of politics, planning, governance and media, among other key aspects that determine whether or not a hazard becomes a disaster.

Cardona (2004, p. 37) defines vulnerability as something that originates in human experience and "represents the physical, economic, political, and social susceptibility or predisposition of a community to damage in the case [of] a destabilizing phenomenon of natural or anthropogenic origin". As a result, disasters do not impact communities equally. This is why Brandt and Botelho (2020, p. 1494) argue that "[e] pidemics are not simply natural events: they are also the result of human actions, in both their emergence and containment". The use of inaccurate terminology is an important example of how language diverts attention and potentially aggravates the situation by increasing confusion and vulnerability.

Brandt and Botelho (2020) observe that COVID-19 has been described in the mainstream media as "a perfect storm"—a phrase that implies an anomalous and unpredictable "worst possible" storm—and yet there are several political and social factors that encourage its spread. According to Sontag (2001), as cited in Brandt and Botelho (2020), the metaphors used to describe pandemics profoundly shape people's experience of an illness. The cultural discourses on diseases such as cancer and AIDS, for instance, produce fear and stigma that potentially marginalise patients and hinder proper care. Therefore, the (often inadvertent) use of counterproductive language in disaster and pandemic situations undermines potentially helpful interventions. Kemp (2020) notes, for example, that the use of the word "isolation" generates fear in people. Similarly, the use of technical terms can be counterproductive, as they are often not understood by many citizens, particularly those who do not have formal education.

In the management and mitigation of natural disasters, ReliefWeb (2020) identifies risk communication and community engagement as priority areas, and therefore "all responders [need] to communicate effectively with communities, counter misinformation, and make sure people can hold them accountable". ReliefWeb (2020) further observes that communicating disaster information in non-vernacular international languages makes marginalised and often illiterate people more vulnerable.

4. Theoretical frameworks

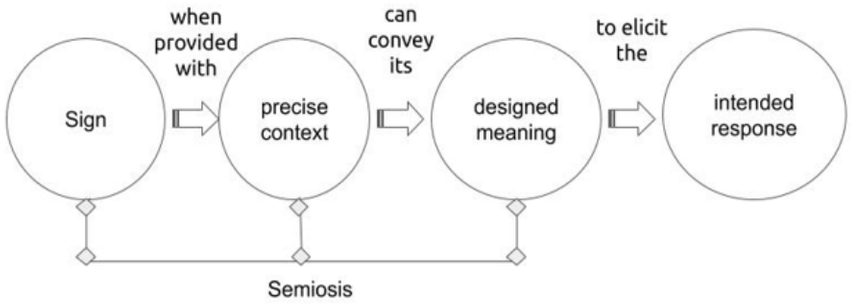
This study is anchored in two theoretical frameworks: semiotics and language change.

Semiotics

Sebeok (2001, p. 3) defines semiotics as "how messages are, successively, generated, encoded, transmitted, decoded, and interpreted, and how this entire transaction (semiosis) is worked upon the context". The main early proponents of semiotics were Swiss linguist Saussure (1857–1913) and American philosopher Peirce (1839–1914) (Peirce, 1965; Saussure, 1966). A critical component of semiotics is the sign. Scollon and Scollon (2003, p. 3) define a sign as "any material object that indicates or refers to something other than itself". Semiotics studies how the use of signs in communication affects the intended outcome.

The semiotics theory consists of three main concepts—sign, context, and meaning—as illustrated in the Figure below. According to Saussure (1966), as cited in Riera (2021), a sign is composed of both the form it takes in physical reality and how it is conceived or interpreted by its receiver. The sign can be vocal, visual, or action. Language is a vocal system of communication that could be spoken or written, and its context involves both the physical and social aspects that make the receiver of the information understand it. The effective use of signs is context-determined. Meaning in semiotics pertains to the interplay between the message and the prior knowledge of the receiver. Effective communication is thus relative to the prior knowledge that the receiver has of the information.

Figure 1: A model of semiotic theory (adapted from Riera, 2021)



In terms of information on pandemics, effective dissemination depends on the capacity of the receivers to understand the language used, together with all the technical aspects, such as the names of the disease. In the case of COVID-19, this process can depend, for example, on whether the listener or reader has prior knowledge of technical and scientific communication. Effective communication happens when

the sign is used appropriately in a context to convey the intended message of the sender. The level of precision in the choice of signs, and the interplay with context and meaning, determine the effectiveness of communication. The use of unique, unfamiliar, and confusing signs affects the process of semiosis by creating barriers to communication. Therefore, language as a sign system should be handled with care to ensure mutual understanding, between the sender and the receiver of the information, particularly when the receiver is expected to take a certain action or set of actions following receipt of the information.

Language change

Language change is a sociolinguistic phenomenon that explains how language changes both synchronically (at one point in time) and diachronically (over time). In simple terms, language change theories seek to explain the continual and dynamic changes that languages undergo as they are used for communication (Hickey, 2001; Lass, 1997; Matsumoto, 2019). The theorisations examine the factors behind language change and the forms of the change. The changes that affect language affect the phonology, morphology, syntax, and semantics of a particular language. Because of the breadth of conceptualisations of language change, this study—guided by the orientation of semiotic theory—used an approach (as explained in Hickey, 2001) that focuses on how language change is occasioned by a mix of morphological (internal) and semantic (external) factors.

Internal changes are caused by, among other elements, the need to achieve simplicity in a language. For example, as speakers of a particular language use it, there is an intuitive inclination to shorten and simplify words that appear long and difficult to articulate. Internal changes can also result from the over-generalisation of a linguistic rule in a particular language. For example, the use of the past tense morpheme *-ed* on words such as catch, run, and eat by some Zimbabwean speakers of English might eventually lead to the wider use of "catched", "runned", and "eated" in Zimbabwean English. External changes mostly affect a language at the semantic level. These semantic changes can be categorised into metonymy and metaphorical changes. Metonymy occurs when the "meanings arise from something closely related to another meaning present in the discourse" (Hickey, 2001, p. 15). Metaphorical change, according to Hickey (2001, p. 15), occurs when "a figurative meaning of a word develops alongside a more literal one".

As seen above in the work of Khumalo (2004) on contact between the Shona and Ndebele languages, language change can also occur on a lexical level as a result of contact between two or more languages. The core conceptual framework that has been developed to account for such language changes is "lexical borrowing". According to Miller et al. (2020, p. 1), lexical borrowing refers to "[t]he transfer of words from one language to another" to meet the lexical needs of the speakers. Grant (2015, p. 431) states that lexical borrowing "constitutes the commonest form of contact-induced

linguistic change". The borrowed terms are then adapted to the phonology of the recipient language (Kiparsky, 2014).

In designing the study on which this article is based, I determined that the two theoretical frameworks just outlined—semiotics and language change (including lexical borrowing)—could be effective for exploring how COVID-19-related communication may have consequences for the management of the pandemic among Shona-speakers in the Zimbabwean city of Gweru.

5. Research design

Study objective and research questions

The study sought to establish the potential impact of language use on efforts to address the COVID-19 pandemic in Gweru. It was specifically guided by two research questions:

- What is the conceptualisation of COVID-19 among Shona-speakers in Gweru?
- How has language change and/or use mitigated or aggravated the risk of COVID-19 among Shona-speakers in Gweru?

Methodology

The research was a case study of COVID-19 communication in the city of Gweru. This city was selected because of its close proximity to my workplace and because of the presence of many people in the central business district (CBD). Such a choice of a case is acceptable in qualitative studies (Guest et al., 2006; Patton, 2002; Yin, 2009).

Media content (including social media), observation, and semi-structured in-depth interviews were the sources of primary data. The data was collected from August to October 2021. The scrutiny of media content focused on digital, print, and online media stories on COVID-19, with a particular focus on identifying new terminologies (and any potential insinuations) linked to the disease. Observation focused on observing the behaviour of the general public in relation to the government's stipulations and protocols on COVID-19 prevention. I took notes while observing behaviour in the Gweru CBD, e.g., in queues at banks, supermarkets, and bus stations. Media data sources consisted of COVID-related content produced by National FM radio and two Gweru-based radio stations, 95.9 Central Radio and 98.4 Midlands; COVID-related reports by *The Herald* newspaper; international news reports focused on COVID in Zimbabwe; and COVID-related videos posted by Zimbabweans on YouTube.

The interviews, conducted in a mix of Shona and English and guided by a semi-structured interview protocol, were carried out with 10 adult participants from the delimited area. Since the study intended to establish the impact of language use, the

sampling had to cater to a cross-section of people. Accordingly, the 10 participants, selected through purposive sampling, had the following characteristics: a 50:50 gender ratio; ages ranging from 22 to 65 years; and a wide range of educational levels: one person with a PhD degree, two with Master's degrees, four with post-secondary diploma qualifications, two with primary-level education, and one with no formal education. The participants were recruited at Midlands State University Gweru City Library, known informally as "Hellenics"; at stores in the Gweru CBD; and outside the Gweru long-distance bus terminus, known as Kudzanai. The bus terminus has several market stalls operated by vendors with little or no formal education.

The interview protocol was composed of questions about: knowledge of the meaning of the term COVID-19; sources of information about COVID-19; whether the interviewee had all the necessary information about COVID-19; measures to protect oneself from COVID-19; experiences with the loss of relatives, friends, or neighbours to COVID-19; perceptions of quarantine and isolation measures; vaccination status and attitudes towards vaccination; and perceptions of the effectiveness of communication on COVID-19 by the media and the Ministry of Health and Child Care. The questions were asked mainly in Shona, with frequent code-switching to English where necessary. Follow-up questions were asked based on initial responses given by the interviewee. Ethical considerations, including informed consent and voluntary participation, were given due regard, following the guidance of Guest et al. (2006), Creswell (2014), Patton (2002), and Yin (2009). The participants provided informed consent and were free to withdraw from the study at any time if they wished.

The analysis of the generated data was guided by semiotics and language change as the theoretical foundations of the study.

6. Findings and discussion

This section discusses the terminologies (a mix of Shona and English words) that have emerged, and/or gained altered meanings/significance, among Shona-speakers in Gweru as a result of the COVID-19 pandemic, and examines the terms through the theoretical lenses of semiotics and language change. The following terms are discussed:

- COVID-19, *kovhidhi*
- *dzihwamupengo*
- *kuzvitsaura wakagara uri kumba* (quarantine, isolation, social distance)
- *korona* (corona, crown)
- *furuwenza* (COVID-19 pandemic)
- *zhing-zhong* (low-cost and low-quality Chinese products)

COVID-19, kovhidhi

The English term "COVID-19" and its "Shonalised" representation "*kovhidhi*" were found to be the most frequently used terms for the disease among the Shona-speaking study participants in Gweru. And it was found in the interviews that the term struck the majority of the interviewees as referring to a mysterious disease—one that was incomprehensible. One interviewee stated that "*hatitombozivi kuti chii* ("we don't even know what it is") (interviewee 8, 2021).

While the meaning of the term COVID-19/*kovhidhi* was found to be reasonably well understood by the better-educated participants (i.e., those with a post-secondary degree or diploma), even they seemed at the same time to be confused by the term's origins. To such citizens, using the term COVID-19/*kovhidhi* was apparently like using the chemical formula for water, H_2O, instead of simply calling it water—and they felt that it was not realistic to refer to the disease in scientific terms and expect the majority of citizens to comprehend it. All interviewees concurred that the name COVID-19/*kovhidhi* had, therefore, not assisted in the control of the disease because it had deepened fear and confusion and made some people indifferent to vaccination and medication.

This shows how the use of language can achieve negative outcomes if it is not properly aligned to the needs and abilities of the citizens. Accordingly, in pandemic situations, health professionals should be supported to immediately translate key terminology into vernacular languages, before seeking to enlist community buy-in. Naming pandemics in the idioms of the communities is not new—as seen earlier in the example of polio. In this study, it was found that the cryptic term COVID-19/*kovhidhi* was reverberating on a daily basis in Gweru, and in Zimbabwean media, as some kind of incomprehensible bringer of death that the citizens had to anxiously await.

It also emerged from the observations and interviews that the government's communication on how to combat the disease had apparently not succeeded in making citizens fully understand their role. The observed behaviours and interactions of people in places such as supermarkets, bus termini, and on public transport showed general disregard for advice from the Ministry of Health and Child Care on the wearing of masks, social distancing, and avoidance of shaking hands. I frequently observed masks being worn strapped around the chin with the mouth and nose uncovered; people shaking hands; and non-adherence to social distancing.

According to one interviewee: "*hatitombozivi kuti tinofanira kudii kuti tirwise chirwere ichi. Kuvhara muromo nokugeza maoko tiri kuita asi vanhu vari kungofa.*" ("We don't even know what we should do to fight this disease. We are trying to cover our mouths and wash our hands, but people are just dying") (interviewee 4, 2021). One likely contributor to the confusion had been statements by leading political

figures in Zimbabwe, reported in the media in the early stages of the outbreak, which presumably complicated the citizens' appreciation of the seriousness of COVID-19. For example, Oppah Muchinguri, Zimbabwe's Minister of Defence, was quoted in the media as saying:

> This coronavirus that has come [is a form of] sanctions against the countries that have imposed sanctions on us. God is punishing them now and they are staying indoors now while their economy is screaming like what they did to ours by imposing sanctions on us. (as quoted in Mutsaka, 2020)

This statement, coming from a high-ranking government official, deflected attention from the need for Zimbabweans to pursue mitigation strategies.

Dzihwamupengo

Another Shona expression that had emerged as a name for COVID-19 was *dzihwamupengo*. Though used only sporadically in the media, the public, and by certain artists (Kwayedza, 2021; Mhanduwe weNhau, 2020; 2021; Murewanhema, 2021), the term is significant for this study because of its potentially rich linguistic elements. The term means "a rabid cough", and is a complex nominal construction that brings together two words: *dzihwa* and *mupengo*. The term is grounded in a metaphor whereby COVID-19 spreads like a raging fire. *Dzihwa* refers to general sickness, and can also be used to refer to an ordinary cold. When something is *mupengo* (crazy), it defies all norms and conventions, because, in Shona, a person who is *mupengo* is one who is mentally ill and is characteristically identified by unpredictability, violence, and anti-social behaviour. Many interviewees indicated that this inclusion of *mupengo* in the term *dzihwamupengo* struck fear in them.

From a different angle, the morphology of *dzihwamupengo (dzihwa + mupengo)* is similar to that of: *chimbwamupengo*, a rabid dog suffering from rabies; *dutumupengo*, a cyclone or hurricane; and *bwowamupengo*, a poisonous mushroom. These are all complex nominal constructions made up of two separate words. Also, based on the semiotic principle of context in communication, the word *dzihwamupengo* can be seen as evoking the word *dutumupengo*—the name given to Cyclone Idai, which caused widespread damage and displacement, and more than 300 deaths, in the eastern parts of Zimbabwe in early 2019 (Chatiza, 2019; Munsaka et al., 2021).

The term *dzihwamupengo* suggests a lack of comprehension of the real nature of the disease, and points to its incurability—just as a mentally ill person is uncontrollable if not treated. The fear element generated by the term *dzihwamupengo* can also potentially be traced to the Shona cultural strategy of deterring people from certain things or places through the use of names. For example, to ward people away from certain places such as sacred mountains, the Shona call those places *kumazivandadzoka* ("a dangerous place one would regret visiting") to discourage those who would want

to explore them. Thus, the term *dzihwamupengo* creates a sense of a highly infectious, incurable, and deadly disease that is not selective based on age, economic status, and so on. As a product of the interplay between the Shona culture and language ecology, the name *dzihwamupengo* portrays how certain Shona-speakers sought to make sense of COVID-19 using familiar linguistic elements. As a word that emerged as an effort to enhance communication, *dzihwamupengo* can be said to ironically achieve the opposite—because it engenders feelings of fear and helplessness.

Kuzvitsaura wakagara uri kumba (quarantine, isolation, social distance)

Closely associated with COVID-19 are the concepts and terminologies around quarantine, isolation, and isolation centres. Though these concepts are not new in local languages, particularly in Shona where isolation and quarantine are known as *kuzvitsaura wakagara uri kumba* ("isolating yourself while staying at home"), the overwhelming use of English when referring to these concepts (many Shona people engage in considerable code-switching between Shona and English) potentially marginalises the illiterate, who might have little or no proficiency in English.

Meanwhile, for the well-educated interviewees who had no difficulty understanding the terms isolation and quarantine, the terms were still problematic because they carried negative connotations of rejection and condemnation of death in isolation. Many of Zimbabwe's COVID-19 deaths occurred in isolation centres, and these incidents were often reported in the media and via word-of-mouth. The isolation centres were associated with, in the words of one well-educated participant, "painful death of patients without their relatives and care" (interviewee 1, 2021). Accordingly, the participant spoke of preferring to suffer silently, without going for testing, so as to avoid being "taken to the isolation centre to die without my relatives and family".

The Herald newspaper had reported numerous escapes from the quarantine centres (*The Herald*, 2020; 2021). Despite sometimes being guarded by soldiers and the police, people had found ways of escaping. According to the interviewees, isolation in the context of COVID-19 equalled death, and hence the negativity associated with the centres.

As a semantic sign, the term "isolation" seems clearly to have gathered new negative semantics in the context of COVID-19. Based on the responses of interviewees and my observations, even if COVID-19 were to end, the concept of isolation (no matter how positive its intentions are) would likely be detested by the majority of citizens in Zimbabwe. Such kinds of semantic shift are quite normal in language and have been noted by other scholars during this COVID-19 pandemic with respect to phrases such as "social distance" (Mehta, 2020).

Korona (corona, crown)

Another prominent Shona word in the discourse of the COVID-19 pandemic is *korona* (corona), a term which was used often in the early days of the disease. It was a word already used in the Shona vocabulary to refer to a crown, or to refer to the thorny flowers used as a hedge in certain homes and institutions. Interviewees indicated that because there had been very little education of the general citizenry on the use of the term in the COVID context, those with low literacy or low access to news media were left confused about the meaning of *korona* in this new context.

Another dimension of the word *korona* in Shona is its use to refer to the biblical crown of thorns that was placed on the head of Jesus. This term *korona* thus played into the hands of some Christian sects, who, in one instance in an audio message distributed on social media, opposed seeking medical attention in response to COVID-19, arguing that it was the crown of thorns that people should just bear as Jesus did. Some sects labelled the COVID-19 vaccines as the mark of the beast, which should not be taken by Christians. This had led one Christian minister, Emmanuel Makandiwa, to refute this position publicly and advise his congregants that they should take the vaccine:

> There is no mark of the beast in the vaccine. Christians are afraid, fearing for their lives and trying to protect themselves and their loved ones from receiving the mark of the beast by receiving the vaccine. This COVID-19 vaccine is not the mark of the beast […] (as quoted in Ncube, 2021).

At the same time, it was found that some people were giving the word *korona* a lighter dimension. Evidence from social media indicated that the word was often being construed positively, because it was bringing couples together at home. The power of the disease to rein in wayward husbands, who never used to spend much time with their families, particularly their wives, was being celebrated. The pandemic was said to be rekindling marital love, because husband and wife were forced to be indoors most of the time—even resulting in the *korona* of conception to some people, owing to repeated sexual intercourse. This dimension was bringing some "comic relief" to the macabre drama of the COVID-19 pandemic.

Furuwenza (COVID-19 pandemic)

Another term found to be used repeatedly in relation to COVID-19 was the Shonalised representation for influenza: *furuwenza*. The mere mention of the pandemic evoked memories of the Spanish influenza pandemic of 1918–19 and the SARS outbreak of 2003. One elderly interviewee recalled harrowing stories she was told by her grandmother about the sudden deaths that occurred during the Spanish influenza pandemic. *Furuwenza* is recorded in both the oral and formal history of Zimbabwe, and equating the COVID-19 pandemic with *furuwenza* was found to be quite chilling among participants because both are respiratory infections. Also, to some Christians, the COVID-19 *furuwenza* is seen as a plague similar to the biblical

plagues sent by God to punish disobedient Israelites. These connotations reflect Khumalo's (2004, p. 106) argument that linguistic changes can be instigated by both "obvious and mysterious" reasons and might be perceived differently by different speakers of the language.

Zhing-zhong (low-cost and low-quality Chinese products)

A majority of the interviewees stated that the Chinese COVID-19 vaccines (which were the predominant ones used in Zimbabwe) were *zhing-zhong*, a derogatory term meaning that they were inexpensive and unreliable—the reputation assigned to numerous other Chinese products on the market in Zimbabwe. This showed that the semiosis of vaccination in the context of COVID-19 in Zimbabwe was entangled with the longstanding negative attitudes of citizens towards Chinese commodities. These attitudes undermined the mitigatory strategies put in place by the government. According to the participants, the initial take-up rate for the vaccines was extremely low.

7. Conclusion and recommendations

The foregoing findings and discussion point to the dynamism of language and the need for planned communication in a pandemic situation. As a semiotic process—an interplay between the sign and context to give meaning—language and the words it is composed of generate meanings that can only be taken for granted at great risk. The findings of this study have demonstrated how the use of language has entrenched fear and confusion in the context of the fight against COVID-19 in the Zimbabwean city of Gweru, and presumably in much of the rest of the country, thus marginalising large segments of the country's population. It seems clear that the hasty and sometimes spontaneous responses from government have led to the neglect of language issues, and yet language is the vehicle of effective communication.

If the evolution of new terminologies is not systematically controlled, and the terms disseminated widely to all stakeholders, including via the media, unintended counterproductive outcomes can emerge. In a pandemic context, the dissemination of information in vernacular languages should be prioritised, and where there is a need, new terminology should be generated in local languages to counter the fear and uncertainty generated by the use of foreign untranslated vocabulary. In addition, the terms used in the vernacular languages must be carefully chosen. This is because the use of terms, acronyms, and concepts that are not fully understood by the target population breeds indifference, suspicion, and fear—often exacerbated by the terms' intersections with complex socio-cultural structures—which discourages people from taking recommended precautions.

References

Brandt, A. M., & Botelho, A. (2020). Not a perfect storm – COVID-19 and the importance of language. *The New England Journal of Medicine*, *382*(16), 1493–1495. https://doi.org/10.1056/NEJMp2005032

Cardona, O. D. (2004). The need for rethinking the concepts of vulnerability and risk from a holistic perspective: A necessary review and criticism for effective risk management. In G. Bankoff, G. Frerks & D. Hilhorst (Eds.), *Mapping vulnerability: Disasters, development and people* (pp. 37–51). Routledge.

Cassim, J. (2020, March 18). Zimbabwe declares coronavirus "national disaster". *Anadolu Ajansı*. https://www.aa.com.tr/en/africa/zimbabwe-declares-coronavirus-national-disaster/1769900

Chatiza, K. (2019). *Cyclone Idai in Zimbabwe: An analysis of policy implications for post-disaster institutional development to strengthen disaster risk management*. Oxfam. https://doi.org/10.21201/2019.5273

Chmutina, K., & von Meding, J. (2019). A dilemma of language: "Natural disasters" in academic literature. *International Journal of Disaster Risk Science*, *10*, 283–292. https://doi.org/10.1007/s13753-019-00232-2

Chmutina, K., von Meding, J., & Bosher, L. S. (2019). Language matters: Dangers of the "natural disaster" misnomer. https://repository.lboro.ac.uk/articles/online_resource/Language_matters_Dangers_of_the_natural_disaster_misnomer/9460862

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches.* (4th ed.). SAGE.

Croke, K., Grossman, G., Larreguy, H. A., & Marshall, J. (2014). *The effect of education on political participation in electoral authoritarian regimes: Evidence from Zimbabwe*. Semantic Scholar.

Finocchiaro, M. (1964). *English as a second language: From theory to practice*. Simon and Schuster.

Grant, A. (2015). Lexical borrowing. In J. R. Taylor (Ed.), *Oxford handbook of the word* (pp. 431–444). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641604.013.029

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). SAGE.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hickey, R. (2001). Language change. In J. Verschueren, J.-O. Östman & J. Blommaert (Eds.), *Handbook of pragmatics*. John Benjamins. https://www.uni-due.de/SHE/Language_Change.pdf

Inter-Agency Standing Committee (IASC). (2006). Protecting persons impacted by natural disasters: IASC operational guidelines on human rights and natural disasters. Brookings-Bern Project on Internal Displacement.

Kemp, R. (2020). Replacing the language of fear: Language and communication in DRC's latest Ebola response. *Humanitarian Exchange* magazine, No. 77. https://odihpn.org/publication/replacing-the-language-of-fear-language-and-communication-in-drcs-latest-ebola-response

Khumalo, L. (2004). Language contact and lexical change: A lexicographical terminographical interface in Zimbabwean Ndebele. *Lexikos*, *14*(2004), 105–117. https://doi.org/10.4314/lex.v14i1.51413

Kiparsky, P. (2014). New perspectives in historical linguistics. In C. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 64–102). Routledge.

