

Visually mapping the interplay between pandemic interest groups and 'the vulnerable' in newspaper accounts, 2018–2022

Stuart Neilson and Neil Kenny

Introduction

The COVID-19 pandemic had an unprecedented impact on society, presenting a valuable potential case study of how media represents a societally important topic. We use a visual mapping approach to illuminate and discuss the patterns that emerge within media narratives and coverage relevant to the COVID-19 pandemic. Our approach involves the creative analysis of 'big data' extracted from media reportage in order to create visually accessible depictions of emerging patterns within the analysis. We analysed the contrasting roles played by 'vulnerable' groups, 'mental health', and people with 'underlying conditions' on the one hand, and social and business interest groups on the other hand, within this manufactured 'collective imaginary' (Bouchard, 2017, p 19). We visually exposed their often-conflicting impacts on shared media or policy narratives regarding public health, risk, and policy responses to the COVID-19 pandemic. The complex interplay of these themes creates an excellent case study for the visualisation of 'post-truth' media messaging (Fuller, 2020), expanding on techniques we have previously explored. The creative analysis, and resulting visually engaging presentation of the salient factors, effectively represent the complex interplay of narrative persuasion and policy response across the COVID-19 pandemic.

This study applied the term frequency and word vector techniques that are common in the 'big data' tools that underlie web searches and recommendation algorithms and guide many other aspects of 21st-century life. To make these underlying quantitative data analyses visible and intelligible to a non-mathematical viewer, this chapter uses a range of visual approaches to data representation to portray patterns and findings in an accessible manner. Our objective is to explore the potential within big data analysis to make observable the metadata underlying media narratives and messaging. Creativity is fostered by tools that are easy to use, converting statistical analysis into play and effort into image making. There is a tension between creating imagery with a high impact, in which the message is simplified and amplified into an aesthetic call for action, and the

1 systematic application of statistical techniques. Any resolution of the tension must
2 not misrepresent findings, and at the same time it must recognise the distorting
3 influence of aesthetic choices.

4
5 **Theoretical frame and study rationale**
6

7 We adopt critical realism as a theoretical lens through which to explore the
8 interplay of patterns that emerge within the data and their implications.
9 Critical realism, as a philosophical framework for social scientific research, uses
10 components of both the positivist and constructivist paradigms (Brown et al,
11 2002). Importantly, the framework of critical realism allows researchers to search
12 for causal factors, which help explain social events but also suggest practical
13 policy recommendations to address social problems (Yucel, 2018). The social
14 world is comprised of layers of phenomena, from the ‘real’, to the social and
15 the cultural (Fletcher, 2017). In essence, there is space for features of positivist
16 ‘causation’ alongside socially constructed shared understandings within cultural
17 spaces (Botha, 2021).

18 The domains of science and health have, across much of the 20th and 21st
19 centuries, been characterised by an emphasis on ‘evidence’ and ‘facts’. However,
20 more recently, with the ‘post-truth’ era of social media and information
21 saturation, there has been an increasing narrative regarding ‘alternative facts’
22 and divergent opinion (Gravino, 2022). The dissemination and interpretation of
23 scientific or medical knowledge is being increasingly challenged by those with
24 vested interests in distorting and denying the evidence-based claims of scientists
25 (Lockie, 2016; Lubchenco, 2017). There has simultaneously been an explosion
26 in conflicting and emotively toned public health messaging attempting to
27 manage public impressions or understandings regarding the pandemic. Evidence
28 from within the European Union (EU) has shown that lobbying by business
29 and professional organisations increased during the pandemic (Junk et al,
30 2021), presumably in an effort to influence short-term policy responses to
31 COVID-19. This is most pronounced among groups or industries which self-
32 identify as being highly impacted by the pandemic, such as the hospitality,
33 sports, and leisure industries, among many other examples. This lobbying
34 has *visually identifiable* impacts within the intersection of economic, public
35 health, political, and media messaging, which illustrate how media narratives
36 are constructed or influenced. The volume of media content focused on the
37 pandemic over a relatively short timeframe is unprecedented in scope and
38 accessibility. Therefore, it is useful to provide a visually clear and accessible
39 analysis of the features of media narratives relevant to COVID-19, shared public
40 understandings, and how vulnerable citizens – who are the stated subjects of
41 concern – have been represented.