Kwayedza. (2021, March 12). Detembo: Dzihwamupengo. [Poem: COVID-19.] https://www.kwayedza.co.zw/detembo-dzihwamupengo

Lass, R. (1997). *Historical linguistics and language change*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620928

Matsumoto, K. (2019). Language variation and change. In P. Heinrich & Y. Ohara (Eds.), *The Routledge handbook of Japanese sociolinguistics* (pp. 199–217). Routledge. https://doi.org/10.4324/9781315213378-13

Maulana, N., Nyadera, I. N., & Wandekha, B. (2020). The generals and the war against COVID-19: The case of Zimbabwe. *Journal of Global Health*, *10*(2), 020388. https://doi.org/10.7189/jogh.10.020388

Mehta, V. (2020). The new proxemics: COVID-19, social distancing and sociable space. *Journal of Urban Design*, *25*(6), 669–674. https://doi.org/10.1080/13574809.2020.1785283

Mhanduwe weNhau. (2020, December 12). Dzihwamupengo riri kurura kuzvikoro nomumamana. [COVID-19 is wreaking havoc in schools and communities.] *YouTube*. https://www.youtube.com/watch?v=Z19JE0pWtys

Mhanduwe weNhau. (2021, January 22). Dzihwamupengo riri kuita humaenzanise munyika. Chigwedere nowamwe wapfa [COVID-19 is meting out justice: Chigwedere [former government minister] and others have died.] *YouTube*. https://www.youtube.com/watch?v=1Qv-Qz3f2uk

Miller J. E., Tresoldi T., Zariquiey, R., Beltrán Castañón, C. A., Morozova, N., & List, J.-M. (2020). Using lexical language models to detect borrowings in monolingual wordlists. *PLoS ONE 15*(12), e0242709. https://doi.org/10.1371/journal.pone.0242709

Montana State University. (2020). SARS-CoV-2 background & origins. *Montana COVID-19 Blog*. https://www.montana.edu/diseaseecologylab/covid19blog/sars-cov-2-background-origins/index.html

Moyo, N. (2017, April 30). MCAZ warns public against "Tsunami". *Sunday News*. https://www.pressreader.com/zimbabwe/sunday-news-zimbabwe/20170430/281762744153865

Munsaka, E., Mudavanhu, C., Sakala, L., Manjeru, P., & Matsvange, D. (2021). When disaster risk management systems fail: The case of Cyclone Idai in Chimanimani District, Zimbabwe. *International Journal of Disaster Risk Science*, *12*(5), 689–699. https://doi.org/10.1007/s13753-021-00370-6

Murewanhema, G. (2021, July 19). Corona yava kuda kuita muonera, chuma chemuzukuru [Collaborative effort is needed to tackle coronavirus]. *Health Times*. https://healthtimes.co.zw/2021/07/19/corona-yava-kuda--kuita-muonera-pamwe-chuma-chemuzukuru

Mutsaka, F. (2020, March 16). Zimbabwe official says coronavirus punishes US for sanctions. *Associated Press.* https://apnews.com/article/africa-international-news-united-states-zimbabwe-virus-outbreak-5289697596801a1907c7b9decddee34a

Ncube, A. L. (2021, July 26). Prophet Emmanuel Makandiwa backtracks on COVID-19 vaccine. *iHarare*. https://iharare.com/prophet-emmanuel-makandiwa-backtracks-on-covid-19-vaccine

Paton, B. (2020, April 9). Social change and linguistic change: The language of COVID-19. *OED blog*. https://public.oed.com/blog/the-language-of-COVID-19/

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). SAGE.

Peirce, C. P. (1965). Basic concepts of Peircean sign theory. In M. Gottdiener, K. Boklund-Lagopoulou & A. P. Lagopoulos (Eds.), *Semiotics*. SAGE.

Popiolek, K. (2020, August 10). Researchers study how COVID pandemic is affecting language change. *MSU Today*. https://msutoday.msu.edu/news/2020/researchers-study-how-covid-pandemic-is-affecting-language-change

ReliefWeb. (2020). Do you speak COVID-19? The importance of language for effective communication across the response. https://reliefweb.int/attachments/663f246a-f6dc-3491-947f-fac59384a270/TWB_PolicyBrief-COVID19.pdf

Riera, J. J. (2021). Semiotic theory. In J. Egbert & M. F. Roe (Eds.), *Theoretical models for teaching and research*. Press Books. https://opentext.wsu.edu/theoreticalmodelsforteachingandresearch/chapter/semiotic-theory

Saussure, F. de (1966). *Course in general linguistics*. Wade Baskin.

Scollon, R., & Scollon, S. W. (2003). *Discourses in place: Language in the material world*. Routledge. https://doi.org/10.4324/9780203422724

Sebeok, T. A. (2001). *Signs: An introduction to semiotics*. University of Toronto Press.

Sontag, S. (2001). *Illness as metaphor and AIDS and its metaphors*. Picador.

The Herald. (2020, May 27). 118 escape from isolation centres. https://www.herald.co.zw/118-escape-from-isolation-centres

The Herald. (2021, July 26). COVID-19 positive immigrants escape from isolation. https://www.herald.co.zw/covid-19-positive-immigrants-escape-from-isolation

WebMD (n.d.). Epidemics, pandemics, and outbreaks. https://www.webmd.com/cold-and-flu/what-are-epidemics-pandemics-outbreaks

Wise, A. (2020, April 17). Military metaphors distort the reality of COVID-19. *Scientific American*. https://blogs.scientificamerican.com/observations/military-metaphors-distort-the-reality-of-COVID-19

Yin, R. K. (2009). *Case study research: Design and methods.* (4th ed.). SAGE.

Factors influencing post-hackathon project continuation in an African corporate setting

Zanele Ratsoga
Non-Resident Fellow, Institute for Pan-African Thought and Conversation (IPATC), University of Johannesburg
https://orcid.org/0000-0002-3388-8689

Mpho Primus
Associate Professor and Deputy Director, Institute for Intelligent Systems, University of Johannesburg
https://orcid.org/0000-0002-5270-3619

Abstract
This article reports on a study examining the factors influencing post-hackathon project continuation in a company with presence in several African countries. The research was conducted as a case study, and focused on hackathon events held by the company between 2018 and 2020. The study identified three core factors that influenced the potential for project continuation after the corporate hackathons: (1) availability of financing; (2) team skills fit and diversity; and (3) degree of project integration into company operations. Where one or more of these elements was insufficiently present, then project continuation became less likely—and the likelihood of project discontinuation increased. The findings are of potential utility to corporate hackathon organisers seeking to increase the levels of project continuation—and, by, extension, return on investment—from their companies' hackathon projects.

Keywords
corporate hackathons, hackathon projects, project continuation, project discontinuation, financing, skills, project integration

Acknowledgement
This article draws on findings from first-listed author Ratsoga's research in the Master of Management in Digital Business programme, University of the Witwatersrand (Wits), Johannesburg.

DOI: https://doi.org/10.23962/ajic.i31.15564

Recommended citation
Ratsoga, Z. & Primus, M. (2023). Factors influencing post-hackathon project continuation in an African corporate setting . *The African Journal of Information and Communication (AJIC), 31*, 1-20. https://doi.org/10.23962/ajic.i31.15564

1. Introduction

Organisations operate in competitive environments and face constant pressure to provide innovative solutions in order to attract and retain customers. Innovation is the application of new technology (and information) to create a new product or service, requiring knowledge that has not previously been utilised to create the product or service in question (Afuah, 2020). The majority of innovations inside a firm or industry are the product of deliberate, purposeful searching for new ideas, which can be discovered through unexpected events, process demands, as well as industry and market developments (Drucker, 2002). Corporate hackathons have become one of the strategies that firms use to speedily develop prototypes which can turn into solutions providing value to customers (Herbsleb et al., 2018).

Hackathons are time-limited events centred around problem-solving and computer programming, with a focus on showcasing prototypes of digital innovations (Cobham et al., 2017). Corporate hackathons are aimed at generating innovative technological ideas, with the ultimate goal being to continue projects beyond the hackathon stage and to launch products that meet customer needs. Despite the many positive views on the usefulness of corporate hackathons, concerns have also been voiced about the lack of economic exploitation of many hackathon outcomes—with questions raised regarding the return on investment, in terms of finances and time, that firms receive from organising these events (Kohne & Wehmeier, 2020; Komssi et al., 2015; Medina & Nolte, 2020; Nolte et al., 2018, 2020).

A key potential return on investment for a corporate hackathon is project continuation, i.e., when participants in the hackathon are able to take forward the innovations they generated in the hackathon after the event. Accordingly, this study explored the factors that lead to post-hackathon project continuation, or non-continuation, within a corporate setting.

This article is organised as follows: Section 2 provides the background, including the literature review, on corporate hackathons and project continuation; section 3 presents the study design; section 4 presents the findings; section 5 discusses and analyses the findings; and section 6 presents conclusions and suggestions for future research.

2. Background and literature review

The word "hackathon" is a fusion of two words: "hack" and "marathon". It resembles a marathon in that effort is focused; but, in this case, the focus is on developing software to bring about a solution (Kohne & Wehmeier, 2020; Komssi et al., 2015). Hackathons are time-limited, problem-oriented computer programming events showcasing prototypes of digital innovation, where people meet and form teams (Lifshitz-Assaf et al., 2021). Each team tries to complete its chosen project of interest, usually within two to five days (Briscoe & Mulligan, 2014; Herbsleb et al., 2018). Hackathons connote the positive sharing of knowledge and the development of a prototype in a time-constrained and pressurised environment (Kohne & Wehmeier, 2020; Komssi et al., 2015; Pe Than et al., 2020; Raatikainen et al., 2013; Soltani et al., 2014). They encourage collaboration, creativity, and quick prototyping with the goal of generating new ideas and solutions (Briscoe & Mulligan, 2014). While hacking is sometimes associated with negativity and cybercrimes, in a hackathon setting it refers to an experimental programming event (Flus & Hurst, 2021).

Hackathons began in technology and programming communities with participants focused on creating software applications through coding in marathon-like settings (Briscoe & Mulligan, 2014; Falk et al., 2021). However these innovation-driven events can be conducted in either corporate, civic, or educational settings (Kienzler, 2016; Komssi et al., 2015; Leclair, 2015). The hackathon typology can also be categorised as innovation-searching of a community-nurturing, issue-oriented, or catalytic nature (Medina & Nolte, 2020). Significant adoption of hackathons emerged in the mid- to late 2000s in the form of competitive events that focused on driving product innovation, with higher prevalence at large technology companies and in university computing environments (Herbsleb et al., 2018).

Hackathons provide numerous short-term benefits for the organisers and participants. They create a platform for problem-solving, idea generation, and prototyping within a short period (Briscoe & Mulligan, 2014; Falk et al., 2021; Komssi et al., 2015). They encourage learning, collaboration, and networking among participants (Briscoe & Mulligan, 2014; Soltani et al., 2014; Tang & Vezzani, 2017). They also stimulate creativity and encourage a culture of innovation (Herbsleb et al., 2018; Kohne & Wehmeier, 2020). Even while these short-term benefits may be quite clear, the long-term impacts and sustainability of hackathon projects are less certain and require interrogation (Medina & Nolte, 2020; Nolte et al., 2020).

Corporate hackathons

Companies in this highly competitive age have been forced to seek innovative ways to decrease the duration of product development cycles (Herbsleb et al., 2018). Corporate hackathons are therefore designed to deliver innovation, new products, and new ventures, and corporates set aside considerable budgets to fund these initiatives as sources of competitiveness (Kitsios & Kamariotou, 2019). Through hackathons,

ideas can be generated, and revenue-making prototypes developed, which answer fundamental problems facing organisations (Komssi et al., 2015). Nolte et al. (2020) report that roughly one-third of all hackathons occur in the corporate space.

Studies have shown that corporate hackathons can encourage effective idea generation, problem-solving, and cross-functional collaboration within an organisation (Herbsleb et al., 2018; Kohne & Wehmeier, 2020). The success of corporate hackathons should be assessed not only on the basis of immediate results, but also on the extent to which they meet the challenge of converting hackathon projects into actual products or initiatives that can deliver value to an organisation (Leemet et al., 2021; Nolte et al., 2020; Pe Than et al., 2020).

In their systematic literature review of 29 peer-reviewed journal papers and 62 peer-reviewed conference papers, Medina and Nolte (2020) identify two types of tangible outcomes, and five types of intangible outcomes, from hackathons:

- Tangible outcomes
  - technical artefacts (e.g., new prototypes, product features, bug fixes), and
  - non-technical artefacts (e.g., visualisations, new or improved documentation, publications).
- Intangible outcomes
  - learning (e.g., about a hackathon, new technology, new industry),
  - networking (e.g., meeting new people, more opportunities to collaborate),
  - interdisciplinary collaboration (e.g., creative ideation),
  - entrepreneurship, or fostering existing enterprises, and
  - fostering awareness about hackathon theme (Medina & Nolte, 2020).

Medina and Nolte (2020) find that entrepreneurial, long-term, post-hackathon project continuation requires both tangible and intangible outcomes working in tandem.

Post-hackathon project continuation

Post-hackathon project continuation refers to activities occurring after the hackathon event that further develop the hackathon idea or prototype, ideally into a product or start-up that can be launched into the market. Many studies of hackathons have focused on the preparation and operational aspects of the events themselves (Briscoe & Mulligan, 2014; Falk et al., 2021; Komssi et al., 2015; Soltani et al., 2014; Stoltzfus et al., 2017). In contrast, there has been relatively little focus on project continuation and achievement of sustained outcomes—with notable exceptions being the work of Leemet et al. (2021), Medina and Nolte (2020), Nolte et al. (2018; 2020), and Pe Than et al. (2020).

Due to the short and intense nature of hackathons, there is a risk of post-hackathon activities being neglected or abandoned (Flores et al., 2018). Komssi et al. (2015) found that, in the overall assessment of a hackathon event, the organisers, participants, audience members, and stakeholders typically rate the event as quite successful, irrespective of the degree of scaling of ideas generated at the hackathon.

According to an archival analysis of the hackathon database Devpost2, which primarily includes hackathons open for anyone to participate in, only 5% of all types of hackathon projects remained active after five months (Nolte et al., 2020). These low project continuation rates were found to be present regardless of whether the hackathon event was positioned as individual learning, community-building, or entrepreneurial opportunity (Nolte et al., 2020). Findings such as these lead to the argument that hackathon outcomes are typically not sustained beyond the hackathon event (Cobham et al., 2017; Mantzavinou et al., 2018; Trainer et al., 2016).

In their analysis of a hackathon database, Nolte et al. (2020) identified nine elements found to influence project continuation. These are team size, team familiarity, prior hackathon participation, skill matching, skill diversity, preparation activities, continuation intention, project complexity, and winning the hackathon. The Nolte et al. (2020) study also found that having smaller teams, team member familiarity with each other, and prior participation in a hackathon all result in a better chance of post-hackathon project continuation. They further found that long-term project continuation is supported where teams have a good skill fit, diverse skill sets, and a strong intention to continue the project.

Also regarding elements that can influence project continuation, the Nolte et al. (2018) study found that consultations with potential stakeholders before and during the hackathon event create awareness that positively contributes to project placement and financial sponsorships. Also, companies that implement structured post-hackathon activities, such as conducting lessons-learned sessions, assessing ideas, and reflecting on prototypes developed during the event, have been found to generate returns on hackathon investment (Flores et al., 2018; Kohne & Wehmeier, 2020). Further to such activities, determinations then need to be made regarding ideas to be further pursued or discarded, and budget allocations can then be provided for ideas that are assessed to have potential (Komssi et al., 2015).

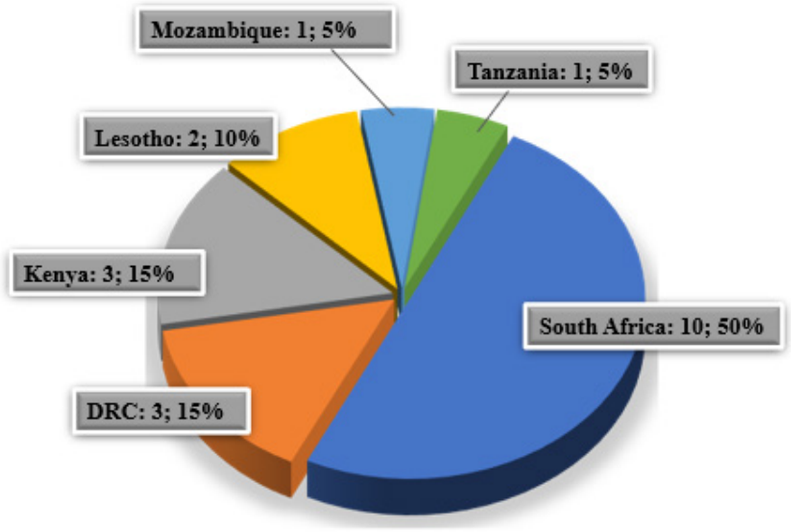
Our study was of an exploratory nature, focused on determining the factors that influence post-hackathon project continuation. Nevertheless, it relied on the baseline provided by the literature review findings, and aimed to contribute to the existing body of knowledge.

3. Study design

Semi-structured interviews were conducted, from November 2021 to February 2022, with 20 hackathon participants across the company's operations in six African countries. Permission was sought and granted by the company's hackathon organisation team to conduct the research and to obtain access to the database of all hackathon participants within the company. Participants were each invited to participate via an email in which the context of the study was laid out. The data was collected using semi-structured interviews targeted at participants who took part in hackathons held from 2018 to 2020. The selection of the participants targeted both instances where there was post-hackathon project continuation and where there was no continuation.

The majority of the participants were based in South Africa, representing 50% of the study sample. This was followed by the DRC and Kenya, each with 30% representation, Lesotho with 10% representation, and Mozambique and Tanzania each with 5% representation, as represented in Figure 1 below.

Figure 1: Number of participants, and percentage of study sample, per country



The South African office of the company chosen for the case study had more employees than the offices in the other African countries, resulting in the South African office having a larger representation of participants who were interviewed. In the study's findings and analysis, the participants' views were found to be similar across all the country locations, i.e., no significant country-specific differences were identified.

An interview guide (see Appendix) was used, containing questions grounded in the literature reviewed as well as other specific questions that we devised. Because the guide was semi-structured, it allowed for probing based on the responses received. As seen in the Appendix, the guide comprised an introductory section with three questions, and then three core sections containing a total of 14 questions. The first section gathered information on the participants: length of service in the company, position in the company, and number of hackathons participated in. The remaining three sections focused on:

- motivations for participation, and project continuation intent;
- project continuation; and
- project discontinuation.

The semi-structured interviews were conducted over Microsoft Teams and audio-recorded, with the participants first consenting to the recording. Each recording was transcribed word-for-word and then subjected to thematic analysis. The identified themes were mapped and interpreted according to the research focus on the factors influencing post-hackathon project continuation or discontinuation.

4. Findings

Status of the 20 hackathon projects

Each of the 20 interview participants was involved in a separate hackathon project, meaning that the study examined the dynamics of post-hackathon continuation and discontinuation with respect to 20 projects. One of the questions in the interview guide asked whether the project was:

- implemented ("completed" in the wording of the interview guide);
- in progress; or
- abandoned.

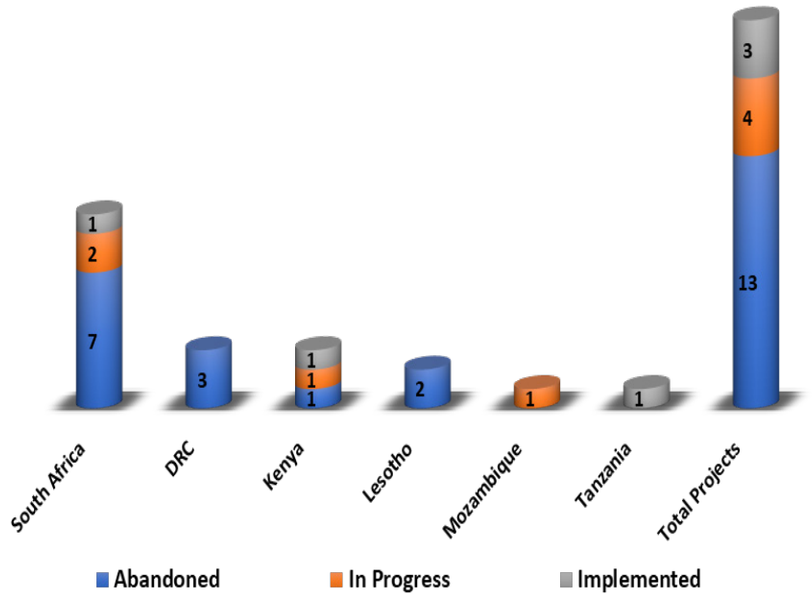
Table 1 represents the status of the 20 participants' projects, as well as each participant's country and hackathon year.

Table 1: Status of the 20 hackathon projects as at February 2022

Participant	Project status	Country	Year of hackathon
Participant 1	abandoned	South Africa	2019
Participant 2	abandoned	South Africa	2019
Participant 3	abandoned	South Africa	2019
Participant 4	abandoned	South Africa	2019
Participant 5	implemented	South Africa	2018
Participant 6	abandoned	South Africa	2019
Participant 7	abandoned	South Africa	2019
Participant 8	in progress	South Africa	2020
Participant 9	abandoned	South Africa	2019
Participant 10	in progress	South Africa	2020
Participant 11	abandoned	DRC	2019
Participant 12	abandoned	DRC	2020
Participant 13	abandoned	DRC	2019
Participant 14	in progress	Mozambique	2020
Participant 15	implemented	Kenya	2020
Participant 16	abandoned	Kenya	2020
Participant 17	in progress	Kenya	2020
Participant 18	abandoned	Lesotho	2020
Participant 19	implemented	Tanzania	2020
Participant 20	abandoned	Lesotho	2020

Among the 20 projects, it was found that 13 (65%) had been abandoned, four (20%) were still in progress, and three (15%) were already implemented. Figure 2 represents the project status by country.

Figure 2: Project status by country (at the time of each interview)



Motivations for hackathon participation

The motivations for hackathon participation that were most frequently cited by the interview participants were:

- to develop solutions;
- to develop something that could be implemented by the company;
- to contribute to the company's culture of innovation;
- to be part of the hackathon, which was something new;
- to have fun;
- to win the hackathon and be rewarded with money.

These motivations are reflected in the following participant statements:

> **Participant 2**: "To really sell a solution that is alternative to day-to-day thinking […] and creating a culture of innovation in my own team. So that was really the driver."
> **Participant 3**: "To be [a] solution provider [for] the issues that we as the Fraud Department had at that time."
> **Participant 15**: "[…] the culture of innovation, the culture of experimenting with new ideas, and the culture of using technology to solve everyday problems."

Participant 19: "My intention, basically, [was that] I wanted to have something which I can see working in the market and impacting the lives of people but also bringing revenue to company."
Participant 14: "Seeing that idea winning, and the people liking that idea, and this idea being implemented within the organisation."
Participant 3: "That first time experience and the vibe and interacting with other people."
Participant 5: "It was the first time the organisation was hosting a hackathon, so I just also wanted to be part of the hackathon. I did not really know what the hackathon was all about, but I just wanted to be part of it."
Participant 9: "To have fun and see if we could win the prize money."
Participant 3: "Win the hackathon and [see] our idea being implemented within the organisation."
Participant 16: "I think the prize money was of course a big motivation."
Participant 18: "Of course money held more weight as an incentive than the other things, because I mean, hey, we live in a world where money moves everything, so it was one of the things that ticked my box to enter the hackathon."

Project continuation intentions

As seen above in the findings on motivations for hackathon participation, several of the interviewees had clear project continuation intentions when entering the hackathon. Across the interviews, the key project continuation intentions were:

- introduction of a new, or enhancing an existing, business process/product;
- having the new or enhanced offering implemented in the company; and/or
- seeing the new enhanced offering being "live in the market".

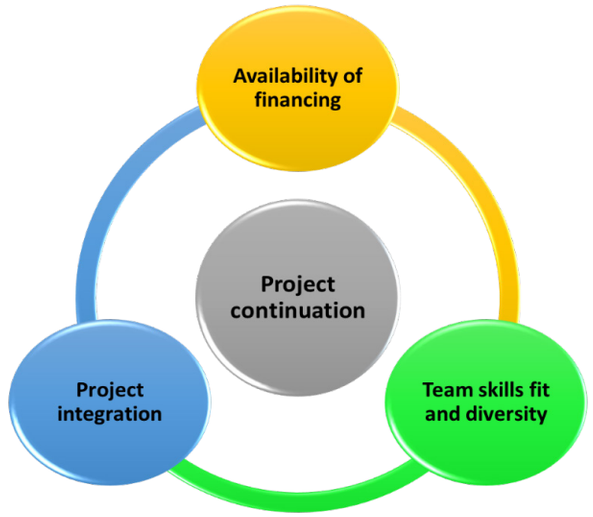
The following are additional examples of participant statements on project continuation intent:

Participant 1: "The intention was to improve the current processes to a better process that will benefit both users and customers."
Participant 7: "Improve customer experience by improving the business process, and also, if we improve the customer experience by streamlining the business process, we will therefore avoid unnecessary calls coming into the call centres."
Participant 2: "Seeing our idea being implemented within the organisation."
Participant 11: "Bring our project to life."
Participant 16: "Seeing [our] product live in the market."

Factors influencing project continuation

In the participant interviews, the three factors most frequently cited as influencing whether or not a hackathon project would continue post-hackathon were: (1) availability of financing; (2) team skills fit and diversity; and (3) degree of project integration into company operations. Figure 3 provides a visualisation of the three main factors found to influence project continuation.

Figure 3: Factors influencing post-hackathon project continuation



Availability of financing

In order for their projects to continue post-hackathon, the participants were required to source funds that could be allocated to their projects, as the organisation had not made funding automatically available for post-hackathon project continuation. Several participants stated that it was challenging to continue working on their projects without an allocated budget, with the projects typically being placed on a backlog list while the focus was on other projects.

Participant 2: "The cost implication you're going to be carrying on your own, and it became clear that the executives were more interested in investing in those hackathon ideas that had won and took ownership on those."
Participant 8: "The item was in the backlog [of company initiatives], but we could not get traction because the focus was on other things."

> **Participant 7**: "There was no budget from the hackathon team to have us seconded […] or any support to help us influence and implement a solution. In literal terms it meant our efforts were at a standstill."
> **Participant 8:** "It would have been great if there was a budget support because I tell you what, that would have accelerated our efforts. We would have scaled up very quickly."
> **Participant 10**: "It would have been a lot easier if there was separate budget for more resources allocation to look at some of the technical challenges and also some of the business requirements analysis and things like that."

Team skills fit and diversity

Several interview participants spoke of the importance, for post-hackathon project continuation, of having selected a team with appropriate skills fit and diversity for the project both during and after the hackathon. Some participants said they discovered that while they may have had the appropriate skills for the project at the hackathon stage, the same team was not fit for post-hackathon project continuation.

> **Participant 6**: "We would have needed a broader team to take the idea through. […] the skills that we had were sufficient for the hackathon, but in terms of implementation, we needed other people involved."
> **Participant 13**: "I think that we underestimated the challenge of the implementation of our own idea."
> **Participant 17**: "Actually I think it was luck or something, but I think I got the best team for [the hackathon]. Now, for implementation, probably I will say [I] just need an additional resource."

One participant, participant 19, said their team was able to successfully pursue project continuation primarily because the implementers were part of the hackathon team:

> **Participant 19**: "Taking it live was very simple, because the same people whom I picked were already working within the same systems."

Project integration

Many interview participants pointed to the importance, for project continuation, of finding ways to ensure that their projects could be integrated into the company's systems and processes. The participants highlighted the need for their direct leadership to support and prioritise the continuation of their hackathon projects, and the challenges posed when there was an absence of such support and prioritisation. The participants spoke of difficulties in getting prioritisation for their projects,

resulting in some of the projects being abandoned or taking longer than anticipated to progress to implementation. The following are some of the statements made:

> **Participant 6**: "Had our manager been also keen on it, there would have been more fight from them in terms of fighting with the other systems to make sure that this thing goes in into production. So, to sum it up, I would say management […] should be as committed as the people that were involved in the hackathon about the idea."
> **Participant 10**: "Prioritisation for the project was an issue because there were other high-impact projects that took precedence."
> **Participant 13**: "I think we could have had more support in terms of accompanying us to implement the project. So the one thing is the pressure on ExCo [Executive Committee] level. Another thing is support to us, to the team, for each level of implementation of the project, which wasn't done."
> **Participant 13**: "[At] ExCo level, it's not a problem of understanding, but it's a problem of priority. They just did not push for it because they did not feel the pressure to do so."

Participant 8 said that the prioritisation success achieved by their project was due to the fact that the company's Chief Risk Officer (CRO) had an interest in it:

> **Participant 8:** "Our idea had a lot of buy-in from our ME [Management Executive]. To start with, he bought into the idea. It also had a buy-in from the CRO, the head of Fraud and Risk. The fact that it had buy-in from that higher-up meant that it received a lot of impetus and momentum, so we could […] proceed."

Some participants highlighted the problem of similar ideas co-existing within the company, and these overlaps only being discovered once project continuation had commenced. For example, participant 5's team discovered that another team within the company had already progressed with a project with similar elements. Ultimately, the teams collaborated towards project continuation, which resulted in a completed project.

> **Participant 5**: "So when we presented the idea, we realised there was another team that had sort of the similar ideas […] although not exactly the same. So, bits and pieces of our ideas. We took some of the ideas from the other team and we sort of fused them together."

However, for participant 9's team, overlap with an existing project meant that their project became unnecessary and had to be abandoned.

> **Participant 9**: "[A project similar to ours was] already being started by a different group that we weren't aware of, and they actually finished it."

Participants also spoke of project integration problems posed by lack of readiness in the company to work with certain new technological elements.

> **Participant 2**: "There was no appetite to explore blockchain, or invest in any blockchain, because of a lack of understanding and no sort of use cases that [were] known before."
> **Participant 3**: "The technology [used by our project], at that time, was new to [the] organisation. There were not a lot of systems that were running on the new technology. So basically, to the company [our project] was […] going to be a Capex expense."
> **Participant 8:** "So there [were] some architectural [elements] that were promised to be in existence that we needed to conform to, and as development went by, we quickly discovered that [the technology] stack was not ready."
> **Participant 14**: "[Our project required an] enabling or creating [of] capability within the team for machine-learning. […] maybe after this capability [is] in place within the organisation, [our] idea can be implemented."

Participant 19, whose team's project had been partially implemented, spoke of the difficulty of integrating into both new and old products.

> **Participant 19**: "The way the project was, we wanted to touch […] several products. We were planning to add value [to almost] all products [the company has], but now the challenge was how to integrate with some old products. So, we had to start with only a few. I think we started dealing with like three products out of like 20-plus products. Which we have now. So, due to integration challenges we had to start with those products which were straight away and simple to do."

Another barrier to project integration at the company, according to participant 7 (whose project had been abandoned), was the absence of centralised repositories for

sourcing the latest data and process documentation necessary to enable hackathon teams to confidently pursue project continuation.

> **Participant 7**: "There's a lack of benchmarked, centralised systems, which makes it very difficult to source the right type of data. You might think you have the right type of data, but […] it probably is not the latest updated data or the most comprehensive data that covers the problem conclusively. […] Documentation of those systems and processes is very lacking, right, so it makes innovation a bit hard because you have to dig really deep. You literally cannot trust that the process that you're looking at is the correct flow of data."

Participant 10, whose team's project had successfully continued beyond the hackathon and was still in progress, said his team's success was due to intentionally re-using existing technology capabilities.

> **Participant 10**: "[We re-used] some of the existing capabilities […] existing technology, in terms of how the model gets built and exposed. So [we are] just re-using technology that's been used on other use cases."

Participant 19, whose team's project had been partially implemented, said the key to their success was adding value to something that already existed, not proposing something entirely new.

> **Participant 19**: "[Our project adds] value to existing products so that customers can be attracted to […] purchase the product with the extra value added."

5. Discussion and analysis

We now discuss and analyse some of the key lessons that emerged from this study's findings.

Availability of financing

Budgets need to be made available for the deployment of corporate hackathon projects that require integration into a company's systems, especially in the case of hackathon projects that involve the use of technology in building prototypes. In the company studied, the hackathon organisers did not allocate any funding to the projects, but rather left it to the participants to seek funding through their own efforts. The hackathon organisers thus anticipated that participants' business units

would budget for the deployment of these projects. Consequently, where there was an absence of budget allocations, the projects could not continue after the hackathon and were ultimately abandoned.

In the company that we studied, there was no indication that systematic determinations were made regarding which projects should be pursued. Where the teams had consulted potential stakeholders within the company before and during the hackathon, those projects tended to find financial support and opportunities for integration within the organisation. Therefore, project continuation was greatly dependent on whether the team had secured sponsorships prior to the hackathon event.

Team skills fit and diversity

For a corporate hackathon project to have potential for success, its team needs to possess the skills appropriate to, and sufficiently diverse for, both the hackathon itself and the subsequent process of taking the project forward within the company. In the case studied, the value of appropriate skills was made particularly evident by the experience of the hackathon team that was made up of system developers who were already working on the systems to be impacted by the hackathon project—making project continuation after the hackathon event quite straightforward. This study's findings on the importance of skills fit and diversity align with the conclusions reached in the Nolte et al. (2020) study mentioned above.

Project integration

In this study, it was found that projects requiring new technology tended to be discontinued due to the realisation that the organisation was not able, or willing, to invest in a new untested technology with unknown business cases. The projects that were most easily continued were those adding value to existing products or re-using existing technology with known use cases. These findings align with conclusions drawn from studies that found that teams who had built prototypes that could be incorporated into current products had a strong likelihood of continuing to work on their projects after a hackathon (Herbsleb et al., 2018; Komssi et al., 2015). This study found that project integration also relied to a great extent on management commitment. Many participants found that, for a variety of reasons, management commitment was difficult to secure. These findings on the importance of corporate commitment to project continuation align with what Leemet et al. (2021) found in their study cited above.

6. Conclusions and potential future research directions

This study's key contribution to the body of knowledge on corporate hackathons is its identification of three core factors influencing post-hackathon project continuation: (1) availability of financing; (2) team skills fit and diversity; and (3) degree of project integration into company operations. Where one or more of these elements is weak, then the chances for project continuation decrease—and, by extension, the likelihood of discontinuation increases. These findings can potentially empower corporate hackathon organisers who seek to increase return on investment on hackathons through increased levels of continuation—leading to successful implementation—of hackathon projects.

This study's findings also suggest potential directions for future research that could be of use to corporate hackathon organisers. One area that could benefit from future research is the relationship between hackathon participants' continuation intentions and their actual continuation behaviour. A deeper understanding of this relationship could enable hackathon organisers to identify hackathon participants who, with the right support, would be most likely to continue the necessary work to deliver on the promise of a functioning product or service solution. Another area that could benefit from study would be the feasibility of requiring all hackathon teams to have, in advance of the hackathon, an in-principle commitment from an internal organisational business unit, prior to the event, that the project will receive priority appraisal, after the hackathon, regarding its potential for development for integration into the company's operations (Falk et al., 2021).

References

Afuah, A. (2020). *Innovation management: Strategies, implementation, and profits* (2nd ed.). Oxford University Press.

Briscoe, G., & Mulligan, C. (2014). Digital innovation: The hackathon phenomenon. *Information, Communication & Society*, *17*(3), 329–347.

Cobham, D., Gowen, C., Jacques, K., Laurel, J., & Ringham, S. (2017). From appfest to entrepreneurs: Using a hackathon event to seed a university student-led enterprise. In *INTED2017 Proceedings* (pp. 522–529). https://doi.org/10.21125/inted.2017.0265

Drucker, P. F. (2002). The discipline of innovation. *Harvard Business Review*, *80*(8), 95–102.

Falk, J., Kannabiran, G., & Hansen, N. B. (2021). What do hackathons do? Understanding participation in hackathons through program theory analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3411764.3445198

Flores, M., Golob, M., Maklin, D., Herrera, M., Tucci, C., Al-Ashaab, A., Williams, L., Encinas, A., Martinez, V., Zaki, M., Sosa, M., & Pineda, K. (2018). How can hackathons accelerate corporate innovation? Paper presented at IFIP WG 5.7 International Conference, APMS, Seoul. https://doi.org/10.1007/978-3-319-99704-9_21

Flus, M., & Hurst, A. (2021). Design at hackathons: New opportunities for design research. *Design Science*, *7*, E4, 1–24. https://doi.org/10.1017/dsj.2021.1

Herbsleb, J., Pe Than, E. P. P., Nolte, A., Filippova, A., Bird, C., & Scallen, S. (2018). Designing corporate hackathons with a purpose. *IEEE Software*, *36*, 15–22. https://doi.org/10.1109/MS.2018.290110547

Kienzler, H. (2016). Learning through inquiry: A global health hackathon. *Teaching in Higher Education*, *22*(4), 455–471. https://doi.org/10.1080/13562517.2016.1221805

Kitsios, F., & Kamariotou, M. (2019). Beyond open data hackathons: Exploring digital innovation success. *Information*, *10*(7), 235, 2–10. https://doi.org/10.3390/info10070235

Kohne, A., & Wehmeier, V. (2020). *Hackathons: From idea to successful implementation* (1st ed.). Springer International. https://doi.org/10.1007/978-3-030-58839-7_1

Komssi, M., Pichlis, D., Raatikainen, M., Kindström, K., & Järvinen, J. (2015). What are hackathons for? *IEEE Software*, *32*, 60–67. https://doi.org/10.1109/MS.2014.78

Leclair, P. (2015). Hackathons: A jump start for innovation. *Public Manager*, *44*(1), 12–15.

Leemet, A., Milani, F., & Nolte, A. (2021). Utilizing hackathons to foster sustainable product innovation –the case of a corporate hackathon series. In *2021 IEEE/ACM 13th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)* (pp. 51–60). https://doi.org/10.1109/CHASE52884.2021.00014

Lifshitz-Assaf, H., Eden, L., & Gukov, I. (2021). The experience of organizing global hackathons: From the perspective of innovation intermediaries. *Journal of Business Venturing Insights*, *15*, e00254.

Mantzavinou, A., Ranger, B., Gudapakkam, S., Hutchins, K., Bailey, E., & Olson, K. (2018). Health hackathons drive affordable medical technology innovation through community engagement. In S. Hostettler, S. Najih Besson & J. C. Bolay (Eds.), *Technologies for development* (pp. 87–95). https://doi.org/10.1007/978-3-319-91068-0_8

Medina, M., & Nolte, A. (2020). What do we know about hackathon outcomes and how to support them? – A systematic literature review. In *Collaboration technologies and social computing: 26th International Conference, CollabTech 2020, Tartu, Estonia, September 8–11, 2020, proceedings* (pp. 50–64). Springer.

Nolte, A., Chounta, I.-A., & Herbsleb, J. D. (2020). What happens to all these hackathon projects? Identifying factors to promote hackathon project continuation. In *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 145, 1–26. https://doi.org/10.1145/3415216

Nolte, A., Pe Than, E. P. P., Filippova, A., Bird, C., Scallen, S., & Herbsleb, J. (2018). You hacked and now what? – Exploring outcomes of a corporate hackathon. In *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW) (pp. 1–23). https://doi.org/10.1145/3274398

Pe Than, E. P. P., Nolte, A., Filippova, A., Bird, C., Scallen, S., & Herbsleb, J. (2020). Corporate hackathons, how and why? A multiple case study of motivation, projects proposal and selection, goal setting, coordination, and outcomes. *Human–Computer Interaction*, *35*(3), 193–235. https://doi.org/10.1080/07370024.2020.1760869

Raatikainen, M., Komssi, M., Bianco, V., Kindström, K., & Järvinen, J. (2013). Industrial experiences of organizing a hackathon to assess a device-centric cloud ecosystem. In *2013 IEEE 37th Annual Computer Software and Applications Conference (COMPSAC 2013)* (pp. 790–799). https://doi.org/10.1109/COMPSAC.2013.130

Soltani, P., Pessi, K., Ahlin, K., & Wernered, I. (2014). Hackathon – a method for digital innovative success: A comparative descriptive study. In *Proceedings of the 8th European Conference on Information Management and Evaluation (ECIME 2014)* (pp. 367–373).

Stoltzfus, A., Rosenberg, M., Lapp, H., Budd, A., Cranston, K., Pontelli, E., Oliver, S., & Vos, R. A. (2017). Community and code: Nine lessons from nine NESCent hackathons. *F1000Research*, *6*, 1–20. https://doi.org/10.12688/f1000research.11429.1

Tang, T., & Vezzani, V. (2017). *Fostering a culture of collaboration through playful design jams.* In *DIS '17 companion: Companion publication of the 2017 ACM Conference on Designing Interactive Systems, Funchal, Madeira Island, Portugal.* https://doi.org/10.1145/3064857.3079153

Trainer, E., Kalyanasundaram, A., Chaihirunkarn, C., & Herbsleb, J. (2016). How to hackathon: Socio-technical tradeoffs in brief, intensive collocation. Paper presented at CSCW'16: Computer Supported Cooperative Work and Social Computing, San Francisco. https://doi.org/10.1145/2818048.2819946

Appendix: Interview guide

Section A: Participant information

a) How many years have you worked in [name of company]?
b) What is the division/functional area and country that you work in?
c) How many hackathon events have you participated in at [name of company]?

Section B: Project continuation intent

1. What was your motivation and intention to participate in the hackathon?
2. How did you and your team prepare for the event, including systems or processes the project would be impacted by?
3. Did you have intentions of working on the project post the event? Why?

Section C: Post-hackathon activities – project continuation

4. Did the organisers provide any guidelines or terms of project continuation? What were those?
5. Do you believe your team had the appropriate skill fit and diversity for the chosen project? Why?
6. Did you encounter any challenges while working on the project? (system/people/processes/budget constraints?)
7. What are the specific activities that contributed to project continuation?
8. Were you required to provide regular project progress status to the organisers?
9. Is the project completed or still in progress or abandoned?
10. If abandoned, what other challenges did you experience?

Section D: Post-hackathon activities – project discontinuation

11. What is the reason for not continuing?
12. What challenges did you encounter that discouraged project continuation? (system/people/processes/budget constraints)
13. Do you feel that there's something that could have been done to encourage project continuation?
14. Will you be participating in the next event? Why?

Evaluation of web-based online agricultural information relevant to Tanzanian maize producers

Ester Ernest Mnzava
Lecturer, Sokoine University of Agriculture, Morogoro, Tanzania
https://orcid.org/0000-0002-3312-3635

Lorette Jacobs
Professor and Chair, Department of Information Science, University of South Africa (UNISA), Pretoria
https://orcid.org/0000-0001-8081-7739

Abstract
This study examined the quality of web-based online agricultural information relevant to the maize industry in Tanzania. Selected online sources were evaluated to assess the agricultural information available in terms of four dimensions of quality, namely: *authority*, *completeness*, *timeliness*, and *understandability*. The study identified a wide variety of web-based online information on maize production, including information on seeds, fertilisers, pesticides, and grain-handling. It was found that the information was of variable quality. Among the 39 online sites studied, several lacked contact information, had outdated content, and contained information that was missing some important details, and none provided weather information. This study contributes to the body of knowledge on online agricultural information in an African context where the agricultural sector is central to national economic development. The online agriculture information evaluation tool used in the study can potentially be of use, in its current form or adapted, to researchers beyond Tanzania.

Keywords
agricultural information, online resources, maize production, authority, completeness, timeliness, understandability, Tanzania

DOI: https://doi.org/10.23962/ajic.i31.15944

Recommended citation
Ernest Mnzava, E. & Jacobs, L. (2023). Evaluation of web-based online agricultural information relevant to Tanzanian maize producers. *The African Journal of Information and Communication (AJIC), 31*, 1-21. https://doi.org/10.23962/ajic.i31.15944

1. Introduction

Advances in information and communication technology (ICT) have created an environment that facilitates the production, exchange, and sharing of information via online platforms (Janc et al., 2019), including extensive information related to agricultural production (Food and Agriculture Organisation of the United Nations (FAO) & International Telecommunication Union (ITU), 2022). Numerous types of agricultural information are available, including information related to pest control, markets, prices, credit, loans, crop storage, inputs, agronomic practices, and agricultural innovations (Adio et al., 2016; Ogessa & Sife, 2017). Whilst much online agricultural information is useful, some is inaccurate and thus useless (Barau & Afrad, 2017; Muhammad et al., 2020), undermined by an absence of editorial review to validate content quality (Verkijika & De Wet, 2018). Variations in the quality of agricultural information are of great concern because information quality in this context affects livelihoods—by influencing individuals' ability to make decisions pertaining to their farming activities in relation to the seed variants to plant, weather patterns, maturity, fertiliser use, and the marketing of agricultural products (Kante et al., 2019).

Although the quality of agricultural information is known to impact agricultural productivity (Kante et al., 2019; Ndimbwa et al., 2020; Wilson & Lewis, 2015), little is known about the quality of different online agricultural information sources that are needed to support decision-making related to agricultural activities. Research has mostly focused on access and use, types of information sources, and the information-seeking behaviour of the farmers (Holt-Day et al., 2020; Magesa et al., 2020; Nkebukwa, 2018; Tumbo et al., 2018; Uwandu et al., 2018).

Despite the necessity of ensuring that reliable and recent online information for agriculture is accessed and shared, the information found on online sources related to agricultural topics is not validated to ensure that only accurate information is shared with stakeholders within the maize industry in Tanzania (Barau & Afrad, 2017). Research conducted by Ndimbwa et al. (2020) in Tanzania's Kyela District revealed that efforts by smallholder farmers to increase maize production were undermined by the inadequate quality of available online agricultural information. Given these weaknesses, this research examined the quality of online agricultural information intended to support Tanzanian maize producers in respect of seed selection, fertiliser application, pesticide use, and grain handling. We developed an online agricultural information evaluation tool, which was used to guide the evaluation of the selected online resources in terms of four information quality dimensions: *authority*, *completeness*, *timeliness*, and *understandability*.

2. Literature review and analytical framework

Information quality is a multidimensional concept that cannot be expressed through a single definition or even through the dimensions that measure it (Al-Hakim, 2007). This is because the term "quality" has no unified definition. Considering the

value of information in knowledge-intensive products and processes, Eppler (2006) cites two frequently used definitions of quality. The first explains quality as including all the features of a product or service that describe its ability to satisfy a given need, while the second describes quality as the extent to which the expectations of the user of a particular product or service are met or exceeded. These definitions set out by Eppler (2006) reveal two main components: a subjective component, as in meeting expectations; and an objective component, as in meeting requirements.

Abumandil and Hassan (2016), in a study in the banking sector, found that authority, completeness, timeliness, and interpretability are the core measures for assessing quality of information. Laumer et al. (2017), in studying enterprise content management systems, find that information quality should be contextualised through two measures: representation (characteristics of information conciseness, namely authority, understandability and presentation) and contextual application (degree of fitness for use in a task, namely timeliness and completeness). Within the context of health studies, Tao et al. (2017) consider accessibility, timeliness, understandability, and authority as the key dimensions to be examined to assess the quality of information available in online sources. Considering the dimensions of quality related to websites, Tate (2019) states that authority, accuracy, objectivity, timeliness, completeness, and the intended audience are the fundamental dimensions to be heeded when measuring the quality of information.

Table 1 summarises information quality dimensions found in some of the core literature, with an indication of which dimensions are foregrounded in which items of literature.

Table 1: Dimensions of information quality

Dimension	Abumandil & Hassan (2016)	Laumer et al. (2017)	Tao et al. (2017)	Tate (2019)
Authority	√	√	√	√
Completeness	√	√	√	√
Understandability		√	√	√
Timeliness		√	√	√
Accuracy	√			√
Interpretability	√			
Objectivity				√
Accessibility			√	
Presentation		√		

From the matrix in Table 1, the four dimensions that we decided to study were authority, completeness, timeliness, and understandability. A dimension cited in

all four pieces of literature featured in Table 1, namely, *authority* of information, is determined by the degree to which the reputation of a person or an organisation responsible for an online source is available to users (Tao et al., 2017). Tate (2019) warns that, on the web, anyone can state that they are providing authoritative information. Therefore, assessing the authority of the information on a website requires an extra eye to ensure the authenticity of the creator of the information. For instance, the use of contact information such as email addresses and phone numbers can be a useful indication of the authority of the individual or organisation that presents online information (Gaillard, 2018).

Completeness refers to whether the information contains all necessary details (Al-Hakim, 2007). It is the quality of being whole (Laumer et al., 2017; Tate, 2019). Information is considered complete if it provides factual information that is verifiable (Al-Hakim, 2007; Laumer et al., 2017). In the context of this research, agricultural information is considered relevant for farmers and other stakeholders of the maize industry in Tanzania if it addresses various activities in the maize production value chain. For example, the selection of maize seeds is determined by various factors, including weather, agroecological location, and maturity period (Dao et al., 2015; Temesgen & Kebena, 2019). This suggests that information made available via the web to support farming activities should provide expanded detail to ensure the application and usefulness of the information.

Timeliness is an important factor in agriculture development because every agricultural activity requires information that is current. Folnovic (2021) explains that because weather influences every activity during the growing season, information on the weather must be timely and provided in real time. Other agricultural information is considered timely when created/posted not more than two years before (Naruka et al., 2017). The timeliness of other agricultural information is influenced by the fact that maize is grown once or twice a year, depending on the agroecological zone. Therefore, information related to agricultural technology, practices, and innovations posted annually may be useful in relation to a second planting and harvest (Mbagwu et al., 2018).