42 While the pandemic has proved both complex and long-lasting, it has also
43 illustrated a previously undetected lack of public health education among citizens
44 in Ireland and internationally (Jarreau et al, 2021). Nearly half of all Europeans

(Kickbusch et al, 2013) and nine out of 10 adults in the US have inadequate health literacy skills and struggle to follow personal or public health information (CDC, 2021). This, unfortunately, has led to many consuming information regarding COVID-19 that is not accurate or trustworthy (Caballero et al, 2020). In other words, the audience for public health messaging during the pandemic are particularly vulnerable to consuming selective data, processed by well-resourced organisations and individuals. These sources may present analyses characterised by a narrow demographic representation, which have the potential to both encode and amplify specific messages under the guise of producing neutral data sets or universal ‘facts’. Such analyses hide impactful decisions in what data to present to the public and how to present it, while hiding such decisions behind a veneer of scientific objectivity (Bryan et al, 2019). In the context of quantitative analysis of media narratives, such degrees-of-freedom decisions dictate which narratives are given media focus, or which slices of research findings are referenced within government messaging.

Newspaper text data

Our data set comprises every newspaper article in five large-circulation national newspapers mentioning key phrases identifying vulnerable cohorts. For contrast, we include the two years preceding and the two and a half years following the first public health measures in Ireland in March 2020. We also contrasted news reports in Ireland with news reports in Australia and New Zealand, where the pandemic response was very different. Following preliminary searches and discussions to expand and refine our keywords, our text corpus consists of the full text of every article mentioning terms related to ‘mental health’ (defined by a search for mental with health, illness, disorder(s), or well-being), ‘underlying conditions’ (underlying or pre-existing condition(s), illness(es), or disease(s)), and ‘vulnerability’ (vulnerable or vulnerability related to health).

We obtained the full text of all articles and wrote short Python and R programs to examine and clean the data, ensuring consistent (e.g. UTF-8 Unicode) character encoding, punctuation codes, date formats, and article delimiters. The cleaning phase provided an opportunity to skim the news, absorbing themes, significant terms, and time trends. The end result of the cleaning process was five plain text files, one per newspaper title, containing the title, date, and article text ready for analysis.

We examined these articles across both time and geography. We categorised time in three epochs: a pre-pandemic phase from January 2018 to March 2020, a pandemic crisis phase from April 2020 to December 2021, and a post-crisis phase from January 2022. We categorised geography into Ireland, represented by the *Irish Examiner*, *Irish Independent*, and *Irish Times*, and Oceania, represented by the *New Zealand Herald* and *The Age* (Australia). In total, our corpus contained 29 million words in 34,249 news items.

Procedures: quantitative models of collective understandings

The wide availability and easy access to large volumes of text has stimulated the development of increasingly sophisticated and complex text analysis tools claiming to capture collective opinions, sentiments, and understandings. We chose to use code in Python and the statistical language R due to our familiarity with these, and their wide selections of well-documented open-source text analysis packages. Implementations of the same procedures are available in many computer languages and packages, and would reproduce the same outcomes. Quantitative word embeddings and their associated algorithms are usually black boxes with mysterious workings. Their outputs are taken on trust, providing the responses to online searches, selecting the posts at the top of social media streams, suggesting the next streaming TV show, locating housing, reviewing employee performance, and recommending prison sentences. ‘Making visible’ the word frequencies and word vectors (the frequency of every word that co-occurs with a word) in accessible form is a window into a large volume of text, and the knowledge or opinions a machine-learning system would claim to unearth.

Despite early warnings, the myth persists that analyses based on large data sets will transform science, bracket human biases or limitations in intelligence or knowledge, and generate insights with ‘the aura of truth, objectivity, and accuracy’ (boyd and Crawford, 2012 p 663). Indeed, analyses based on big data findings continue to be regarded as an oracle of the public zeitgeist. In both publicity puff pieces and academic papers, complex analyses based on machine-learning research are regularly presented as ‘value-neutral’ and possessing powerful potential for knowledge (Birhane et al, 2022). The reality, however, is that these analyses are based on social or information systems *designed by* humans. As such, the big data is likely also to reflect existing societal barriers, use the language of existing social categories or metrics, and perpetuate existing understandings or biases within their predictions or findings (Weidinger et al, 2021).

The current chapter harnesses these *weaknesses or challenges* related to the use of big data, reversing our analysis to use big data to *reveal* existing understandings, power structures, and influences within media reportage and ecology. Media text can be taken as a selective snapshot of society’s current opinion that is collective but not representative, due to the influence of dominant and powerful voices. This unrepresentative collective both reflects *and* forms opinion, with the potential to codify and magnify bias, exclusion, and harmful expressions.

Our approach to interrogating this data set is, firstly, exploratory and playful, with the intent of fostering a wide variety of imaginative readings of the data, and an unrestrained dialogue in imagery. We winnow the large volume of quantitative and visual findings into those we feel are objectively substantiated expressions of our opinions. We are very conscious that we have formalised subjective storytelling within an envelope of collective truths. This *truths envelope* contains diverse, contrary, mutually incompatible, and opposing views. The objective envelope expressed in media narratives is a collective envelope that shrinks to

narrower ranges of ‘objective truth’ with greater knowledge (when some opinions are widely recognised as invalid), and also as a consequence of compromise, compliance with more authoritative voices, and submission to (or silencing by) dominant voices.

The aim of our analysis is to expose the inner structures of word models, thus producing relatively easily understood visualisations. Within the images that are the outcome of our analyses we hope to explore the potential over-representation of powerful groups that could function to exclude vulnerable or less empowered people in society. This is prescient given the pandemic has been shown to have had a profound effect on vulnerable, disabled, and neurodivergent groups who may be less likely to be represented in media text. An additional and complementary aspect may be to uncover aspects indicating how these latter groups are rarely the creators or controllers of media narratives. Identifying these lacunae is an opportunity to consult these missing voices and bring to the foreground the lived experience of psychiatric care, health care, and the needs of vulnerable people during and following the COVID-19 pandemic. In Ireland, examples of such voices might be mental health patients, people with underlying conditions, Travellers, migrants in Direct Provision, prisoners, ethnic minorities, and neurodivergent and other vulnerable people.

Data analysis: making visible

One artistic influence on our approach is Philipp Steinwebber and Andreas Koller’s 2007 exhibition ‘Similar Diversity’ (Lima, 2011), a large-scale information graphic visualising commonalities and differences in the world’s five largest religions, derived from the text of their holy books. The clarity and surface-level simplicity of their infographics are windows into a detailed analysis of word frequencies and word embeddings. *Word embeddings* – the vector of co-occurring words defining the contexts in which a word is used – are said to represent meanings and understandings (Jurafsky and Martin, 2000), presented without interpretation by the artists as a supposedly objective visualisation of commonality and difference. Steinwebber and Koller’s results are reproducible, but, in ‘making visible’ rather than interpreting, they invite us not to trust the objectivity of the algorithm, but to perceive how it works (Arnall, 2014).