Understandability is the degree to which information is easy to comprehend and apply (Al-Hakim, 2007). Readers should be able to understand certain information if they have some knowledge of the given topic. This suggests that understanding

information relates to the way in which it is presented. For instance, agricultural research reports carrying accurate information may not be easily understood by farmers because of the format used to express their content and because of the high level of language used (Lamptey et al., 2017).

Though the other dimensions were not considered for the purposes of this study, they do carry value as they were mentioned by key authors. Briefly, *interpretability* is the degree to which information is presented based on the ability of the consumer to understand it (Al-Hakim, 2007). *Accuracy* reflects the extent to which information is reliable and free from errors (Al-Hakim, 2007; Tate, 2019). It is determined by the professional experience and educational qualifications of the creator/author of the information (Brock University Library, 2023). *Presentation* refers to the way in which information is presented so that it is easily interpreted by a reader (Al-Hakim, 2007). *Accessibility* is the extent to which information is available without limitations related to time, special skills/abilities/technology, or accessibility (Eppler, 2006), so that it can be used by individuals to obtain access to information that will meet maize crop farming information needs.

3. Study design and methodology

A mixed-methods approach was used to determine the information quality of selected web-based online resources. Linked to the pragmatism paradigm, the exploratory sequential design was used to collect qualitative and quantitative data in two different phases. Qualitative data collection and analysis was used to verify the quality dimensions to be used for assessing agricultural information for maize production. Through systematic sampling of web-based online resources related to the Tanzanian agricultural sector and, in particular, the country's maize industry, we identified a sample of online resources. Online resources were identified via the use of keyword-searching. The keyword entries, in both Kiswahili and English, included "*mahindi*" (maize), "maize", "zea mays Tanzania" ("Zea mays" is the cientific name for maize), *"kilimo cha mahindi*" (maize cultivation), "*zao la mahindi*" (maize crop), "*mbegu za mahindi*" (maize seeds), and "fall armyworms maize" (fall armyworms is a maize pest that attacks cereal crops, including maize).

For each word search, 50 search results were included in the sample group, which yielded a total of 500 online resources. Spink et al. (2006) suggest that a 10% sample size should suffice in quantitative research to draw a reputable sample. Using simple

random sampling, every 10th online resource was identified to be part of the sample group. Thus, 50 sites were initially selected to be part of the sample size. These 50 sites were further analysed to ensure that they included information in either, or both, Swahili and English, as official languages of Tanzania, as well as information relevant to the maize value chain. Finally, 39 online resources were selected for inclusion in the study (see Appendix 2).

The information quality of the online resources was assessed, in December 2021, through application of a tool that we developed for this study: an online agricultural information evaluation tool (see Appendix 1). The tool was based on this study's four chosen dimensions of information quality: authority, completeness, timeliness, and understandability.

4. Findings and discussion

This section provides details on the findings that resulted from assessing the 39 selected web-based online resources in terms of the four identified quality dimensions: authority, completeness, timeliness, and understandability.

Quality dimension 1: Authority

Tate (2019) explains that assessing the authority of the information on a website requires that contact information such as email addresses and phone numbers can be checked for accuracy as an indication of the authority of the individual or organisation that presents information. As seen in Table 2, not all 39 online resources complied with the dimensions of authority in terms of (1) providing the names of authors, creators, or contributors; or (2) making contact information available. Table 2 shows that 31 of the 39 online resources included the names of the authors of, or contributors to, the contents. However, only 25 had contact information such as email addresses and phone numbers.

Table 2: Authority of information

Dimension	Attributes	Number of resources that complied (n=39)
Authority	Names of authors, creators, or contributors	31
	Authors', creators', or contributors' contact information (e.g., phone numbers, email addresses)	25

The lack of contact information for the creators of information provided on websites is perceived by authors such as Huff et al. (2015) as a serious limitation, because it suggests that the authors may not be willing to take responsibility for the validity and reliability of the contents shared on a website. Furthermore, the lack of contact details limits the ability of users to obtain access to more detailed content relating to the information posted on a website (Tate, 2019, p. 53). This suggests that contact information in online resources has implications for the users, information providers, and owners of a website. In instances where contact information is available, users are more inclined to trust the quality of information available on websites and to regard the information as authoritative. By allowing users to engage with creators/organisations through the provision of contact details, website owners can receive and benefit from feedback on their websites.

Quality dimension 2: Completeness

Maize production requires different types of information, including information about the weather, seeds, fertilisers, pesticides, and grain handling (Abass et al., 2014; Malozo, 2021; Mattah et al., 2015; Saïdou et al., 2018). The content of the 39 online resources was analysed to determine the completeness of the maize production information that was presented. Components of the maize value chain that were evaluated are presented in Figure 1.

Figure 1: Types of information evaluated

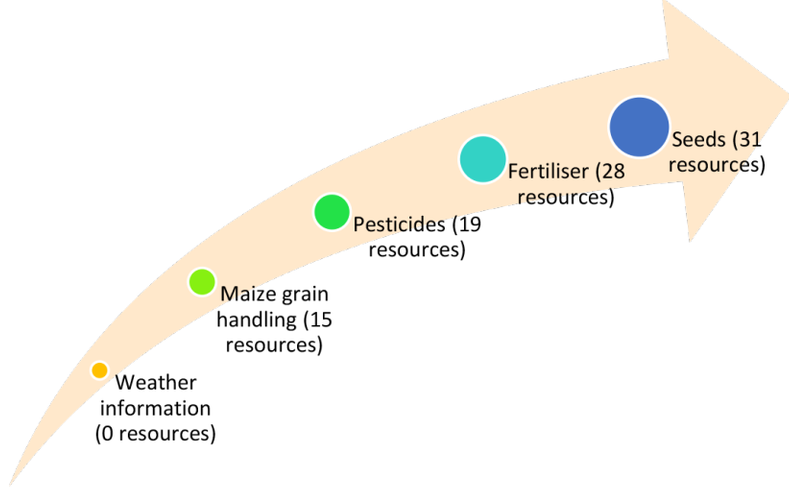


Table 3 sets out in more detail the findings regarding the completeness of the information provided in the 39 online resources.

Table 3: Completeness of information

Information on seeds (n=31)	
Topic	**Number of resources that complied**
1. Variety	29
2. Agroecological zone	12
3. Sowing time	15
4. Seeding rate (kg/ha)	20
5. Planting/sowing spacing	23
6. Maturity period (in days)	14
7. Yield/ha	16
Information on fertiliser (n=28)	
Topic	**Number of resources that complied**
1. Type	28
2. Time/stage of application	16
3. Soil information	4
4. Method of application	14
5. Application rate (kg/ha)	16
6. Interval/frequency of application	10
Information on pesticides (n=19)	
Topic	**Number of resources that complied**
1. Pest type	19
2. Pesticide type	17
3. Plant stage	6
4. Method of application	8
5. Application rate (litres/ha)	6
6. Frequency of application	1
Information on grain handling (n=15)	
Topic	**Number of resources that complied**
1. Cleaning	8
2. Drying	7
3. Storage conditions	6
4. Storage facilities	9

It was evident from the information provided about fertilisers via the online resources that none of the resources had complete information about the type of fertiliser given for maize farming. Most of the online resources lacked information on fertiliser application details (stage of fertiliser application, soil information, method of application, application rate, and frequency). This is undesirable, as inadequate information on fertilisers has negative implications for their use (Abebe & Feyisa, 2017). While authors such as Adiaha and Agba (2016) emphasise the importance of knowledge related to fertilisers and their use in enriching crops with nutrients during the growing period, authors such as Burke et al. (2019) lament the lack of comprehensive web information on the effective use of fertilisers throughout the maize production process in Zambia. The lack of information on fertilisers for maize farmers in Tanzania is of equal concern. A lack of knowledge on the use of fertilisers can negatively affect the environment and humans. Within Tanzania, the over-use of fertilisers can lead to toxins in the water and an imbalance in the crop food chain. Humans can be affected as chemicals presented in the fertilisers can cause cancer and other chronic diseases.

In terms of information on pesticides, information, such as that relating to the method of application and to application rate, was found to be under-represented. This is undesirable because it creates the strong possibility that pesticides will be used improperly. The effectiveness of a pesticide is affected by various factors including pest type, pesticide type, time of application, plant stage, method of application, application rate, and frequency (Damalas & Koutroubas, 2016; Mattah et al., 2015). According to Janc et al. (2019), an awareness of essential pesticide-related information (e.g., on application frequency, method of application, amount of fertiliser to be used, and pest threshold) contributes to the proper use of the pesticides. The lack of such information in online resources referring to pesticides implies that website information on this topic is insufficiently comprehensive. Incorrect pesticide use can affect the soil's physicochemical properties and thus the growth of maize plants. The faulty use of pesticides also negatively influences the calcium, magnesium, potassium, and manganese levels found in maize, leaving the end product without sufficient subsistence for consumption (Jallow et al., 2017; Mubushar et al., 2019). Another possible consequence is that such pesticides, when consumed by animals and humans, lead to imbalances in the animals' nervous systems, and illnesses linked to hormones or the endocrine system in humans.

With respect to the evaluation of the information on the correct handling of maize grains, it was found that the online resources did not provide sufficient detail on drying and cleaning maize grains or on storage conditions and facilities. In respect of the four main types of grain seeds that are used in Tanzania (SC Nduna, SC Shine, SC Sico, and SC Shungu), farmers need to be provided with sufficient detail on the management of these grains, so as to ensure an optimal harvest. On the specific topic of maize-cleaning, detail on the removal of foreign materials such as stones, dirt,

damaged grains, and other crops' grains is needed for use by agricultural extension officers and farmers (Akatuhurira et al., 2021). Also imperative is information on methods for removing moisture so as to avoid the deterioration of maize grains. Sai et al. (2019) explain that Tanzanian farmers can increase crop production by 13% if they know how to remove moisture. Details on how the moisture level in grains can be measured using grain moisture meters can make a valuable contribution in assisting agricultural extension officers to guide farmers on the measurement and control of moisture as a foreign material.

Finally, it was found that none of the reviewed online resources provided any weather information. Jack and Tobias (2017), Hilary et al. (2017) and Mugonya et al. (2021) point to negative impacts on maize farming caused by insufficient information on weather conditions. Information on possible rainfall, the influence of global warming, and temperature fluctuations has a direct and important influence on decisions pertaining to the best seed varieties to use, sowing time, pest control, and processes and procedures for maize grain handling. Msemo et al. (2021) propose that for weather information, specific and specialised online resources should be linked to official resources such as the Tanzania Meteorological Authority (TMA) website. Information provided by an organisation such as the TMA can serve as an authoritative, relevant, and reliable main source of weather information for various activities related to maize crop farming (TMA, n.d.).

Quality dimension 3: Timeliness

As indicated by Mubofu and Malekani (2020), timeliness is one of the essential criteria in determining the quality of an online information resource. In the agricultural sector, access to timely and reliable information contributes to agricultural production and to the adoption of agricultural inputs such as seeds, pesticides, and fertilisers. In evaluating the timeliness of the 39 selected online sources, we found, as seen in Table 4, that the majority—18(46.2%) of the resources—had content that was last created/updated in 2018-19, and that only 13 (33.3%) contained information that had been either created or updated in 2020 or 2021.

Table 4: Timeliness of information in the resources (n=39)

Year of most recent information addition	Number (percentage) of resources
2018-19	18 (46.2%)
2020-21	13 (33.3%)
No date	8 (20.5%)

Naruka et al. (2017) posit that agricultural information on topics such as seed information, pesticides, and fertilisers can be considered up-to-date if it has been created or posted within the previous two years. It was concerning that only one-third of the online resources had content created or updated during the two years leading up to the data collection in December 2021. Also of concern was the fact that eight of the resources did not provide any date of publication or date of information update. Up-to-date information is required by Tanzanian farmers, as farming practices are undergoing extensive transformation. As in many other African countries, Tanzanian farming communities are intensifying their efforts towards the use of modern inputs and improved seeds, in order to improve crop yields. Increased agricultural productivity is seen as a key catalyst for enhanced national employment and for more reliable food access in both rural and urban communities. As expressed by Wineman et al. (2020), agriculture in Tanzania is seen as the "engine" that generates income opportunities. Up-to-date online information aimed at fostering the expansion of agricultural opportunities is therefore imperative.

Quality dimension 4: Understandability

As per the evaluation instrument in Appendix 1, understandability relates to the languages and formats in which online information resources are provided. In terms of language, as shown in Table 5, it was found that 35 (90%) of the online resources used only Kiswahili, 3 (8%) employed both Kiswahili and English, and 1 (2%) used only English. Thus, most of the sites used only Kiswahili, so as to ensure direct access to information for maize industry stakeholders. This is advantageous to farmers in Tanzania, whose first language is Kiswahili (Isayam et al., 2018). Ndimbwa et al. (2020) assert that the language in which information is disseminated influences the extent to which information is usable.

Table 5: Language use in the resources (n=39)

Language(s) used	Number (percentage) of resources
Kiswahili	35 (90%)
Kiswahili and English	3 (8%)
English	1 (2%)

Regarding the information formats used (text, graphics, videos), the evaluation found, as shown in Table 6, that 22 (56.4%) of the resources were text-only, 13 (33.3%) carried a mixture of text and graphics, and 4 (10.3%) were video-only. Numerous scholars, including Flores and Sun (2018), have reported that the use of graphics, or a combination of text, graphics, and videos, is more effective than text alone in communicating meaning.

Table 6: Format use in the resources (n=39)

Format(s) used	Number (percentage) of resources
Text	56.4% (22)
Text and graphics	33.3% (13)
Video	10.3% (4)

What was not assessed, as it fell outside of the scope of the research, was the ability of farmers to access and read the content in the selected online information resources. This brings forth the possibility of further research to determine the internet usage levels and literacy levels of Tanzanian maize production stakeholders.

5. Conclusions

The findings from this study revealed that among the 39 web-based online resources studied, several of the resources lacked contact information, presented outdated content, and/or contained information that was missing some important details—and none provided weather information. These findings have implications for several maize production stakeholder groupings in Tanzania.

One such grouping is the information providers. Included in this stakeholder category are agricultural officers, who are important sources of agricultural information and advice for farmers in Tanzania. For these officers, web-based online resources are part of their communication infrastructure, and relevant and timely information is therefore imperative to ensure that they are able to advise farmers effectively. Also in this information provider category are the managers and content creators responsible for the online agricultural information resources—resources which, as this study has shown, are in many cases in need of substantial improvement. This study's findings are also significant for Tanzanian government departments and legislators, whose agricultural support policies and programmes are dependent on maize producers having access to high-quality information. Finally, and most importantly, there are the targets for the information: the farmers and other participants in the Tanzanian maize production value chain, who require online information resources that are authoritative, complete, timely, and understandable.

The online agricultural information evaluation tool that was developed and used in this study can be of value to all of the stakeholders just mentioned, because it draws attention to four important dimensions of information quality: authority, completeness, timeliness, and understandability. Since, as seen in this study, the quality of online information resources targeted at Tanzanian maize producers is questionable, all relevant stakeholders need to be empowered with information evaluation tools that enable them to identify useful and reliable online agricultural information.

The inadequate quality of web-based online agricultural information related to the Tanzanian maize industry revealed in this study requires further research to investigate how persons or organisations that are responsible for such information can be empowered to improve its quality. In addition, given the rapid advances in non-web-based digital resources accessible through non-web-based mobile apps and platforms providing interactive content to mobile devices, research is needed on the quality of the agricultural information being made available via these means to maize farmers in Tanzania and elsewhere on the continent. Given that mobile phones are the most-used digital tools for information access on the continent, mobile apps can be expected to play a central role in supporting the information needs of the continent's maize farmers. The information quality evaluation tool proposed by this study can be used, with adaptations to make it appropriate to the mobile context, to evaluate the quality of the agricultural information being made available to maize farmers who are reliant on non-web-based mobile platforms.

References

Abass, A. B., Ndunguru, G., Mamiro, P., Alenkhe, B., Mlingi, N., & Bekunda, M. (2014). Post-harvest food losses in a maize-based farming system of semi-arid savannah area of Tanzania. *Journal of Stored Products Research*, *57*, 49–57. https://doi.org/10.1016/j.jspr.2013.12.004.

Abebe, Z., & Feyisa, H. (2017). Effects of nitrogen rates and time of application on yield of maize: Rainfall variability influenced time of N application. *International Journal of Agronomy*, *2017*, 1–10. https://doi.org/10.1155/2017/1545280

Abumandil, M. S., & Hassan, S. (2016). Information quality and decision-making effectiveness: A study of banks in Palestine. *International Review of Management and Marketing*, *6*(7S), 127–132.

Adiaha, M. S., & Agba, O. A. (2016). Influence of different methods of fertiliser application on the growth of maize (*Zea mays* L.) for increase[d] production in south Nigeria. *World Scientific News*, *54*, 73–86.

Adio, E. O., Abu, Y., Yusuf, S. K., & Nansoh, S. (2016). Use of agricultural information sources and services by farmers for improve productivity in Kwara State. *Library Philosophy and Practice*, 1456, 1–16.

Al-Hakim, L. (2007). Information quality factors affecting innovation process. *International Journal of Information Quality*, *1*(2), 62–176. https://doi.org/10.1504/ijiq.2007.014607

Arbuckle Jr, J. G., Morton, L. W., & Hobbs, J. (2015). Understanding farmer perspectives on climate change adaptation and mitigation: The roles of trust in sources of climate information, climate change beliefs, and perceived risk. *Environment and Behavior*, *47*(2), 205–234. https://doi.org/10.1177/0013916513503832

Barau, A. A., & Afrad, M. S. I. (2017). An overview of social media use in agricultural extension service delivery. *Journal of Agricultural Informatics*, *8*(3), 50–61. https://doi.org/10.17700/jai.2017.8.3.395

Brock University Library. (2023). *Research support.* Research Support – Brock University Library.

Burke, W. J., Frossard, E., Kabwe, S., & Jayne, T. S. (2019). Understanding fertiliser adoption and effectiveness on maize in Zambia. *Food Policy*, *86*, 101721, 1–12. https://doi.org/10.1016/j.foodpol.2019.05.004

Damalas, C. A., & Koutroubas, S. D. (2016). Farmers' exposure to pesticides: Toxicity types and ways of prevention. *Toxics*, *4*(1), 1–10. https://doi.org/10.3390/toxics4010001

Dao, A., Sanou, J., Gracen, V., & Danquah, E. Y. (2015). Identifying farmers' preferences and constraints to maize production in two agroecological zones in Burkina Faso. *Agriculture & Food Security*, *4*(1), 1–7. https://doi.org/10.1186/s40066-015-0035-3

Eppler, M. J. (2006). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes*. Springer Science & Business Media.

Flores, J., & Sun, J. (2018). Information quality awareness and information quality practice. *Journal of Data and Information Quality*, *10*(1), 1–18. https://doi.org/10.1145/3182182

Folnovic, T. (2021, June 12). *Importance of weather monitoring in farm production.* AGRIVI. https://blog.agrivi.com/post/importance-of-weather-monitoring-in-farm-production

Food and Agriculture Organisation of the United Nations (FAO), & International Telecommunication Union (ITU). (2022). *Status of digital agriculture in 47 Sub-Saharan African countries*. https://doi.org/10.4060/cb7943en

Gaillard, D. (2018, February, 8). Why you should display your contact details on a website? Samurai Web Experts. https://www.samurai-experts.com/spotlight/why-contact-details-are-important-on-a-website/

Hilary, R. S., Sseguya, H., & Kibwika, P. (2017). Information quality, sharing and usage in farmer organizations: The case of rice value chains in Bugiri and Luwero Districts, Uganda. *Cogent Food and Agriculture*, *3*(1), 1–16. https://doi.org/10.1080/23311932.2017.1350089

Holt-Day, J., Curren, L., & Irlbeck, E. (2020). USDA agricultural checkoff programs' YouTube presence and video quality. *Journal of Agricultural Education*, *61*(2), 190–202. https://doi.org/10.5032/jae.2020.01190

Huff, D., Edmond, D., & Gillette, C. (2015). *2015 B2B web usability report*. Huff Industrial Marketing and KoMarketing. https://komarketing.com/files/b2b-web-usability-report-2015.pdf

Isaya, E. L., Agunga, R., & Sanga, C. A. (2018). Sources of agricultural information for women farmers in Tanzania. *Information Development*, *34*(1), 77–89. https://doi.org/10.1177/0266666916675016

Jack, K., & Tobias, J. (2017). Seeding success: Increasing agricultural technology adoption through information. Growth Brief. International Growth Centre (IGC). https://www.theigc.org/sites/default/files/2017/12/IGCJ5833-Agriculture-growth-brief-171214-Web.pdf

Jallow, M. F., Awadh, D. G., Albaho, M. S., Devi, V. Y., & Thomas, B. M. (2017). Pesticide knowledge and safety practices among farm workers in Kuwait: Results of a survey. *International Journal of Environmental Research and Public Health*, *14*(4), 1–15. https://doi.org/10.3390/ijerph14040340

Janc, K., Czapiewski, K., & Wójcik, M. (2019). In the starting blocks for smart agriculture: The internet as a source of knowledge in transitional agriculture. *Journal of Life Sciences*, *90*, 1–12. https://doi.org/10.1016/j.njas.2019.100309

Kante, M., Oboko, R., & Chepken, C. (2019). An ICT model for increased adoption of farm input information in developing countries: A case in Sikasso, Mali. *Information Processing in Agriculture*, *6*(1), 26-46. https://doi.org/10.1016/j.inpa.2018.09.002

Lamptey, R. B., Sambo, I. A., & Hassan, A. A. (2017). Disseminating and promoting agriculture information through library and information services in Ghana. *Qualitative and Quantitative Methods in Libraries*, *5*(4), 901–907.

Laumer, S., Maier, C., & Weitzel, T. (2017). Information quality, user satisfaction, and the manifestation of workarounds: A qualitative and quantitative study of enterprise content management system users. *European Journal of Information Systems*, *26*(4), 333–360. https://doi.org/10.1057/s41303-016-0029-7

Magesa, M. M., Michael, K., & Ko, J. (2020). Access and use of agricultural market information by smallholder farmers: Measuring informational capabilities. *The Electronic Journal of Information Systems in Developing Countries*, *86*(6), 12134. https://doi.org/10.1002/isd2.12134

Malozo, M. (2021, February 21). Digital access to weather information helps young Tanzanian farmers build resilience. https://www.climatelinks.org/blog/digital-access-weather-information-helps-young-tanzanian-farmers-build-resilience

Maresma, A., Ballesta, A., Santiveri, F., & Lloveras, J. (2019). Sowing date affects maize development and yield in irrigated Mediterranean environments. *Agriculture*, *9*(3), 67–69. https://doi.org/10.3390/agriculture9030067

Martin, R., & Kahamba, J. S. (2017). Access and use of information communication technologies by agricultural extension agents in Tanzania. *Journal of the Open University of Tanzania*, *24*(2), 82–96.

Mattah, M. M., Mattah, P. A., & Futagbi, G. (2015). Pesticide application among farmers in the catchment of Ashaiman irrigation scheme of Ghana: Health implications. *Journal of Environmental and Public Health*, *2015*, 1–7. https://doi.org/10.1155/2015/547272

Maziku, P. (2019). Determinants for post-harvest losses in maize production for small holder farmers in Tanzania. *African Journal of Applied Research (AJAR)*, *5*(1), 1–11.

Mbagwu, F. C., Benson, O. V., & Onuoha, C. O. (2018). Challenges of meeting information needs of rural farmers through internet-based services: Experiences from developing countries in Africa. Paper presented to IFLA WLIC 2018 Kuala Lumpur. https://library.ifla.org/2195/1/166-mbagwu-en.pdf

Misaki, E., Apiola, M., & Gaiani, S. (2016). Technology for small scale farmers in Tanzania: A design science research approach. *The Electronic Journal of Information Systems in Developing Countries*, *74*(1), 1–15. https://doi.org/10.1002/j.1681-4835.2016.tb00538.x

Msemo, H. E., Taylor, A. L., Birch, C. E., Dougill, A. J., & Hartley, A. (2021). The value of weather and climate information to the Tanzanian disaster risk reduction sector using nonmonetary approaches. *Weather, Climate, and Society*, *13*(4), 1055–1068. https://doi.org/10.1175/wcas-d-21-0005.1

Mubofu, C., & Malekani, A. W. (2020). Agricultural information sources, channels and strategies for sharing agricultural research findings among farmers in Iringa district in Tanzania. *Library Philosophy and Practice*, 4041, 1–14.

Mubushar, M., Aldosari, F. O., Baig, M. B., Alotaibi, B. M., & Khan, A. Q. (2019). Assessment of farmers on their knowledge regarding pesticide usage and biosafety. *Saudi Journal of Biological Sciences*, *26*(7), 1903–1910. https://doi.org/10.1016/j.sjbs.2019.03.001

Mugonya, J., Kalule, S. W., & Ndyomugyenyi, E. K. (2021). Effect of market information quality, sharing and utilisation on the innovation behaviour of smallholder pig producers. *Cogent Food and Agriculture*, *7*(1), 1–17. https://doi.org/10.1080/23311932.2021.1948726

Muhammad, A. H., Siddique, A., Youssef, A. E., Saleem, K., Shahzad, B., Akram, A., & Al-Thnian, A. B. S. (2020). A hierarchical model to evaluate the quality of web-based e-learning systems. *Sustainability*, *12*(4071), 1–23. https://doi.org/10.3390/su12104071

Naruka, P. S., Verma, S., Sarangdevot, S. S., Pachauri, C. P., Kerketta, S., & Singh, J. P. (2017). A study on role of WhatsApp in agriculture value chains. *Asian Journal of Agricultural Extension, Economics & Sociology*, *20*(1), 1–11. https://doi.org/10.9734/ajaees/2017/36498

Ndimbwa, T., Mwantimwa, K., & Ndumbaro, F. (2020). Channels used to deliver agricultural information and knowledge to smallholder farmers. *IFLA Journal*, *47*(2), 153–167. https://doi.org/10.1177/0340035220951828

Nkebukwa, L. L. (2018). Access and use of agricultural information in Muleba and Misenyi districts, Tanzania. *Business Education Journal*, *2*(1), 1–11.

Ogessa, C. M., & Sife, A. S. (2017). Newspaper coverage of agricultural information in Tanzania. *University of Dar es Salaam Library Journal*, *12*(1), 12–26.

Sai, M. V. S., Revati, G. D., Ramya, R., Swaroop, A. M., Maheswari, E., & Kumar, M. M. (2019). Knowledge and perception of farmers regarding pesticide usage in a rural farming village, Southern India. *Indian Journal of Occupational and Environmental Medicine*, *23*(1), 32–36. https://doi.org/10.4103/ijoem.ijoem_121_18

Saïdou, A., Balogoun, I., Ahoton, E. L., Igué, A. M., Youl, S., Ezui, G., & Mando, A. (2018). Fertiliser recommendations for maize production in the South Sudan and Sudano-Guinean zones of Benin: Improving the profitability, sustainability and efficiency of nutrients through site specific fertiliser recommendations in West Africa Agro-Ecosystems. *Nutrient Cycling in Agroecosystems*, *1*(110), 361–373. https://doi.org/10.1007/978-3-319-58792-9_13

Spink, A., Jansen, B. J., Blakely, C., & Koshman, S. (2006). A study of results overlap and uniqueness among major web search engines. *Information Processing & Management*, *42*(5), 1379–1391. https://doi.org/10.1016/j.ipm.2005.11.001

Tanzania Meteorological Authority (TMA). (n.d.). https://www.meteo.go.tz/

Tao, D., LeRouge, C., Smith, K. J., & De Leo, G. (2017). Defining information quality into health websites: A conceptual framework of health website information quality for educated young adults. *JMIR Human Factors*, *4*(4), e25. https://doi.org/10.2196/humanfactors.6455

Tate, M. A. (2019). *Web wisdom: How to evaluate and create information quality on the Web*. CRC Press. https://doi.org/10.1201/b22397

Temesgen, T., & Kebena, T. (2019). Effects of varieties and intra row spacing on yield of maize (Zea mays L.) under supplementary irrigation at Guliso, Western Ethiopia. *International Journal of Environmental Sciences & Natural Resources*, *19*(5), 144–151. https://doi.org/10.19080/ijesnr.2019.19.556024

Tumbo, S. D., Mwalukasa, N., Fue, K. G., Mlozi, M. R., Haug, R., & Sanga, C. (2018). Exploring information seeking behavior of farmers in information related to climate change adaptation through ICT (CHAI). *International Review of Research in Open and Distributed Learning*, *19*(3), 300–319. https://doi.org/10.19173/irrodl.v19i3.3229

Uwandu, C. N., Thomas, K. A., & Okoro, C. M. (2018). Utilization of agricultural information sources and adoption of animal and crop technologies among farming households in Imo State, Nigeria. *Journal of Agricultural Extension*, *22*(1), 143–155. https://doi.org/10.4314/jae.v22i1.13

Verkijika, S. F., & De Wet, L. (2018). Quality assessment of e-government websites in Sub-Saharan Africa: A public values perspective. *The Electronic Journal of Information Systems in Developing Countries*, *84*(2), 1–17. https://doi.org/10.1002/isd2.12015

Wheeler, S. A., Zuo, A., Bjornlund, H., Mdemu, M. V., Van Rooyen, A., & Munguambe, P. (2017). An overview of extension use in irrigated agriculture and case studies in south-eastern Africa. *International Journal of Water Resources Development*, *33*(5), 755–769. https://doi.org/10.1080/07900627.2016.1225570

Wilson, R. T., & Lewis, J. (2015). *The maize value chain in Tanzania*. South African Institute of International Affairs. https://saiia.org.za/wp-content/uploads/2016/06/008-Tanzania_maize.pdf

Wineman, A., Jayne, T. S., Modamba, E. I., & Kray, H. (2020). The changing face of agriculture in Tanzania: Indicators and transformation. *Development Policy Review*, *38*(6), 685–709. https://doi.org/10.1111/dpr.12491

Appendix 1: Online agricultural information evaluation tool

Note: The decision as to whether a particular information resource is of sufficient quality or not remains in the hands of the information user. While this tool seeks to support the safe use of web-based online agricultural information, it does not replace the role of agricultural professionals in dispensing information or advice. Therefore, information users are encouraged to seek professional guidance when they feel that they will benefit from doing so.

Dimension	Attributes	Information available/not available
Authority	Are the names of authors, creators, or contributors available?	
	Does the online source present contact information such as phone numbers, email addresses of creators / authors / contributors?	
Completeness	To what extent does the information resources provide detailed information relevant to the maize value chain (in terms of the following)?	
a. Seeds	1. Variety	
	2. Agroecological zone	
	3. Sowing time	
	4. Seed rate (kg/ha)	
	5. Planting/sowing spacing	
	6. Maturity period (in days)	
	7. Yield/ha (depends on factors such as planting space)	
b. Fertiliser(s)	1. Type	
	2. Time/stage of the application	
	3. Soil information	
	4. Method of application	
	5. Application rate (kg/ha)	
	6. Interval/frequency of application	
c. Pesticide(s)	1. Pest type	
	2. Pesticide type	
	3. Time/stage of the application	
	4. Plant stage	
	5. Method of application	
	6. Application rate (e.g. litres/ha)	
	7. Interval/frequency of application	
d. Grain handling	1. Cleaning	
	2. Drying	
	3. Storage conditions	
	4. Storage facilities	
e. Weather information	1. Location	
	2. Publication time (not older than a month)	
Timeliness	Does the online source provide a date of the last update of information on the site? Has the information available on the online source been updated recently (at least within the last two years for maize value chain information and in the last month for weather information)?	
Understandability	Is the information provided in the online information resource in a language that can be understood by the readers?	
	Is information presented in the online resource supported by graphical (photo / animation) or video information to enhance understanding?	

Appendix 2: The 39 web-based online resources studied

Name of online resource	Resource location (in December 2021)
Kilimo Solution Investments Tanzania (Agriculture Solution Investments Tanzania)	https://web.facebook.com/218482752173471/posts/utunzaji-na-uhifadhi-wa-nafaka-ya-mahindiutanguliziasilimia-80-ya-mahindi-yanayo/563297451025331/?_rdc=1&_rdr
Farmers Market	http://farmersmarket.co.tz/kanuni-5-muhimu-za-kufanya-kilimo-bora-cha-mahindi/
Mjasiriamali hodari (Strong Entrepreneur)	https://mjasiriamalihodari.blogspot.com/2017/10/kilimo-bora-cha-mahindi-growing-maize.html
Kilimo bora (Good Agriculture)	https://agricomlizy.blogspot.com/2019/09/kilimo-cha-mahindi.html
Farm Radio International Resource Pack 104	http://scripts.farmradio.fm/sw/radio-resource-packs/104-chickens-cowpea-sorghum/taarifa-za-kina-uzalishaji-wa-mahindi/
Farm Radio International Resource Pack 110	http://scripts.farmradio.fm/sw/radio-resource-packs/110-farm-radio-resource-pack/taarifa-za-kina-kupunguza-upotevu-wa-mahindi-baada-ya-mavuno/
Jifunze Kilimo (Learn Agriculture)	http://jifunzekilimotz.blogspot.com/20http://jifunzekilimotz.blogspot.com/2018/10/kilimo-bora-cha-mahindi.html8/10/kilimo-bora-cha-mahindi.html
Mitiki-Kilimo Kwanza (Teak-First Agriculture)	http://mitiki.blogspot.com/2010/08/kilimo-bora-cha-mahindi.html
Jukwaa la Kilimo (Agriculture Platform)	https://kilimoforum.wordpress.com/2016/08/17/kilimo-bora-cha-mahindi/
Kangetakilimo (Kangeta Agriculture)	https://kangetakilimo.co.tz/sw/makala-zote/113-kilimo-cha-mahindi.html
MSMEs Information Portal	https://m.facebook.com/msmeinfotz/posts/1841079969476210/
Shamba Darasa Mkusi (Field Class Mkusi)	https://www.youtube.com/@shambadarasaTV
E-Sokoni (In the Market)	https://vc4a.com/ventures/e-sokoni-2/
Halimashauri ya Wilaya ya Handeni (District Council of Handeni)	http://www.handenidc.go.tz/how-do-i-single/kulima-kilimo-bora-cha-zao-la-mahindi
Maisha Daily (Life Daily)	https://maishadaily.wordpress.com/2016/12/16/mahindi/
Frank Sungau	https://www.youtube.com/watch?v=jbs1TyS07uo
Morgiculture tz	https://www.mogriculture.com/2016/09/kilimo-cha-mahindi/
Mtandao wa Ushauri kwa Wakulima (Farmers' Advice Network)	https://www.youtube.com/watch?v=ty3kHrQU1lw
Kilimo Bora (Good Agriculture)	http://kilimoboratz2.blogspot.com/2018/01/zifahamu-mbegu-bora-mpya-za-mahindi.html

Name of online resource	Resource location (in December 2021)
Kangetakilimo (Kangeta Agriculture)	https://kangetakilimo.blogspot.com/2017/08/kilimo-bora-cha-mahindi-na-kangeta.html
Kilimo Online (Agriculture Online)	http://kilimoonline.blogspot.com/2018/01/kilimo-cha-kisasa-cha-mahindi.html
Kilimo For Life (Agriculture for Life)	https://kilimoforlifetz.blogspot.com/2018/10/kilimo-bora-na-cha-kitaalamu-cha.html
Maisha Daily (Life Daily)	http://maishadaily1.blogspot.com/2018/05/kilimo-bora-cha-mahindi.html
Modern Agriculture	http://brilliantus.blogspot.com/2016/09/kilimo-cha-kisasa-cha-mahindi.html
Muakilishi (Representative)	https://www.muakilishi.com/article/kilimo-kanuni-5-muhimu-za-kilimo-bora-cha-mahindi#
Mkulima Mdadisi (Inquisitive Farmer)	https://mkulimamdadisi.blogspot.com/2016/07/kilimo-cha-mahindi.html
Mkulima Tanashati (Smart Farmer)	http://mkulimatanashati.blogspot.com/2018/01/kilimo-cha-mahindi.html
Vincent Munde	https://www.youtube.com/channel/UCw8hCwmSy1gGyEYZ9BJavJw/playlists?view=1
Hodari Agricultural Consultants (Strong Agricultural Consultants)	https://www.facebook.com/hodarikilimo/posts/872403239770419
Farmers Market	http://farmersmarket.co.tz/hizi-ndizo-mbegu-za-mahindi/
Mkulima Mbunifu (Innovative Farmer)	https://mkulimambunifu.org/kilimo/namna-ya-kuhifadhi-mahindi-kulinda-ubora-na-kukidhi-mahitaji-ya-chakula/
Zao la Mahindi (Maize Crop)	http://africasoilhealth.cabi.org/wpcms/wp-content/uploads/2016/09/Maize_technology_brief1.pdf
Muviza	https://www.facebook.com/muviza.muviza/
Kilimo Hifadhi Shadidi Endelevu Katika Mseto wa Mazao ya Mahindi na Mikunde (Sustainable Agriculture for Mixed Farming in Maize and Legumes)	https://simlesa.cimmyt.org/wp-content/uploads/Final-CASI-booklet-revised-9-Feb-2019-1.pdf
Channel Ten	https://www.youtube.com/watch?v=cecIP1IS2MM
Kilimo Bora Tanzania (Good Agriculture in Tanzania)	http://kilimoboratanzania5.blogspot.com/2016/10/kilimo-cha-mahindi.html
Shambani Solutions (Farm Solutions)	http://shambanisolutions.blogspot.com/
Mkulima Mwenzako (Your Fellow Farmer)	https://mkulimamwenzako.wordpress.com/
Mkulima Mbunifu (Innovative Farmer)	https://mkulimambunifu.org/

Exploring COVID-19 public perceptions in South Africa through sentiment analysis and topic modelling of Twitter posts

Temitope Kekere
PhD student, Data Science for Social Impact group, Department of Informatics, University of Pretoria
https://orcid.org/0000-0002-5362-2685

Vukosi Marivate
Professor and ABSA UP Chair of Data Science, Department of Computer Science, University of Pretoria
https://orcid.org/0000-0002-6731-6267

Marié Hattingh
Professor, Department of Informatics, University of Pretoria
https://orcid.org/0000-0003-1121-8892

Abstract
The narratives shared on social media during a health crisis such as COVID-19 reflect public perceptions of the crisis. This article provides findings from a study of the perceptions of South African citizens regarding the government's response to the COVID-19 pandemic from March to May 2020. The study analysed Twitter data from posts by government officials and the public in South Africa to measure the public's confidence in how the government was handling the pandemic. A third of the tweets dataset was labelled using valence aware dictionary and sentiment reasoner (VADER) lexicons, forming the training set for four classical machine-learning algorithms—logistic regression (LR), support vector machines (SVM), random forest (RF), and extreme gradient boosting (XGBoost)—that were employed for sentiment analysis. The effectiveness of these classifiers varied, with error rates of 17% for XGBoost, 14% for RF, and 7% for both SVM and LR. The best-performing algorithm (SVM) was subsequently used to label the remaining two-thirds of the tweet dataset. In addition, the study used, and evaluated the effectiveness of, two topic-modelling algorithms—latent dirichlet allocation (LDA) and non-negative matrix factorisation (NMF)—for classification of the most frequently occurring narratives in the Twitter data. The better-performing of these two algorithms, NMF, identified a prevalence of positive narratives in South African public sentiment towards the government's response to COVID-19.

Keywords
sentiment analysis, sentiment classification, topic modelling, social media, Twitter, natural language processing (NLP), COVID-19, South Africa, government response, public perceptions

Acknowledgements
We wish to thank University of Pretoria and Canada's International Development Research Centre (IDRC) for its funding of (first-listed author) Kekere's research, and ABSA for its sponsorship of (second-listed author) Marivate's industry chair and activities related to this study.

DOI: https://doi.org/10.23962/ajic.i31.14834

Recommended citation
Kekere, T., Marivate, V., & Hattingh, M. (2023). Exploring COVID-19 public perceptions in South Africa through sentiment analysis and topic modelling of Twitter posts. *The African Journal of Information and Communication (AJIC), 31*, 1-27. https://doi.org/10.23962/ajic.i31.14834

1. Introduction

The World Health Organisation (WHO) declared COVID-19 a global pandemic on 11 March 2020 (Ghebreyesus, 2020) following an outbreak in Wuhan, China, in late 2019 (Relman, 2020). The South African government declared it a national disaster on 15 March 2020, and took decisive measures to combat the spread of the disease in the country (Ramaphosa, 2022). The measures implemented by President Cyril Ramaphosa's government to stem infection rates in South Africa included setting up emergency response systems for affected people; imposing travel bans on high-risk countries; requiring self-isolation and quarantine of travellers; school closures; national tracking and tracing of infected people; reducing social gatherings; and reducing the movement of people through the implementation of lockdown Alert Levels 1–5. Alert Level 5 indicated high COVID-19 spread and restricted movement, and Alert Level 1 indicated low spread of the virus and the resumption of everyday activities.

Social media was one of several platforms utilised by the South African government to provide the public with information on the symptoms, status, and spread of this highly infectious disease, and on regulations and restrictions instituted to curb the spread of the pandemic. The public also utilised social media platforms such as Twitter to express their sentiments about the government's response to the pandemic. Extracting these social media discourses can serve as a proxy for understanding the public's views. Such evolving conversations provide indications of the perceptions of those active on social media in respect of the effectiveness of government interventions.

Sentiment analysis of South African Twitter data can provide information on how the government's responses to COVID-19 were perceived. In addition, topic modelling of the same data can reveal the topics of discourses that were linked to COVID-19 and government measures. These topics can highlight critical conversational issues and underlying narratives during the pandemic. They can indicate how the populace perceived the information passed on by government officials on social media.

This study explored the ability of natural language processing (NLP) machine-learning algorithms to measure the public's perception of the South African government's COVID-19 responses through sentiment analysis and topic modelling tasks. We deployed four widely used sentiment classifier algorithms in this study: logistic regression (LR), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost). For topic modelling tasks, we deployed two other frequently used algorithms: latent dirichlet allocation (LDA) and non-negative matrix factorisation (NMF).

The study generated results from the four sentiment analysis and two topic modelling methods applied to South African Twitter data, including empirical results showing the performance of the methods applied. The rest of the article follows this structure: section 2 reviews literature on sentiment analysis and topic modelling; section 3, focused on methodology, describes the data collection and modelling techniques applied in the study; section 4 provides the findings of the study; and section 5 offers conclusions and makes suggestions for future research.

2. Literature review

Sentiment analysis

Sentiment analysis, which determines whether a word or phrase indicates a positive, negative, or neutral sentiment, is an NLP tool that can measure public opinion on social media. Such measurements based on social media can complement analysis from additional data sources, such as news media and government reports, in providing insights into factors that drive citizens' perceptions of, and behaviours in response to, government regulations and policies. In the context of the COVID-19 pandemic, Mendez-Brito et al. (2021) have identified numerous studies seeking to understand public sentiment towards government interventions as national governments try to balance economic, social, and political goals while limiting the spread of the highly infectious disease.

Three main kinds of methods can be employed in undertaking a sentiment analysis task: lexicon analysis; use of machine-learning algorithms; and hybrid methods.

Lexicon analysis

This lexicon-based approach (Moussa et al., 2020; Taboada et al., 2011) can also be referred to as dictionary-based (Cruz et al., 2015) or corpus-based (Rice & Zorn, 2021). This method works with a corpus or group of words and their valence scores in a dictionary. A valence score indicates whether the text it represents has a positive, negative, or neutral sentiment. From this valence dictionary, a given text that requires sentiment classification receives a reference for its scores. The text's overall valence score, usually an aggregate, determines the polarity sentiment (polarity) of the text (Hu & Liu, 2004; Mohammad et al., 2013; Nielsen, 2011; Stone & Hunt, 1963). For example, the word "good" will receive a positive sentiment score, while "bad" receives a negative score. A word with a neutral sentiment will have a score of zero (in the middle of the range). Some sentiment scores have a range of say 0 to 5, with 5 being the score that represents the positive end of the range (spectrum) while 0 is the other extreme (a negative word). In terms of this range example, a summed sentiment score of 3 would stand for a neutral word.

The earliest dictionary built to extract the opinion of a given piece of text was the General Inquirer (Stone & Hunt, 1963), and others include the multi-perspective question answering (MPQA) subjectivity lexicon (Wiebe et al., 2005), the NRC valence, arousal, and dominance (VAD) lexicon (Mohammad, 2018), the Yelp restaurant review corpus (Kiritchenko et al., 2014) and valence aware dictionary and sentiment reasoner (VADER) lexicons (Hutto & Gilbert, 2014).

A lexicon analysis study by Domalewska (2021) explores public opinion on COVID-19 economic relief measures implemented by the Polish government between March and June 2020. The textual data or corpus comprised Polish Twitter tweets and Facebook posts. Domalewska (2021) created a manually curated dictionary of Polish words to label the sentiments in the text corpus. The curation was necessary because existing dictionaries or word lists used to label the sentiment of a tweet or post were not available in Polish. Domalewska (2021) finds that, despite the Polish government's disbursement of EUR49.24 billion in economic relief, many Poles were critical of the solution implemented. The findings also show that the negative sentiment decreased over time as the Polish people complained less about the shortfalls of the economic relief (Domalewska, 2021).

Use of machine-learning algorithms

Machine-learning algorithms learn the pattern or relationship between a text or tweet and its corresponding label (its sentiment). For every given text, a label is assigned that refers to the sentiment of the text, tweet, or post. There are three different types of training in machine learning: supervised learning (Rustam et al., 2021), unsupervised learning (Jiang et al., 2022), and reinforcement training (Park et al., 2022).

Supervised learning

In supervised learning, the machine-learning algorithms supervise the training process based on learning example pairs—with each pair composed of text and its corresponding sentiment label (target). Human annotators often provide labels. A study by Gulati et al. (2022) compares the performance of seven sentiment classifier algorithms on COVID-19 tweets. Four of the seven classifiers tested—linear SVM, perceptron, the passive-aggressive classifier (PAC), and LR—are found to outperform the multinomial naïve Bayes (MNB), Bernoulli naïve Bayes (BNB), and adaptive boosting (AdaBoost) classifiers. The study also finds that unigram, bigram, and trigram features do not improve the performance of the three top-performing classifiers.

A study by Kuyo et al. (2021) evaluates the performance of three variants of the naïve Bayes (NB) sentiment classifier—MNB, BNB, and Gaussian naïve Bayes (GNB)—on an international corpus of COVID tweets and responses to governments' lockdown regulations. The study combines machine-learning algorithms with three different n-grams to examine which of the three classifiers' performance is better. N-grams (Jurafsky & Martin, 2018) are sequences of words wherein a single word is a unigram; two sequential words are bigrams; and three sequential words are trigrams. The Kuyo et al. (2021) study finds that the BNB classifier (with bigrams) is more effective than the MNB and GNB classifiers (Kuyo et al., 2021).

A similar study, by Aljameel et al. (2021), compares the performance of three sentiment classifiers—SVM, k-nearest neighbours (KNN), and NB—in determining public opinions towards government's COVID-19 measures in Saudi Arabia. Using a corpus of Arabic tweets from the country's five regions, the study finds that SVM (with bigrams) performs better than KNN and NB (Aljameel et al., 2021).

Unsupervised learning

This type of machine learning involves training algorithms to find patterns in clusters of text, to determine sentiment analysis, without the use of target pairs. An example of this approach is the study by Jiang et al. (2022), in which tweets were projected into high-dimensional vectors (or matrices) in order to find clusters of words, using: Kmeans++ algorithms; the synthetic minority oversampling technique (SMOTE) algorithm to balance the dataset; and a convolutional neural network (CNN) to find salient textual features. In this study, the result is then fed into a bidirectional long-short term memory (Bi-LSTM) sequence processing model, in order to capture the contextual information needed to label the text (Jiang et al., 2022).

Reinforcement learning
This machine-learning technique uses policies that reward the algorithm for a correct output and penalise it when the output is wrong. An example of this approach is the Chai et al. (2020) study, in which the researchers manually generate descriptions for the sentiment classes and where, for each class, the algorithm has to select a text span (a portion of text) guided by a policy that rewards it when the correct description is chosen.

Hybrid methods
Hybrid sentiment analysis methods use a combination of lexicon-based and machine-learning methods. The process involves using a dictionary or corpus to look up the polarity of the text, and then the polarity becomes the label for the algorithm to learn. In the end, the algorithm provides a classification for each piece of text based on an average of the results of two or more algorithms. The Soumya and Pramod (2021) study provides an example of the hybrid method. In their study, the researchers hybridise SentiWordNet (Baccianella et al., 2010; Esuli & Sebastiani, 2006), a publicly available lexicon of positive, negative, and neutral words, together with three separate classifiers—RF, linear SVM, and radial basis function (RBF)—in order to carry out a fine-grained sentiment analysis of Malaysian tweets. The researchers translate the words in SentiWordNet into Malay using Google Translate, and then extract features using, among other tools, bag-of-words (BoW) and term frequency–inverse document frequency (TF-IDF). The study finds that RF outperforms linear SVM and RBF (Soumya & Pramod, 2021).

Sentiment analysis algorithms deployed in this study
This study adopted a hybrid methodology that combined lexicon analysis and machine-learning techniques for sentiment analysis. The four sentiment classifier algorithms used—LR, SVM, RF, and XGBoost—were selected as baseline models for machine-learning algorithms that have proven their effectiveness across various tasks. Initially, one-third of the tweet dataset was labelled using VADER lexicons, providing ground truth for training the models. The study then employed supervised training to enable the classifiers to learn from the labelled data. It then evaluated and compared the performance of each model. This process ensured that the algorithms were trained and tested on previously unseen data, enhancing the validity of the study results. Having built a sentiment model, we applied it to label two-thirds of the remaining tweets dataset.

Topic modelling
Topic modelling in NLP is the process of discovering topics or themes in an extensive collection of unstructured text by using statistical methods and unsupervised machine learning. Topic modelling provides a means of summarising large clusters of documents into a set of topics. In topic modelling, a collection of words represents a topic in the text. The goal is to find the summary description of each collection of words in a manner that allows for further analysis—while retaining semantic relationships between texts for tasks such as sentiment/text classification, summarisation, and similarity or relevance judgments. Among the most-used topic modelling algorithms are latent semantic indexing (LSI) (Deerwester et al., 1990), latent dirichlet allocation (LDA) (Blei et al., 2003), and non-negative matrix factorisation (NMF). We used LDA and NMF in this study.

LDA is a statistical generative model that extracts topics from a given corpus. It models documents as discrete dirichlet distributions over topics. In other words, LDA assumes that the topics of a document and the words in those topics follow a dirichlet distribution. Given a set of corpora with D number of documents, LDA can discover T number of topics in the document. The T number has to be set manually. LDA finds a set of weights between the documents and topics, and between topics and words.

NMF is a statistical method that reduces the input of the corpora to extract meaningful topics. Through factor analysis, it decomposes the document–term matrix into two smaller matrices that contain only positive numbers. Given a set of corpora with matrix M, NMF factorises M into smaller matrices T and H. Matrix T is a cluster of topics extracted from the document. Matrix H is the weights of topics discovered from the document. The number of topics is calculated by optimising T and H over an expectation-maximisation (EM) objective function, until both matrices converge.

In a topic-modelling study focused on South Africans' reactions to the government's efforts to manage the COVID pandemic, Mutanga and Abayomi (2022) use LDA to discover topics in a tweet corpus. Nine topics are extracted, as follows: lockdown; fifth-generation (5G) conspiracy theory; staying home; alcohol; South African National Defence Force (SANDF) and police violence; tracing of daily statistics; South African Presidential address; essential workers; and Bill Gates conspiracy theories. Since conspiracy theories and misinformation can bring about a lack of confidence in the government's efforts, this study is valuable in highlighting areas where the government can direct its communication efforts (Mutanga & Abayomi, 2022).

Nwankwo et al. (2020) use two statistical models—LDA and bidirectional encoder representations from transformers (BERT)—to model topics in a tweet corpus containing fake news about COVID-19 in Sub-Saharan Africa. Zamani et al. (2020)

use topic modelling to explore public concerns from a tweet corpus in the US with respect to COVID-19. Wan et al. (2021) apply the bidirectional and auto-regressive transformers (BART) statistical model to cluster COVID-19 topics in a corpus of US digital media news content.

A study by Yan et al. (2021) uses LDA to reveal topics found in Reddit posts that indicate public fear and resentment towards COVID contact-tracing measures in Canada. Tao et al. (2020) apply LDA to Twitter data collected globally to generate a visualisation of the evolution of COVID-19 topics from March to July 2020. Xing et al. (2021) use topic-modelling of posts on Twitter and on the Chinese micro-blogging platform Weibo to explore cultural differences between social media users in the US and China as factors in public acceptance of government COVID interventions in the two countries. Hou et al. (2021) conduct topic modelling, at different intervals, with data from Weibo in order to track the evolution of Chinese public opinion on COVID matters.

Topic modelling algorithms deployed in this study

LDA was one of the two topic modelling algorithms we used for our study, because it can be statistically validated. The other topic modelling algorithm we used, NMF, was chosen because of the non-negative constraints it has on matrices, leading to the meaningful extraction of themes.

3. Methodology

Dataset collection

As seen in Table 1, the dataset of COVID-19-related tweets used in this study was composed of three subsets, which we named NICD, Gauteng Province and Healthza. The NICD (National Institute for Communicable Diseases) subset consisted of tweets from the NICD (a government agency), from then-Health Minister Zweli Mkhize, from the South African Presidency, and from the public in response. The Gauteng Province subset consisted of tweets from then-Gauteng Premier David Makhura, from then-Gauteng MEC (Member of the Executive Council) for Health Bandile Masuku, and from members of the public in response. The Healthza subset consisted of tweets from the national Health Minister and the NICD, and from the public in response.

The NICD subset was collected by Marivate et al. (2021), and the Gauteng Province and Healthza subsets were compiled following the same pattern of data collection (Marivate et al., 2021).

Table 1: Twitter data subsets used in the study

Name of subset	Size	Number of tweets	Number of tweets (after pre-processing)	Number of labels	Description
NICD	147 MB	919,019	869,851	None	Tweets from NICD, Health Minister, and Presidency, and from public in response.
Gauteng Province	10.9 MB	71,313	65,730	None	Tweets from Gauteng Premier and Gauteng Health MEC, and from public in response.
Healthza	54.8MB	309,153	295,035	295,035	Tweets from the national Health Minister and NICD, and from public in response.

Data preprocessing

The tweets from the three data subsets contained text mixed with numbers; misspellings; emoticons and emojis; special characters; universal resource locators (URLs); hashtags; and mentions and abbreviations in languages other than English. Before processing, we removed URLs, emojis, emoticons, and digits from the data, and converted emoticons and emojis to their text equivalent. Next, we implemented the natural language toolkit (NLTK) libraries (Loper & Bird, 2002) and the Terrier information retrieval package (Ounis et al., 2005) to remove stop words and punctuation. We also expanded contractions, removed non-English and duplicate tweets, and made all text entirely lowercase. Finally, the text was tokenised and fed into a machine-learning pipeline containing the six algorithms used in the study: four for sentiment analysis, and two for topic modelling.

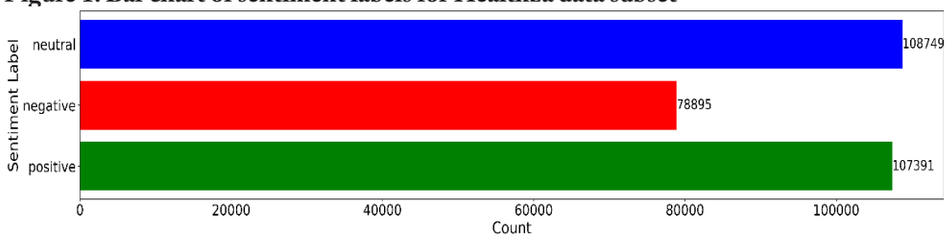
Dataset annotation

The dataset primarily consisted of unlabelled texts requiring positive, negative, or neutral sentiment assignments. The Healthza data subset (comprising 295,035 tweets) underwent labelling using VADER lexicons, an approach incorporating qualitative and quantitative methods for sentiment analysis. VADER is a lexicon- and rule-based sentiment analysis tool specifically attuned to social media sentiments.

For our annotation, we set the VADER compound score threshold as follows: a score greater than 0.05 was considered positive, less than -0.05 was deemed negative, and any score within the range between positive and negative was marked as neutral. As shown in Figure 1, this process resulted in 108,749 tweets labelled neutral, 78,895 labelled negative, and 107,391 labelled positive.

This approach to labelling served two primary purposes. First, it provided a meaningful classification of the sentiments, because VADER lexicons are suited for social media content. Second, it provided an extensive set of examples for the supervised training of the machine-learning algorithms used in this study. As supervised learning involves teaching models to recognise patterns in the text by providing examples of tweet-sentiment label pairs, the substantial volume of labelled tweets contributed significantly to efficient training of the algorithms.

Figure 1: Bar chart of sentiment labels for Healthza data subset



As seen in the bar chart in Figure 1, the dataset skews towards positive and neutral sentiments, with negative samples not represented nearly as much as the other two classes. To reduce this imbalance and ensure unbiased machine-learning model training, we employed SMOTE, which is an effective technique for generating synthetic samples for minority classes. SMOTE was used to augment the negative class. This method identifies the nearest neighbours of minority class instances and generates synthetic data points through interpolation. These new instances then add to the original dataset.

The systematic application of SMOTE allowed us to achieve a more balanced distribution of sentiment classes, enhancing the fairness of our dataset. Furthermore, by setting a random seed, we ensured the reproducibility of our results. This balancing of the dataset enabled us to improve the overall performance of our machine-learning models by mitigating initial biases, resulting in improved accuracy and generalisation. The models were then trained and evaluated on this balanced dataset, leading to improved performance.

Feature extraction and selection

Feature extraction and selection are vital to building text classification models. This study used two popular vectorisers, BoW and term frequency–inverse document frequency (TF-IDF), to measure the importance of the word vectors. BoW, a simple text representation method, generates a vector based on word counts in the document while disregarding word order. However, this approach results in a sparse dataset due to multiple term occurrences for a single word. On the other hand, TF-IDF combines term frequency of a word with the logarithm of its inverse, providing a

more informative representation of word importance in the corpus. The importance of the weights of the words in the corpus is defined by the following equation:

$$w_{w,d} = tf_{w,d} idf_c \quad (1)$$

where $tf_{w,d}$ is the occurrence of terms or words in the documents, and idf_c is the logarithm of the ratio of total number of documents to the number of documents with terms in the corpus.

The four sentiment classifier algorithms

In this study, we built four classifiers—using LR, SVM, RF, and XGBoost—for sentiment analysis. Leveraging a third of the tweets labelled with VADER lexicons, we trained these algorithms on tweet-sentiment label pairs, enabling them to learn the sentiment patterns in the tweets. Accuracy served as the performance measure, with the sentiment polarity of each text assigned as positive, negative, or neutral, forming three target classes. We then used the entire corpus to extract features for the classifiers.

To evaluate the performance of the trained classifiers, we tested them on unseen Twitter data. The comparison resulted in a ranking based on error rates, allowing us to determine the most effective model. We subsequently utilised this top-performing model to label the remaining unlabelled data. The following sections detail the functioning of each of the four algorithms.

Logistic regression (LR)

LR is widely used for binary classification but can be modified to solve multi-class classification tasks. For example, a logistic regression model for two output classes, negative or positive labels, is described by the probability that:

$$\log\left(\frac{p}{1-p}\right) = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \quad (2)$$

where:

1. p is the probability of the target class
2. a, b are coefficients of x
3. x is the predictor of sentiment (a BoW, for example)
4. n is an integer.

Should the value p be greater than 0.5, then the sentiment belongs to one of the classifications, say, positive class, while probabilities lower than 0.5 take the other binary class, a negative label in this case. However, for a multi-class problem we use

multinomial logistic regression (Jurafsky & Martin, 2018) to determine whether the target class is positive, negative, or neutral. In other words, softmax regression or maxent classifier categorises each observation into one of the *K* classes where only one class is considered correct. Should the class *c* be correct or selected, the output p_c will be set to $p_c = 1$ and other output probabilities to zero. The decision boundary is defined by the softmax function expressed as:

$$\text{softmax}(v_i) = \frac{e^{v_i}}{\sum_j^k e^{v_i}} \; 1 \leq i \; \leq K \tag{3}$$

where *v* is a vector of dimensionality *K*. The softmax function is the dot product between the weights vector *w* and input vector *x*. The probabilities of each of the output or target classes, y_k, can be expressed as:

$$p(y_k = 1|\text{x}) = \frac{e^{w_k.\mathbf{x}+b_k}}{\sum_j^k e^{w_k.\mathbf{x}+b_k}} \tag{4}$$

Let *W* be the weight matrix with *K* rows, where each row represents the weight vector for the *k-th* class. The matrix *W* has dimensions *[K* x *f]*, with *K* being the number of output classes and *f* the number of input features. Also let *b* be the bias vector. Then the predicted class probabilities can be computed as:

$$\hat{y} = \text{softmax}\,(Wx + b) \tag{5}$$

Support vector machines (SVM)

SVM algorithms find a line that maximises the separation of data points to their classes in an n-dimensional space. The line is called a hyperplane. The minimum distance between two or more hyperplanes is called support vectors. The hyperplane predicts the sentiment of the tweet defined as:

$$H: w^T(x) + b \tag{6}$$

where *w* is a vector, *x* stands for the data points, and *b* stands for the bias.

In multi-class classification, there will be *n* hyperplanes, and the algorithm selects the hyperplane with the largest value.

$$H = \max_{i=1\ldots n} \{h_i|\, B_i\} \tag{7}$$

$$B = \min_{i=1\ldots m} |w \cdot x + b| \tag{8}$$

Random forest (RF)

RF is a decision tree algorithm that grows classification trees. The input vector grows multiple trees in the forest. The forest chooses the trees with the most votes to perform classification. RF is calculated by:

$$RF\; f_i = \frac{\sum_{j \in T} norm\; f_{ij}}{T} \tag{9}$$

In equation 9, *T* stands for all the trees in the forest, and *norm fij* stands for the normalised features that are important for every *i* in tree *j*. RF will not overfit the model if there are enough trees in the forest.

Extreme gradient boosting (XGBoost)

XGBoost is an optimised distributed gradient boosting for predictive modelling. It minimises the loss function by averaging weak classifiers for either binary or multi-class classification expressed as:

$$f(x) = \sum_{i=1}^{n} l(y_1 \widehat{y_1}^t) + \sum_{k=1}^{K} \alpha\, f(k) \tag{10}$$

where *l* defines the loss function, *K* stands for the sum of the trees, *n* is the number of rows in the data, *a* is the regularisation term, and *f* is the tree.

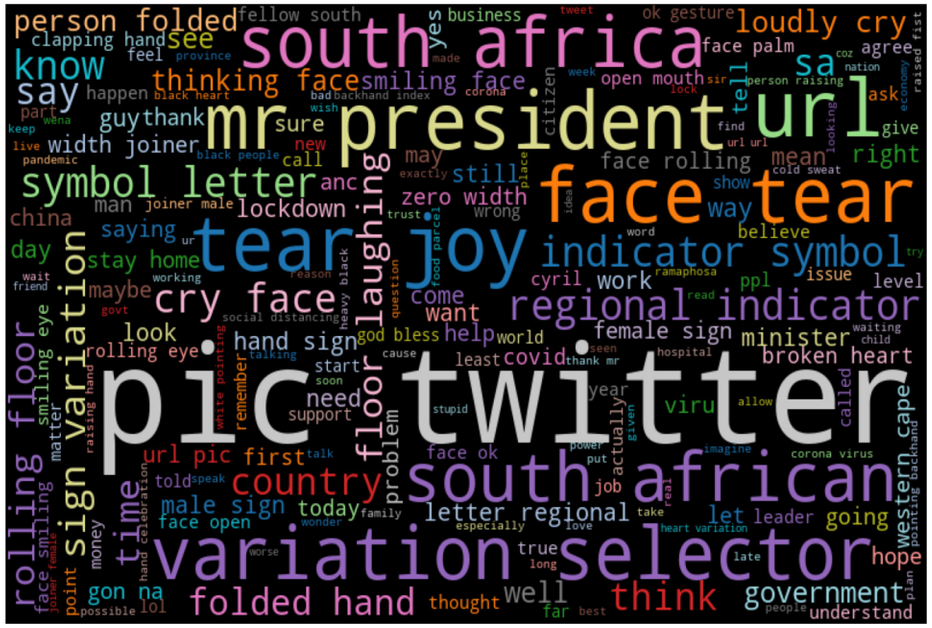
4. Results

Exploratory analysis

Before carrying out the downstream NLP tasks, we explored each of the three data subsets based on the compilation of a word cloud. A word cloud is a visualisation of text data, showing the frequency of words in a corpus: the higher the frequency, the bigger the word in the cloud.

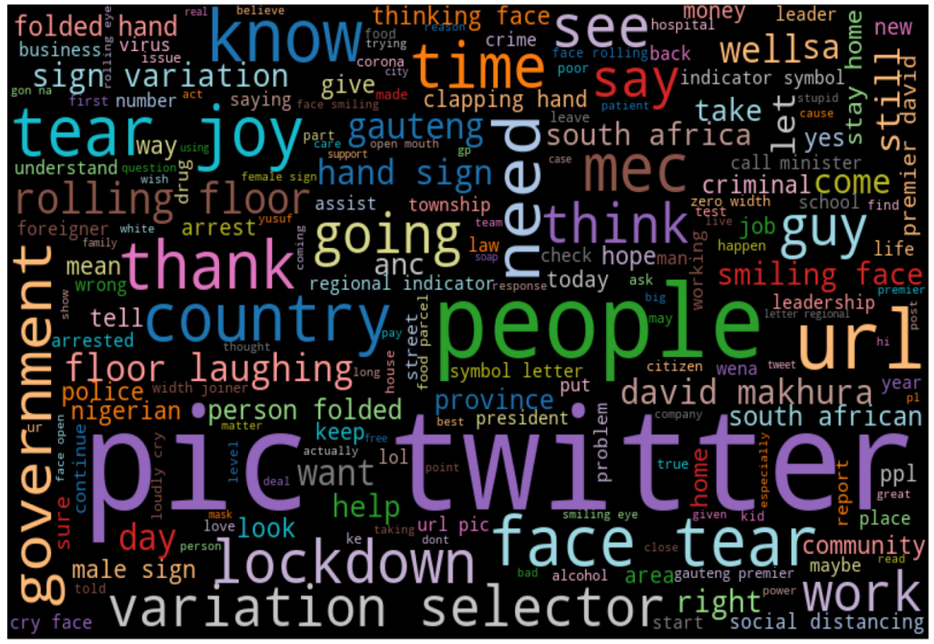
We generated one word cloud for each of the three data subsets, as shown in Figures 2, 3 and 4.

Figure 2: Most frequently occurring words in NICD data subset



In Figure 2, we see that in the NCID data subset, among the most prominent texts (some of which were emoticons converted to text) were “pic”, “twitter”, “variation selector”, “south africa”, “south african”, “mr president”, “tear”, “joy”, “face tear” and “url”.

Figure 3: Most frequently occurring words in Gauteng Province data subset



In Figure 3, we see that in the Gauteng Province data subset, among the most prominent texts (some of which were emoticons converted to text) were “pic”, “twitter”, “people”, “url”, “face tear”, “government”, “lockdown”, “variation”, “selector”, “country”, “thank”, “going”, “tear”, “joy”, “know”, “need”, “time”, “see”, “say”, “mec”, “think” “guy” and “work”.

The “mec” text stands for “Member of Executive Council (MEC)”, which is a provincial government cabinet minister in the South African system. The MEC being referred to was the then-MEC of Health for Gauteng Province, “David Makhura”, whose name also appears in Figure 3, below the word “people”.

Figure 4: Most frequently occurring words in Healthza data subset

In Figure 4, we see that in the Healthza data subset, among the most prominent words (some of which were emoticons converted to text) were "pic", "twitter", "url", "variation", "selector", "south africa", "lockdown", "country", "people", "need", "think", "know", "minister", "government", and "time".

Performance of the four sentiment classifiers

This section describes the performances of the four classifiers—LR, SVM, RF, and XGBoost—in sentiment classification of tweets related to the government's COVID-19 response in South Africa. As stated above, in order to mitigate the effects of class imbalance in the labelled dataset, we deployed SMOTE.

We utilised the TF-IDF vectoriser for feature selection, enabling a nuanced representation of the tweet content for analysis. Coupled with the use of SMOTE for dataset balancing, this approach aimed to reduce bias in the algorithmic training. Furthermore, we conducted a five-fold cross-validation to confirm the classifiers' robustness and generalisability. The importance of cross-validation in sentiment classification lies in its ability to offer a more reliable performance metric. It splits the data into several folds or subsets and, in each round, holds out one subset for testing while training the model on the remaining folds. Repeating this process until each fold has served as a test set assures a comprehensive evaluation that is less impacted by specific data splits.

Thus, we measured the classifiers' effectiveness based on their precision in classifying sentiments. Table 2 highlights the performance of the classifiers, indicating their accuracies and F1-scores.

Table 2: Performance of the classifiers in terms of training accuracy, testing accuracy and F1-score using a vectoriser (TF-IDF), SMOTE, and five-fold cross-validation

Classifier	Training accuracy	Testing accuracy	F1-score
Extreme gradient boosting (XGBoost)	0.84	0.83	0.82
Random forest (RF)	1.00	0.86	0.86
Logistic regression (LR)	0.91	0.93	0.93
Support vector machines (SVM)	0.91	0.93	0.93

Training accuracy, testing accuracy, F1-score

Accuracy is the fraction of correctly predicted polarities in the dataset.

$$Accuracy = \frac{true\ predictions}{total\ number\ of\ labels} \quad (11)$$

Sensitivity, specificity, and precision

For an unbalanced dataset with multi-class classification, accuracy is not enough to measure performance. We also had to evaluate sensitivity, specificity, and precision. Sensitivity or *recall* is the fraction of the number of correct predictions (true positives) over the proportion of positives. It is also known as the true positive rate.

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (12)$$

Specificity is the proportion of correct negative predictions over the total number of negatives.

$$Specificity = \frac{TN}{(TN + FP)} \quad (13)$$

Precision or positive predictive value is the number of correct positive predictions divided by the total number of positive predictions.

$$Precision = \frac{TP}{(TP + FP)} \tag{14}$$

True positive (TP) values are instances where the actual label is positive and correctly predicited as positive. False positive (FP) values are items where the actual label is negative, but the model incorrectly predicted the label as psoitive. True negative (TN) values are samples where the actual result is negative and is correctly predicted as negative. False negative (FN) values are items where the actual label is positive but incorrectly predicted as negative. The F1-score is the harmonic mean of precision and recall. It is micro-averaged over a multi-class classification. A weighted F1-score is reported in this study. The formula below calculated the F1-score.

$$F1 = 2\left(\frac{precision \times recall}{precision + recall}\right) \tag{15}$$

The weighted F1-score computes the F1-score for each class separately and then calculates a weighted average of those scores based on the samples in each class. This ensures that the metric is not biased towards the smaller classes. The weighted F1-score is expressed as:

$$F1_w = \sum C_w F1_i \tag{16}$$

where $F1_w$ is the weighted F1-score, C_w is the weight of class i calculated as the number of samples in class i divided by the total number of samples, and $F1_i$ is the F1-score for class i.

There are ways to calculate the accuracy, misclassification rate, true positive, false positive rate, true negative rate, precision, and prevalence from a confusion matrix. For example, in a confusion matrix, the samples featured in the cells diagonally from top left to bottom right represent correct predictions. The misclassified predictions are in the other cells.

Confusion matrices

We generated confusion matrices for each of the four algorithms—with the matrices based on the test data, i.e., the 59,007 tweets (from among the 295,035 labelled tweets) reserved for testing.

Figure 5: Confusion matrices for XGBoost and RF

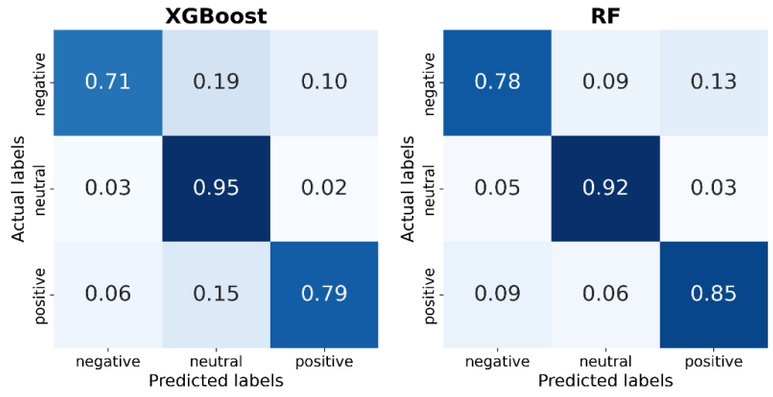


Figure 5 presents the normalised confusion matrices for the performance of the XGBoost and RF classifiers, with correct predictions as percentages. For XGBoost, the accuracy for negative, neutral, and positive classes stood at 71%, 95%, and 79%, respectively. For RF, the accuracies were 78% for negative classes, 92% for neutral, and 85% for positive.

Figure 6: Confusion matrices for SVM and LR

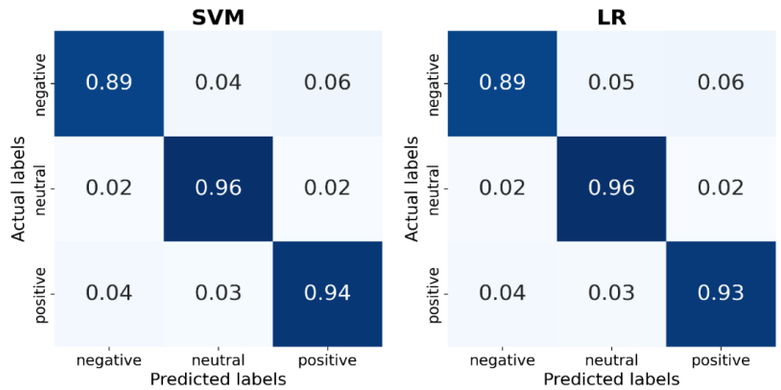


Figure 6 presents the normalised confusion matrices for the performance of the SVM and LR classifiers. Both algorithms accurately classified negative and neutral tweets, respectively, 89% and 96% of the time. However, for the positive class, SVM outperformed LR slightly, with a correct prediction rate of 94% compared to LR's 93%.

The SVM model was chosen to label the NICD and Gauteng Province data subsets, because its confusion matrix showed that it outperformed the other three models.

Figure 8: SVM distribution of class labels for the remaining two data subsets (NICD and Gauteng Province)

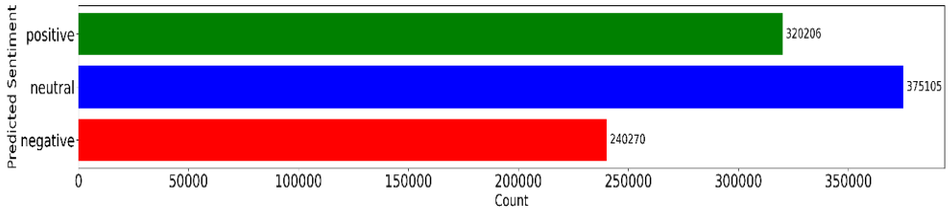


Figure 8 shows the sentiment distribution from the predictions of the SVM model when classifying the NICD and Gauteng Province data subsets. The predictions show that there were more neutral and positive tweets in both datasets than negative tweets. We now proceed to examine the words that that were contained in the predictions generated by the SVM model from classification of the NICD and Gauteng Province data subsets.

Figure 9: Top 20 words classified as negative by SVM

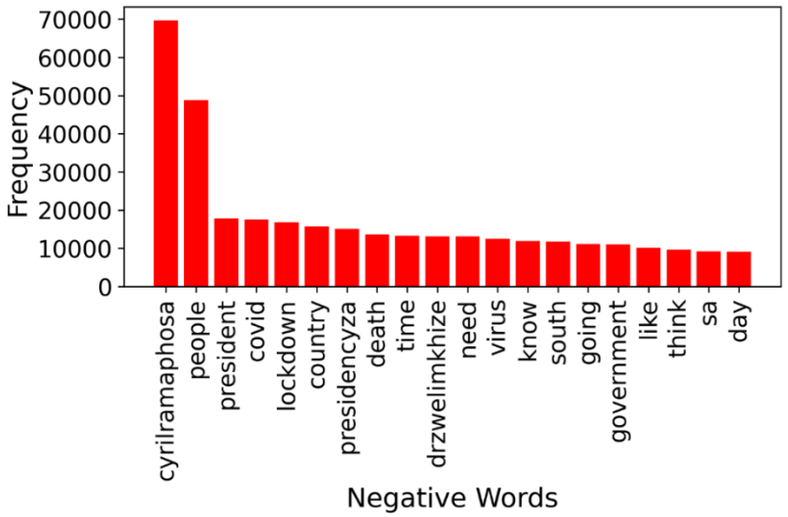


Figure 9 shows the top 20 words that the SVM model considered as negative, along with their frequency of occurrence. Topping the list was "cyrilramaphosa", which is not an inherently negative word, but which presumably often appeared within negative contexts.

Figure 10: Top 20 words classified as neutral by SVM

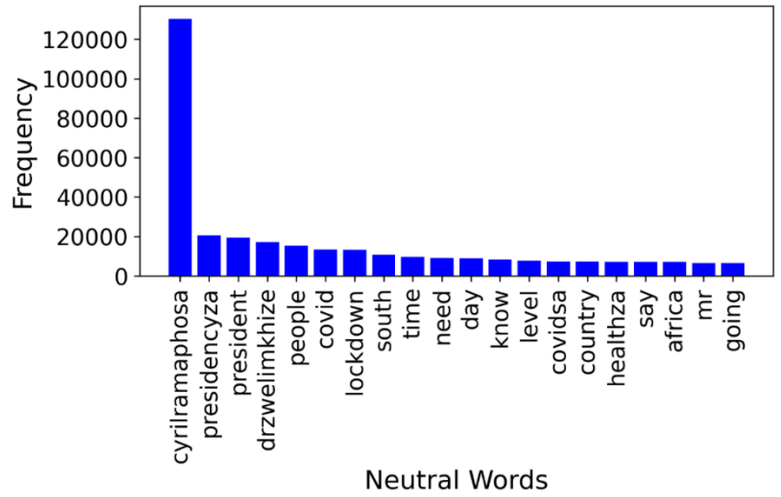


Figure 10 shows all the words classified by the model as neutral, including words such as "covid", "covidsa" and "lockdown" that would typically be expected to carry a negative sentiment. Because the SVM confusion matrix (in Figure 6) shows that SVM had 96% accuracy in labelling neutral words, it is relatively safe to assume that the context was indeed neutral for these instances of those three words as represented in Figure 10.

Figure 11: Top 20 words classified as positive by SVM

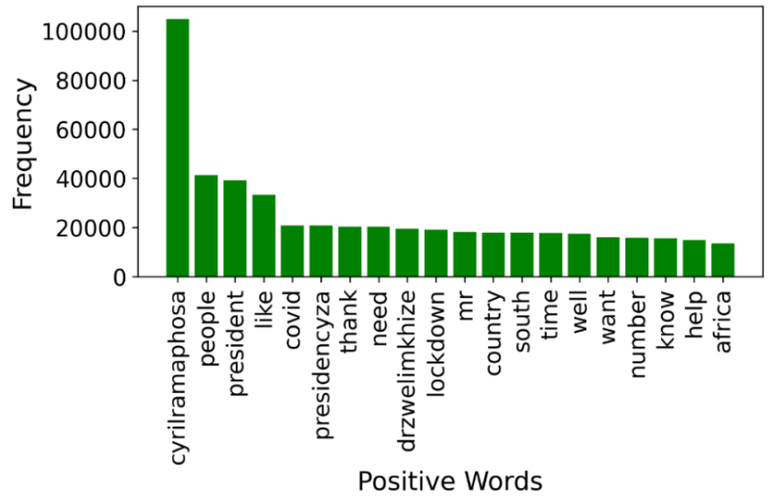


Figure 11 suggests that the SVM model was able, based on contextual factors, to recognise positive sentiment for words that typically would be expected to have

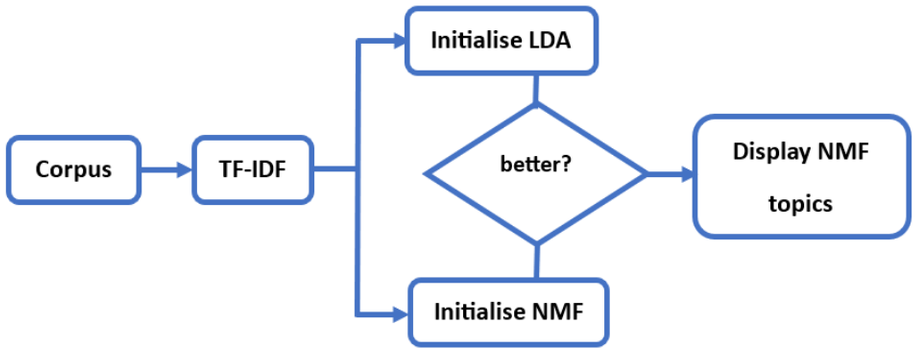
negative sentiment (e.g., covid, lockdown), and for neutral-seeming word such as "number". The instances of positive sentiment for "number" presumably resulted, for example, from tweets applauding a falling number of infections.

When looking across the three Figures (9, 10, 11), it can be noted that the words "cyrilramaphosa", "people", "president", "covid" and "lockdown" were represented in all three of the class labels (negative, neutral and positive), which is evidence that these words took on varying sentiments depending on their textual contexts.

Findings from topic modelling

As explained in the opening sections of this article, the study also included application of two topic modelling algorithms, in order to uncover the main themes, across the entire dataset, in the South African public's Twitter discourses on the government's handling of the pandemic. The two topic modelling algorithms used were LDA and NMF. The flow chart for topic modelling is provided in Figure 12.

Figure 12: Topic modelling flow chart



From the cleaned tweet corpus, a TF-IDF matrix was constructed. The parameters for the topic modeling were then systematically varied to evaluate which combinations produced meaningful topic distributions. In this study, the topics generated by NMF appeared to represent better algorithm performance than those generated by LDA—in other words, the topics produced by NMF were more coherent than those produced by LDA. Table 3 highlights the top 10 topics selected by NMF, along with the sentiments associated with each topic, and provides a thematic example for each one.

Table 3: Topics extracted by NMF

	Extracted topic	Sentiment	Thematic example
1.	Cyrilramaphosa presidencyza covidsa levellockdown alcohol locksouthafricadown dayoflockdown lockdownsa lockdownsouthafrica lockdownextension	Positive (0.06)	The lockdown announced by President Cyril Ramaphosa, to reduce the spread COVID, includes a ban on the sale of alcohol.
2.	people home need virus like country government black think stay	Negative (-0.11)	People are forced to stay at home by the government in our country because of the COVID virus.
3.	president mr ramaphosa nation well country address dear proud decision	Positive (0.25)	President Cyril Ramaphosa has addressed the nation and we are proud of his decisions aimed at keeping us well.
4.	presidencyza drzwelimkhize healthza governmentza mbalulafikile myanc dlaminizuma sapoliceservice minister juliussmalema	Positive (0.07)	The Presidency, Health Minister Zweli Mkhize, [Transport Minister] Fikile Mbalula, African National Congress (ANC), [Minister of Cooperative Governance and Traditional Affairs] Nkosazana Dlamini-Zuma, South African Police Service, [Economic Freedom Fighters leader] Julius Malema.
5.	south africa africans fellow african god country bless covid may	Positive (0.13)	Fellow South Africans, may God bless this country.
6.	lockdown day level week end going extended back extension month	Neutral (0.03)	This lockdown level is going to be extended to the end of the month.
7.	time right today like take address need nation long think	Positive (0.07)	The time is right today to address the nation.
8.	case number covid test death cape province recovery new confirmed	Positive (0.07)	The number of recovery cases from COVID in the Western Cape Province has been confirmed.
9.	thank god leadership minister work bless sir great mr may	Positive (0.63)	Thank God and bless you, minister, for your leadership and great work.
10.	know like want say need let well think right dont	Positive (0.16)	I want to say that we need to think well.

The predicted sentiments shown in Table 3 were the average sentiment values, with -1 indicating negative, 0 neutral, and 1 positive. With respect to the average values, the range between -0.05 and 0.05 was treated as representing neutral sentiment. The data suggested a prevailing positive sentiment among South Africans towards the government's pandemic response.

5. Conclusions and future research

The study has demonstrated the value of sentiment lexicons for training simple machine-learning models. It has also shown that these simple models can perform sentiment classification of social media tweets. In the study, the built model classified social media tweets into positive, negative, or neutral. The tweets were posts from the South African government, and from the public who responded, during the COVID-19 pandemic. The study examined the public's sentiment as South Africa's government imposed lockdown measures to reduce the spread of the virus.

Furthermore, the study showed that SMOTE effectively addresses dataset imbalances, and that TF-IDF is effective for feature selection and for identifying words that the model should pay attention to, and for cross-fold validation in enhancing model generalisation—as evidenced by the consistency of the F1-scores of test results when TF-IDF was used. The study also highlighted how simple topic modelling, such as that performed by NMF, could identify prevalent discussions during the COVID-19 pandemic. The topic-modelling findings suggested a general satisfaction, among the South African public, with the government's pandemic response. Such a finding can offer a valuable insight for policymakers and public stakeholders seeking to evaluate their interventions. Future work could involve exploring techniques for negation handling, or testing of the applicability of the models used in this study in domains beyond health and public safety.

References

Aljameel, S. S., Alabbad, D. A., Alzahrani, N. A., Alqarni, S. M., Alamoudi, F. A., Babili, L. M., Aljaafary, S. K., & Alshamrani, F. M. (2021). A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. *International Journal of Environmental Research and Public Health*, *18*(1), 1–12. https://doi.org/10.3390/ijerph18010218

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Chai, D., Wu, W., Han, Q., Wu, F., & Li, J. (2020). Description based text classification with reinforcement learning. Paper presented to International Conference on Machine Learning. http://proceedings.mlr.press/v119/chai20a/chai20a.pdf

Cruz, L., Ochoa, J., Roche, M., & Poncelet, P. (2015). Dictionary-based sentiment analysis applied to a specific domain. In *SIMBig 2015: Information Management and Big Data* (pp. 57–68). Springer. https://doi.org/10.1007/978-3-319-55209-5_5

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASI1%3E3.0.CO;2-9

Domalewska, D. (2021). An analysis of COVID-19 economic measures and attitudes: Evidence from social media mining. *Journal of Big Data*, *8*(1), 1–14. https://doi.org/10.1186/s40537-021-00431-z

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf

Ghebreyesus, T. A. (2020, March 11). WHO Director-General's opening remarks at the media briefing on COVID-19. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Gulati, K., Kumar, S. S., Boddu, R. S. K., Sarvakar, K., Sharma, D. K., & Nomani, M. (2022). Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic. *Materials Today: Proceedings*, *51*(1), 38–41. https://doi.org/10.1016/j.matpr.2021.04.364

Hou, K., Hou, T., & Cai, L. (2021). Public attention about COVID-19 on social media: An investigation based on data mining and text analysis. *Personality and Individual Differences*, *175*, 110701. https://doi.org/10.1016/j.paid.2021.110701

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/1014052.1014073

Hutto, C., & Gilbert, E. (2014, May). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225). https://doi.org/10.1609/icwsm.v8i1.14550

Jiang, W., Zhou, K., Xiong, C., Du, G., Ou, C., & Zhang, J. (2022). KSCB: A novel unsupervised method for text sentiment analysis. *Applied Intelligence*. https://doi.org/10.1007/s10489-022-03389-4

Jurafsky, D., & Martin, J. H. (2018). N-gram language models. In *Speech and language processing* [draft book]. https://web.stanford.edu/~jurafsky/slp3/3.pdf

Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. https://doi.org/10.3115/v1/S14-2076

Kuyo, M., Mwalili, S., & Okang'o, E. (2021). Machine learning approaches for classifying the distribution of Covid-19 sentiments. *Open Journal of Statistics*, *11*(5), 620–632. https://doi.org/10.4236/ojs.2021.115037

Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. arXiv preprint cs/0205028. https://doi.org/10.3115/1118108.1118117

Marivate, V., Moodley, A., & Saba, A. (2021). Extracting and categorising the reactions to COVID-19 by the South African public - A social media study. Paper presented to 2021 IEEE AFRICON. https://doi.org/10.1109/AFRICON51333.2021.9571010

Mendez-Brito, A., El Bcheraoui, C., & Pozo-Martin, F. (2021). Systematic review of empirical studies comparing the effectiveness of non-pharmaceutical interventions against COVID-19. *Journal of Infection*, *83*(3), 281–293. https://doi.org/10.1016/j.jinf.2021.06.018

Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. https://doi.org/10.18653/v1/P18-1017

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. arXiv preprint arXiv:1308.6242.

Moussa, M. E., Mohamed, E. H., & Haggag, M. H. (2020). A generic lexicon-based framework for sentiment analysis. *International Journal of Computers and Applications*, *42*(5), 463–473. https://doi.org/10.1080/1206212X.2018.1483813

Mutanga, M. B., & Abayomi, A. (2022). Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development*, *14*(1), 163–172. https://doi.org/10.1080/20421338.2020.1817262

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903. https://doi.org/10.48550/arXiv.1103.2903

Nwankwo, E., Okolo, C., & Habonimana, C. (2020). Topic modeling approaches for understanding COVID-19 misinformation spread in Sub-Saharan Africa. Paper presented to AI for Social Good Workshop. https://crcs.seas.harvard.edu/files/crcs/files/ai4sg_2020_paper_70.pdf

Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., & Johnson, D. (2005). Terrier information retrieval platform. In Advances in Information. Paper presented to 27th European Conference on Information Retrieval (ECIR 2005), Santiago de Compostela, Spain, 21–23 March. https://doi.org/10.1007/978-3-540-31865-1_37

Park, S.-J., Chae, D.-K., Bae, H.-K., Park, S., & Kim, S.-W. (2022). Reinforcement learning over sentiment-augmented knowledge graphs towards accurate and explainable recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. https://doi.org/10.1145/3488560.3498515

Ramaphosa, C. (2022, March 15). President Cyril Ramaphosa: Measures to combat coronavirus COVID-19 epidemic. https://www.gov.za/speeches/statement-president-cyril-ramaphosa-measures-combat-covid-19-epidemic-15-mar-2020-0000

Relman, D. A. (2020). To stop the next pandemic, we need to unravel the origins of COVID-19. *Proceedings of the National Academy of Sciences*, *117*(47), 29246–29248. https://doi.org/10.1073/pnas.2021133117

Rice, D. R., & Zorn, C. (2021). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods*, *9*(1), 20–35. https://doi.org/10.1017/psrm.2019.10

Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE*, *16*(2), e0245909. https://doi.org/10.1371/journal.pone.0245909

Soumya, S., & Pramod, K. (2021). Fine grained sentiment analysis of Malayalam tweets using lexicon based and machine learning based approaches. Paper presented to 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE). https://doi.org/10.1109/ICNTE51185.2021.9487741

Stone, P. J., & Hunt, E. B. (1963). A computer approach to content analysis: Studies using the general inquirer system. In *Proceedings of the May 21–23, 1963, Spring Joint Computer Conference.* https://doi.org/10.1145/1461551.1461583

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Tao, G., Miao, Y., & Ng, S. (2020). COVID-19 topic modeling and visualization. In *IV 2020: 24th International Conference Information Visualisation*. https://doi.org/10.1109/IV51561.2020.00129

Wan, X., Lucic, M. C., Ghazzai, H., & Massoud, Y. (2021). Topic modeling and progression of American digital news media during the onset of the COVID-19 pandemic. arXiv:2106.09572. https://doi.org/10.1109/TTS.2021.3088800

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, *39*(2), 165–210. https://doi.org/10.1007/s10579-005-7880-9

Xing, Y., Li, Y., & Wang, F.-K. (2021). How privacy concerns and cultural differences affect public opinion during the COVID-19 pandemic: A case study. *Aslib Journal of Information Management*, *73*(4), 517–542. https://doi.org/10.1108/AJIM-07-2020-0216

Yan, C., Law, M., Nguyen, S., Cheung, J., & Kong, J. (2021). Comparing public sentiment toward COVID-19 vaccines across Canadian cities: analysis of comments on Reddit. *Journal of Medical Internet Research*, *23*(9), e32685. https://doi.org/10.2196/32685

Zamani, M., Schwartz, H. A., Eichstaedt, J., Guntuku, S. C., Ganesan, A. V., Clouston, S., & Giorgi, S. (2020). Understanding weekly COVID-19 concerns through dynamic content-specific LDA topic modeling. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science* (pp. 193–198). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.nlpcss-1.21

CRITICAL INTERVENTION

China's digital transformation: Data-empowered state capitalism and social governmentality

Wayne Wei Wang
PhD Candidate, University of Hong Kong; and Fellow, CyberBRICS Project, Fundação Getulio Vargas (FGV) Law School, Rio de Janeiro
https://orcid.org/0000-0002-2278-9933

Abstract
The article scrutinises the trajectory of China's establishment of a digital state, rooted in a "whole-of-nation" system—or aptly termed (party–)state capitalism. The author illustrates the path of formulating and enforcing strategies to digitalise public services—including, importantly, the digital identity infrastructure—via institutional concentration that exemplifies both the positive and the exclusionary nature of social big data in streamlining administrative procedures. Two catalysts are spotlighted in China's digital transformation: quasi-neoliberal market processes, and technology's social change spillover effects. The author points to the fact that, since its inception, the contemporary Chinese state has created a cybernetic justification for "social governmentality", as a means to redress potential informational imbalances in the process of ruling the state polity. For the Chinese administrative hierarchy, data provides the means to execute a top-down correctivist paradigm for steering societal conduct, a paradigm integrated into (but also to some extent in tension with) data-empowered state capitalism.

Keywords
China, state capitalism, data silos, digital government, identity infrastructure, social big data

Acknowledgement
This article is a synthesis drawing on portions of a forthcoming CyberBRICS[1] project report.

DOI: https://doi.org/10.23962/ajic.i31.16296

Recommended citation
Wang, W. W. (2023). China's digital transformation: Data-empowered state capitalism and social governmentality. *The African Journal of Information and Communication (AJIC), 31*, 1-13. https://doi.org/10.23962/ajic.i31.16296

1 See https://cyberbrics.info. The BRICS countries are Brazil, Russia, India, China, and South Africa.

1. Introduction

A core component of socialism with Chinese characteristics is the term "modernisation" (*xian dai hua*, 现代化), which was ideologically embedded in the earliest-phase Four Modernisations (四个现代化) national strategy and would continue to be central to the Chinese economic and political agenda on poverty reduction and the growth of the economy.[2] Among the Four Modernisations, Scientific and Technological (S&T) Modernisation was central. Then-Paramount-Leader Deng Xiaoping stated that the mastery of S&T was imperative for socialist modernisation (Schneider, 1981). This signalled the embryonic form of Chinese cybernetics in the national innovation system and the institutional intention of Marxism to include technology under political authority (Cong & Thumfart, 2022, p. 4). The terminology "Fifth Modernisation", namely Modernising the State Governance System and Capacity, was coined in 2013 (Central Committee of the Chinese Communist Party (CCP), 2013), and focused on enabling the adjustment of the state–government–market relationship so as to improve Chinese statecraft. In the years since, technology has gradually evolved from being an innovative element driving economic growth to being a governance tool that rules society, thus transforming the bureaucratic structure of China's public administration.

The article probes the impetus propelling China to cultivate its digital state, embodying a form of digital sovereignty that distinctly diverges from the perceived "Western Path" of postcolonial datafication, also claimed as digital colonialism (Cong & Thumfart, 2022, p. 2). Through tracing the genealogy of Chinese law and policy reforms in support of digital transformation, this article shows how China has arrived at its "whole-of-nation" system ("举国体制")—or, more precisely, its system of (party–)state capitalism (Hsueh, 2016; Y. Huang, 2008; Milhaupt & Zheng, 2014; Pearson et al., 2021)This system underlies the development of numerous strategies to digitise public services—particularly through the digital identity infrastructure—as an institutional concentration that illustrates both the positive and the exclusionary nature of social big data. Using social big data can enable governments to improve weak areas of public administration (e.g., shifting from inefficient and neglectful administration to efficient and targeted administration). However, it can also

2 Before establishing the long-term "Reform and Opening-up" tactic, Deng furthered the Four Modernisations, first proposed by Zhou Enlai in 1954, as an economic means of strengthening China's agriculture, industry, defence, and science and technology sectors.

advance the use of digital state infrastructure as a form of observation or social governmentality (Foucault, 2001, p. 201, on the concept of "governmentality"), with a high degree of data granularity.

2. Driving forces in digitalisation of public services

In 2001, when the United Nations first released the e-Government Index, China was classified as having "minimal e-gov capacity", with a relatively low ranking of 1.04 compared to the then global average of 1.62, while the 2022 UN E-Government Survey found that China now has a "very high" E-Government Development Index (EGDI) of 0.8119 (out of 1.0) (UN, 2022, p. 214)—a major shift over a period of 20 years. I now turn to documenting how China has completed its substantial and phased digital transformation at a pace that is described by some as a Schumpeterian industrial policy model of technological leapfrogging and catching-up (Lee, 2022).

Stage one: State informatisation (the 1990s)

The Chinese government first employed state informatisation—state-directed IT-facilitated modernisation (Zou, 1997, p. 6)—to establish a top-down telecommunications network across all levels of government, with the aim of managing economic databases and improving administrative efficiency. For example, the Golden Projects (三金工程) of the 1990s, all e-government focused, included the Golden Bridge Project (金桥工程) focused on internet provision, the Golden Customs Project (金关工程) linking customs entities, and the Golden Card Project (金卡工程) that established a national bank card network (Zhao, 1995). In 1999, China began the Government Internet Access Project (政府上网工程) as an e-government incubator.

Although these IT systems of the 1990s empowered a government-centric, bureaucratic, confidential state, with office automation only in the internal administration and without IT embedded in general governmentality, the state informatisation established the model of "state-directed co-developed standardised interconnection" (Zou, 1997, p. 6) to underpin all levels of the administration's more statistics-informed, macro-control capabilities.

Stage two: E-government data silos (2000–2014)

The Chinese E-Government Guidance and Framework were released in 2002 and 2006, respectively, to accelerate institutional changes and improve the service quality and efficiency of central and local governments (National Informatisation Leading Group, 2006; Office of the Central Committee of the CCP, 2002), resulting in the expansion of informatisation architecture from the Golden Projects to Two Nets, One Website, Four Databases, and Twelve Golden Projects (两网，一站，四库，十二金工程). The "two nets" were the government intranet and the government extranet; the "one website" was the gov.cn domain; the "four databases" were the Population, Legal Entity, Macro-Economy, and Spatial Geography and Natural

Resources databases; and the "twelve Golden Projects" were focused on e-government in numerous sectors, including Finance, Social Welfare, and Agriculture.[3]

Accordingly, the conventional information asymmetry of power in Chinese hierarchical systems was somewhat reduced. However, the hierarchy's organisational structure resulted in different parts of government forming data silos because both vertical and horizontal administrative departments built and maintained data pools independently, without standardising inter-unit data sharing protocols (Zheng, 2007, p. 119). Meanwhile, the increase in civic participation in (new) social media—facilitated by, for example, the launch in 2009 of the Sina Weibo microblogging Gov Account (Yu, 2016)—sparked demand for increased public access to information, and, at the same time, increased state interest in responding to public sentiments to enhance the credibility of the government.

Stage three: Towards Digital China (since 2014)

The deployment of e-government continued in conjunction with the 13th five-year plan (2016–20), through the adjustment and integration of national databases on demography, legal entities, natural resources, spatial geography, macroeconomics, cultural content, and social credits (H. Huang, 2020b, p. 12). Some databases, including the macroeconomic database, were scrapped during construction, while some, like the demographic database, were merged with a more centralised system (H. Huang, 2020a, pp. 50–51). Due to the lack of audits and evaluations of government websites, the Chinese government's efforts to expand e-services and bridge the digital divide between rural and urban areas were largely unsuccessful (H. Huang, 2020a, pp. 51–52). In response, there were administrative campaigns to build a New Media Matrix for Government Affairs and palmtop services (Office of the State Council, 2018a), by developing multidirectional channels, including gov.cn, WeChat public accounts, Weibo public accounts, and WeChat mini programmes and administrative mobile apps.[4]

In 2015, China adopted the Internet Plus initiative focused on using the internet as an empowering infrastructure to enhance mass productivity and galvanise state innovation and entrepreneurship (State Council, 2015b). As part of the initiative, the national Internet + Public Services plan was devised and implemented in accordance

3 See http://www.e-gov.org.cn/article-166340.html

4 See the New Media Matrix for Government Affairs at https://app.www.gov.cn/govdata/zwxmtjz.html. (Government-operated accounts—*zheng wu hao*, 政务号—authenticated and managed in the name of primary government departments, function as media accounts, such as those verified for governmental interaction with the public on Weibo. Analogously, WeChat Public Accounts serve as a distinctive feature within the WeChat application, empowering individuals or organisations to disseminate content, cultivate a following, and sell their goods and services. Complementarily, WeChat Mini Programmes, intrinsically mobile applications that operate within the WeChat ecosystem, offer accessibility without necessitating separate downloads or installations.)

with the procedural design model employed by user-friendly e-commerce platforms (Office of the State Council, 2018b). A quintessential example is China's Internet courts, which have essentially platformised electronic litigation for the parties involved, online case filing by the courts, and online court hearings presided over by judges.[5] At the same time, the central government introduced the national Internet + Regulations System (the National Integrated Online Regulatory Platform).[6] This system, enabled in part by Alibaba's cloud computing services (Aliyun), provided *ex ante* regulatory prediction, regulatory effectiveness assessment, credit modelling, and user interface (UI) design.[7] One example is Hangzhou's City Brain that adopts Aliyun's Super Artificial Intelligence, automatically deploying public resources and correcting bugs in the city's operation.[8] It marked a significant shift in the way the Chinese government reshaped bureaucratic governance, from office automation to automated-decision-making-assisted administration, by utilising emerging technological tools and market-driven mechanisms.

In 2021, Section Five of China's 14th five-year plan (2021–25) emphasised the importance of accelerating digital growth and creating a Digital China (*Xinhua News Agency*, 2021). This call for a Digital China had been preceded by a series of policy documents promoting the development of a national integrated online government service platform,[9] an "all-in-one" portal,[10] and a cross-provincial digital public service platform.[11] The Digital China strategy primarily hinges upon the fortification of digital infrastructure and data resource systems, with digital governance serving as a corollary facet within this overarching agenda. This is predicated upon a symbiotic public–private collaboration, leveraging user-centric amenities, such as accessibility—a salient attribute previously associated predominantly with private platform business models in China. In 2022, the 14th Five-Year Plan's ancillary schemes concerning State Informatisation and Public Services amplified the provision of accessibility

5 See Beijing Internet Court at https://www.bjinternetcourt.gov.cn/; Guangzhou Internet Court at https://ols.gzinternetcourt.gov.cn/; and Hangzhou Internet Court at https://www.netcourt.gov.cn/

6 See the National Integrated Online Regulatory Platform portal at http://www.jianguan.gov.cn/

7 See the overview of the solution demo for the Aliyun-powered Internet + Regulations model at https://cn.aliyun.com/solution/govcloud/intetregulation?from_alibabacloud=

8 See the Hangzhou City Brain Co. LTD. at https://www.cityos.com

9 See the 2018 Guiding Opinions of the State Council on Speeding up the Construction of the National Integrated Online Government Service Platform (国务院关于加快推进全国一体化在线政务服务平台建设的指导意见), the 2019 Provisions of the State Council on Online Government Service (国务院关于在线政务服务的若干规定), and the 2021 Guide to the Construction of the Mobile End of the National Integrated Government Service Platform (全国一体化政务服务平台移动端建设指南).

10 See the 2018 Implementation Plan on Deepening "Internet + Government/Public Service" and Promoting the Reform of "one Web Portal, one Door and one Time" of the Government Affairs Service (进一步深化"互联网+政务服务"推进政务服务"一网、一门、一次"改革实施方案).

11 See the 2020 Guidance of the General Office of the State Council on Accelerating the "Cross-provincial" Government Services (国务院办公厅关于加快推进政务服务"跨省通办"的指导意见).

for the disabled and marginalised demographics. This enhancement was previously evaluated via the recalibration of web and application. The private sector also follows the trend. A typical example is the "Caring Mode" (*guan huai mo shi*, 关怀模式) of WeChat, particularly prototyped for the elderly and visually impaired groups.

Hence, originating from state informatisation and progressing through inter-departmental data silos, all while pioneering user-friendliness, China has accomplished a rapid and phased digital transition in the public sector over the past few decades, utilising a resource-concentration industrial economic model. Particularly, the spillover effects of platformisation—including incorporating super apps like WeChat—have prompted the societal implications of digital technology to embed into administrative hierarchies, thereby acting as the impetus for the digital transformation of public services.

3. Streamlining of government data stewardship

A power-concentrated body, the National Informatisation Leading Group, was created in 1996 with the goal of overseeing China's transition to digital technology (Office of the State Council, 1996). The Ministry of Industry and Information Technology (MIIT) absorbed the Informatisation Office of the State Council in 2008 (Wang, 2014). It was decided in 2014 to reclassify MIIT's Informatisation Promotion Department as the Cyberspace Administration of China (CAC), signalling that one of the CAC's major responsibilities (especially in its Information Development Bureau) would be to coordinate China's e-government initiatives, in concert with other agencies playing a variety of roles in digitising public services (H. Huang, 2020a, p. 50).

The 2018 Plan for Deepening the Reform of Party and State Institutions specified that the Central Cyberspace Affairs Commission (CCAC), a CCP body, would oversee the CAC and the country's cybersecurity and informatisation (Central Committee of the CCP, 2018).

The determination of critical information infrastructure in China has had a gradual evolution from the Cybersecurity Law of 2016 to a complex system of rules involving the Personal Information Protection Law (PIPL), the Data Security Law (DSL), and Regulations on Critical Information Infrastructure (CII) Security Protection (CII Regulation), all of which were put in place in 2021. These laws demonstrate that the national security involved in data management is regarded as more relevant than the advantages to be gained from market-based data.

A broad variety of other government organisations, each with a particular role to play in the process, have also been included in the digitisation of public/government services. The National Development and Reform Commission, more precisely the State Information Centre, makes plans that primarily reflect sharing information resources at the national level (H. Huang, 2020a, p. 51). At the same time, the General Office of the State Council is responsible for overseeing the disclosure of government information (i.e., the Operational Centre for gov.cn) (Office of the State Council, 2008). The Ministry of Industry and Information Technology continues to collaboratively coordinate the development of telecommunications, the internet, and private communication networks, in accordance with the 2018 Plan for Deepening the Reform of Party and State Institutions (Central Committee of the Chinese Communist Party (CCP), 2018).

As the examples above demonstrate, the central government and the CCP have taken numerous steps to streamline government data stewardship. Yet because traditional Chinese administration paid little attention to cross-departmental collaboration, information and/or data ramparts continue to exist. Traditionally, administrations did not simply perceive power divisions as the rationale for a physical separation of data, but instead as an inherently different model of data structures, whose tightness could reduce risks of accountability.

China's digital transformation has, in recent years, undergone a transition from governing society with data to regulating data for society. Data was traditionally viewed as a catalyst for emerging technologies such as big data analytics, artificial intelligence, and blockchain technology in the context of the national innovation system. In 2015, the State Council issued the Action Plan for Promoting the Development of Big Data (State Council, 2015a), which clearly stated that data is a "fundamental strategic state resource". In March 2020, the CCP Central Committee and the State Council released Opinions on Strengthening the Institutional Mechanism for Market-Based Allocation of Elements (Central Committee of the Chinese Communist Party (CCP) & State Council, 2020), which defined data as a key market element. In a sense, the Chinese bureaucracy sees data as both a source of bottom-up information feedback and a complement to its top-down state-dominated capitalism—and, as never before, as a productivity material that can be subject to state-drawn ownership rules and thus constructed as a data market.

Following the Internet Plus initiative, the First Research Institute of the Ministry of Public Security piloted Version 1.0 of the Cyber Trusted Identity (CTID) system in 2016.[12] With the original aspiration of privacy-enhanced technology at its core,

12 See the brief history of the CTID platform at http://www.anicert.cn/platform.html?md=2

the CTID platform desensitises and de-identifies the personal identifiers in the legal identity files, generates an irreversible data file devoid of plaintext information, and maps the documents to the legal identity files one by one.[13] The CTID uses SIM card digital identity, blockchain, two-dimensional code, cryptography, and facial recognition to enable online identification without requiring the user to disclose written and plaintext personal information (Bao & Wu, 2020, p. 117; Gao et al., 2021, p. 91), allegedly resulting in reduced breaches of sensitive personal data.

During the COVID epidemic, China's Health Code system with a three-colour dynamic interface, which was developed based on the underlying CTID technology in 2020, further strengthened the process of using digital identity for social governance. The former constitutes a derivative technological advancement predicated upon the latter (CTID, 2021). Since it is connected to the National Integrated Big Data Platform and relies on widely used payment apps like Alipay and WeChat Pay (Liang, 2020, p. 1), it has a federal character in use. Health codes go beyond identification, and their combination of personal information, health status, residence, travel history, exposure history, and other supporting data plays a socio-technical role in epidemic risk ranking and migration control in epidemic prevention (Cong, 2021, p. 4).

According to the latest policy document (Office of the State Council, 2022), China's extensive collection of big data is focused on two main categories: basic databases and thematic databases. The basic databases are expected to contain data on dimensions such as population, legal persons, natural resources, economy, and electronic certificates.[14] The thematic databases are expected to cover dimensions such as health care, government services, social security, ecological and environmental protection, the credit system, emergency management, and supervision of state-owned assets. The process of centralised data collection and cleansing is known as fusion and aggregation (*gui ji*, 归集). Governmental data integration attempts to facilitate transformative data fusion by distinguishing at the policy level between narrowly defined government data (data generated and collected in the course of government operations), industrial data (sectoral data collected by official units), public data (data collected and generated by public utilities), and social big data (data collected and generated by third-party internet platforms).[15]

13 See the definition of the CTID platform at http://www.anicert.cn/identity.html

14 In China, Electronic Certificates (*dian zi zheng zhao*, 电子证照) typically refer to all kinds of licences, certificates, approvals, appraisal reports, office results, and other documents issued by various units in accordance with the law and with legal effect. For example, there are ID cards, marriage certificates, bank repayment flow certificates, business licences, etc.

15 The aforementioned initiative, commonly referred to as "Internet + Regulations", compelled certain platforms to disclose specific operational data, including, for instance, the identifying information and tax-related particulars of platform operators, in accordance with Article 28 of the 2018 Chinese E-Commerce Law.

4. Conclusions

This article has identified two driving forces behind the Chinese digital transformation: applying quasi-neoliberal market mechanisms, and using technology's spillover effects to influence governmentality. The modern Chinese state has developed a distinct cybernetic rationale for social governmentality since the age of Four Modernisations—with cybernetic social governmentality tools used as a means to redress a potential imbalance of information in the process of ruling the state polity (where local administrators once disregarded the upward institutional feedback of data and statistics pertinent to governance).

To a certain degree, a discernible absence of normative congruence can be observed across the aforementioned approaches of improving public administration, transforming towards digital government, and exercising social governmentality, stemming from the inherent tension between utilising big data for social governmentality and monetising said data. This tension has also manifested geographically in the perception of data sovereignty within China. Data sovereignty with Chinese characteristics signifies a conceptual fusion between preserving the political fabric and invigorating economic growth, which, in turn, accentuates an inherent dichotomy/tension that exists between cybersecurity and digital economy.

As a result of using social big data for platformising public administration, Chinese state capitalism has been invigorated, whereby data gained from grassroots civic participation in e-government services provides real-time information feedback loops, as well as data that can be monetised. Technology assumes a compulsory intermediary role, facilitating the administrative ability to extract and cultivate governance capabilities centred around efficiency, directly derived from evidence-based behavioural data.

Technology is perceived by the state as serving as: a probative catalyst for ideology; a geopolitical symbol of sovereign independence; an economic element of innovation; and, most frequently, a normative instrument of government.

In the case of the public sector's digital transformation, data is frequently more than just data; it represents a top-down correctivist paradigm of social relations. The state's constitutional foundations and, as a result, citizens' rights, may be somewhat vulnerable as a result of the digital non-scarcity that develops when the state gathers large amounts of data on individuals—through, for example, decentralised/federalist abuse of digital technology (Horwitz, 2022). This dichotomy between positive and exclusionary use of social big data—between culturally and economically reimagined monitoring—thus deserves further normative and empirical investigation so that it can be understood and remedied at a micro level, at a behavioural level, at a cognitive level, and, most significantly, at an institutional level beyond socio-technical imagination.

References

Bao, M., & Wu, Y. (2020). 如何证明你是你——数字身份识别在金融中的应用 [How to prove that you are you?—Application of digital identification in finance]. 金融市场研究 *[Financial Market Research]*, *5*, 113–120.

Central Committee of the Chinese Communist Party (CCP). (2013). 关于全面深化改革若干重大问题的决定 [Decision on Several Major Issues of Comprehensively Deepening Reform]. http://www.gov.cn/jrzg/2013-11/15/content_2528179.htm

Central Committee of the Chinese Communist Party (CCP). (2018). 深化党和国家机构改革方案 [Plan for Deepening the Reform of Party and State Institutions]. http://www.gov.cn/zhengce/2018-03/21/content_5276191.htm#1

Central Committee of the Chinese Communist Party (CCP), & State Council. (2020). 关于构建更加完善的要素市场化配置体制机制的意见 [Opinions on Strengthening the Institutional Mechanism for Market-based Allocation of Factors]. http://www.gov.cn/zhengce/2020-04/09/content_5500622.htm

Chi, S. (2021, May 20). Alibaba: Punishment and collaboration. *Verfassungsblog*. https://verfassungsblog.de/alibaba/

Cong, W. (2021). From pandemic control to data-driven governance: The case of China's Health Code. *Frontiers in Political Science*, *3*. https://doi.org/10.3389/fpos.2021.627959

Cong, W., & Thumfart, J. (2022). A Chinese precursor to the digital sovereignty debate: Digital anti-colonialism and authoritarianism from the post-Cold War era to the Tunis Agenda. *Global Studies Quarterly*, *2*(4), ksac059. https://doi.org/10.1093/isagsq/ksac059

Cyber Trusted Identity (CTID). (2021). 公安部一所参与制定的个人健康信息码国家标准发布 [A national standard for personal health information codes, developed with the participation of the Ministry of Public Security, was released]. Weixin Official Accounts Platform. https://tinyurl.com/y9jqkmwm

Foucault, M. (2001). *Power*, vol. 3 (J. D. Faubion (Ed.), R. Hurley (Trans.)). The New Press.

Gao, L., Li, J., Guo, W., Kang, L., Li, N., & Wang, Y. (2021). 创新社会治理 赋能数字经济——可信数字身份创新成果巡礼 [Innovative social governance empowering the digital economy—Trusted digital identity innovation achievement tour]. 警察技术 *[Police Technology]*, *3*, 90–93.

Horwitz, J. (2022, June 23). Chinese officials punished for changing health codes of bank depositors. *Reuters*. https://www.reuters.com/article/china-banks-henan-idINL4N2YA03D

Hsueh, R. (2016). State capitalism, Chinese-style: Strategic value of sectors, sectoral characteristics, and globalization. *Governance*, *29*(1), 85–102. https://doi.org/10.1111/gove.12139

Huang, H. (2020a). 中国"数字政府"的政策演变——兼论"数字政府"与"电子政务"的关系 [The policy evolution of "digital government" in China—The relationship between "digital government" and "e-government"]. 行政论坛 *[Administrative Tribune]*, *27*(03), 47–55.

Huang, H. (2020b). 数字政府:政策、特征与概念 [Digital government: Policies, characteristics and concepts]. 治理研究 *[Governance Studies]*, *36*(3), 6–15.

Huang, Y. (2008). *Capitalism with Chinese characteristics: Entrepreneurship and the state* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511754210

Lee, K. (2022). *China's technological leapfrogging and economic catch-up: A Schumpeterian perspective*. Oxford University Press. https://doi.org/10.1093/oso/9780192847560.001.0001

Liang, F. (2020). COVID-19 and Health Code: How digital platforms tackle the pandemic in China. *Social Media + Society*, *6*(3), 2056305120947657. https://doi.org/10.1177/2056305120947657

Milhaupt, C. J., & Zheng, W. (2014). Beyond ownership: State capitalism and the Chinese firm. *Georgetown Law Journal*, *103*, 665–722.

National Development and Reform Commission. (2021a). 十四五"公共服务规划 [The 14th Five-Year Public Service Plan]. https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/202201/P020220110357049883156.pdf

National Development and Reform Commission. (2021b). 十四五"推进国家政务信息化规划 [The 14th Five-Year National Informatisation Plan]. http://www.gov.cn/zhengce/zhengceku/2022-01/06/5666746/files/cbff2937df654b44a04c6ef9d43549e9.pdf

National Informatisation Leading Group. (2006). 国家电子政务总体框架 [The Overall Framework of National E-Government]. https://www.mct.gov.cn/whzx/zxgz/whbwlaqhxxhgz/xxhjs_whaq/201111/t20111129_800690.htm

Office of the Central Committee of the Chinese Communist Party (CCP). (2002). 国家信息化领导小组关于我国电子政务建设指导意见 [Guidance from the National Informatisation Leading Group on the Construction of E-government in China]. https://www.neac.gov.cn/seac/zcfg/200406/1075051.shtml

Office of the State Council. (1996). Notice Concerning Establishing the State Council Leading Group for Informatization Work. https://digichina.stanford.edu/work/state-council-general-office-notice-concerning-establishing-the-state-council-leading-group-for-informatization-work/

Office of the State Council. (2008). 国务院办公厅主要职责内设机构和人员编制规定的通知 [State Council Office Circular on the Regulations on the Establishment and Staffing of the Main Functions]. http://www.gov.cn/zhengce/content/2008-07/17/content_7616.htm

Office of the State Council. (2013). 关于进一步加强政府信息公开回应社会关切提升政府公信力的意见 [Opinions on Strengthening Government Information Disclosure and Enhancing the Credibility of the Government]. http://www.gov.cn/zhengce/content/2013-10/18/content_1219.htm

Office of the State Council. (2018a). 关于推进政务新媒体健康有序发展的意见 [Opinions on Promoting the Healthy and Orderly Development of New Media for Government Affairs]. http://www.gov.cn/zhengce/content/2018-12/27/content_5352666.htm

Office of the State Council. (2018b). 进一步深化“互联网+政务服务”推进政务服务“一网、一门、一次”改革实施方案 [Implementation Plan on Deepening “Internet + Government/Public Service” and Promoting the Reform of “One Web Portal, One Door and One Time” of Government Affairs Service]. http://www.gov.cn/zhengce/content/2018-06/22/content_5300516.htm

Office of the State Council. (2022). 全国一体化政务大数据体系建设指南 [Guidelines for the Construction of a National Integrated Big Data System for Government Services]. http://www.gov.cn/zhengce/content/2022-10/28/content_5722322.htm

Pearson, M., Rithmire, M., & Tsai, K. S. (2021). Party–state capitalism in China. *Current History*, *120*(827), 207–213. https://doi.org/10.1525/curh.2021.120.827.207

Schneider, L. A. (1981). Science, technology and China's four modernizations. *Technology in Society*, *3*(3), 291–303. https://doi.org/10.1016/0160-791x(81)90001-4

State Council. (2015a). 关于印发促进大数据发展行动纲要的通知 [Action Plan for Promoting the Development of Big Data]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm

State Council. (2015b). 关于积极推进“互联网+”行动的指导意见 [Guiding Opinions on Actively Promoting Internet Plus Actions]. http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm

State Council. (2021). 十四五”数字经济发展规划 [The 14th Five-Year Digital Economy Plan]. http://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm

Third Research Institute of the Ministry of Public Security. (2018). eID数字身份体系白皮书 [White Paper on the Digital Identity System of eID]. https://tinyurl.com/2zjkknzg

United Nations (UN). (2022). *E-Government survey 2022: The future of digital government*. https://desapublications.un.org/sites/default/files/publications/2022-11/Web%20version%20E-Government%202022%20November%2010.pdf

Wang, Y. (2014, March 10). 我国信息化领导小组的发展和演变 [The Development and Evolution of the Informatisation Leading Group]. *CNII*. https://web.archive.org/web/20180208182757/http://www.cnii.com.cn/informatization/2014-03/10/content_1319483.htm

Wen, C., & Yachang, H. (1999). 网络社会第一站——从启动“政府上网工程”谈起 [The first stop of the Internet society—Starting with the launch of the “Government Internet Access Project”]. 中国计算机用户 *[China Computer Users]*, *05*, 37–39.

Xinhua News Agency. (2019, November 5). 中共中央关于坚持和完善中国特色社会主义制度 推进国家治理体系和治理能力现代化若干重大问题的决定 [Decision of the Central Committee of the Chinese Communist Party on Several Major Issues Concerning the Adherence to and Perfection of the Socialist System with Chinese Characteristics and the Advancement of the Modernization of the State Governance System and the Ability to Govern]. http://www.gov.cn/zhengce/2019-11/05/content_5449023.htm

Yu, Y. (2016, January 28). 政务微博待补三大短板 [Three shortcomings of government microblogs to be filled]. *People.Cn* [人民网]. http://politics.people.com.cn/n1/2016/0128/c1001-28090656.html

Zhao, Z. (1995). 信息安全与“三金工程” [Information security and “Three Golden Projects”]. 科技导报 *[Science & Technology Review]*, *13*(9504), 29–32. http://www.kjdb.org/CN/abstract/abstract5566.shtml

Zheng, L. (2007). Enacting and implementing open government information regulations in China: Motivations and barriers. In *Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance* (pp. 117–120). https://doi.org/10.1145/1328057.1328083

Zhong, J. (2020, September 17). 数字政务“云钉一体”的实践 [The integration of cloud and Dingding for government affairs]. *China Daily*. https://caijing.chinadaily.com.cn/a/202009/17/WS5f631fc6a3101e7ce9725000.html

Zou, J. (1997). 加快推进国家信息化 [Accelerating state informatisation]. 信息系统工程 *[CC News]*, *11*, 6–7.

The African Journal of Information and Communication (AJIC)

Published by the LINK Centre
University of the Witwatersrand (Wits)
Johannesburg, South Africa
https://www.wits.ac.za/linkcentre

ISSN 2077-7213 (online version)
ISSN 2077-7205 (print version)