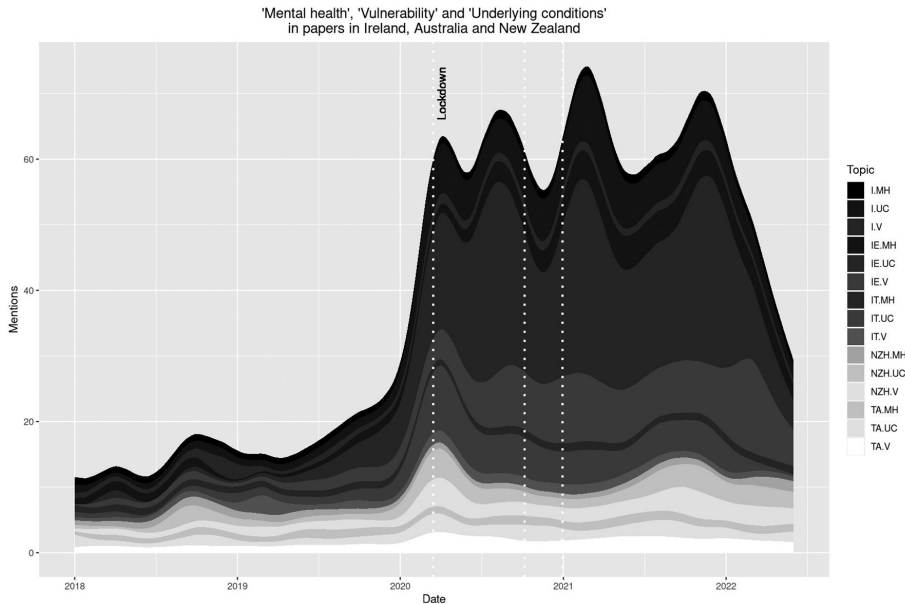
The frequency of words and phrases (including nouns and names) represent one measure of relative dominance and power in written media. The quantities tell us who writes, who is quoted, and whose interests are identified. Our aim is to visually represent these features of the patterns within the objective envelope of the media through data-driven representations using a range of techniques, which we discuss later in this chapter. We present a suite of case studies of how our data can be used creatively to present patterns in a manner more easily accessible or discernible to a general viewer. Each case study provides exemplars of how art can effectively use visual images to represent artefacts of meaning, which accessibly represent actual patterns within the data so that quantitative

variables influencing socially constructed narratives become overtly clear to the reader.

Case 1: Waxing and waning concern, and extreme variability in word frequencies

The word frequency of our three search terms varies over time periods and between geographic regions, as displayed in Figure 3.1. Our immediate observation was that Ireland displayed a substantial rise in interest or concern in our chosen terms during the pandemic crisis in the *Irish Examiner* (publishing 463 per cent of the pre-pandemic baseline number of items), *Irish Independent* (168 per cent), and *Irish Times* (157 per cent). The increased interest during the crisis was followed by a large fall in the post-crisis phase in the *Irish Examiner* (−70 per cent), *Irish Independent* (−36 per cent), and *Irish Times* (−49 per cent). In contrast, the rising and falling interest in Oceania was more restrained, with a pandemic crisis rise in the *New Zealand Herald* (publishing 134 per cent of the pre-pandemic number of items) and *The Age* (158 per cent), followed by smaller post-crisis changes in the *New Zealand Herald* (a continuing 7 per cent rise in concern) and *The Age* (−25 per cent). It is clear that Ireland exhibited a significant rise in concern regarding ‘mental health’, people with ‘underlying conditions’, and ‘vulnerable’ people, and

Figure 3.1: Frequency of newspaper mentions of search terms identifying ‘mental health’ (MH), ‘underlying conditions’ (UC), and ‘vulnerability’ (V)



Note: Newspapers are the *Irish Independent* (I), *Irish Examiner* (IE), *Irish Times* (IT), *New Zealand Herald* (NZH), and *The Age* (TA).

that this concern declined with the declining sense of pandemic crisis. Australia (as represented by *The Age*) exhibited a more restrained change in interest, and New Zealand (as represented by the *New Zealand Herald*) exhibited the most moderate change in concern, allied with the strictest COVID-19 response in any English-speaking country.

It is also visually apparent that rising concern coincided with the imposition of COVID-19 restrictions in Ireland, particularly the three national lockdowns limiting movement, gatherings, and access to employment and leisure.

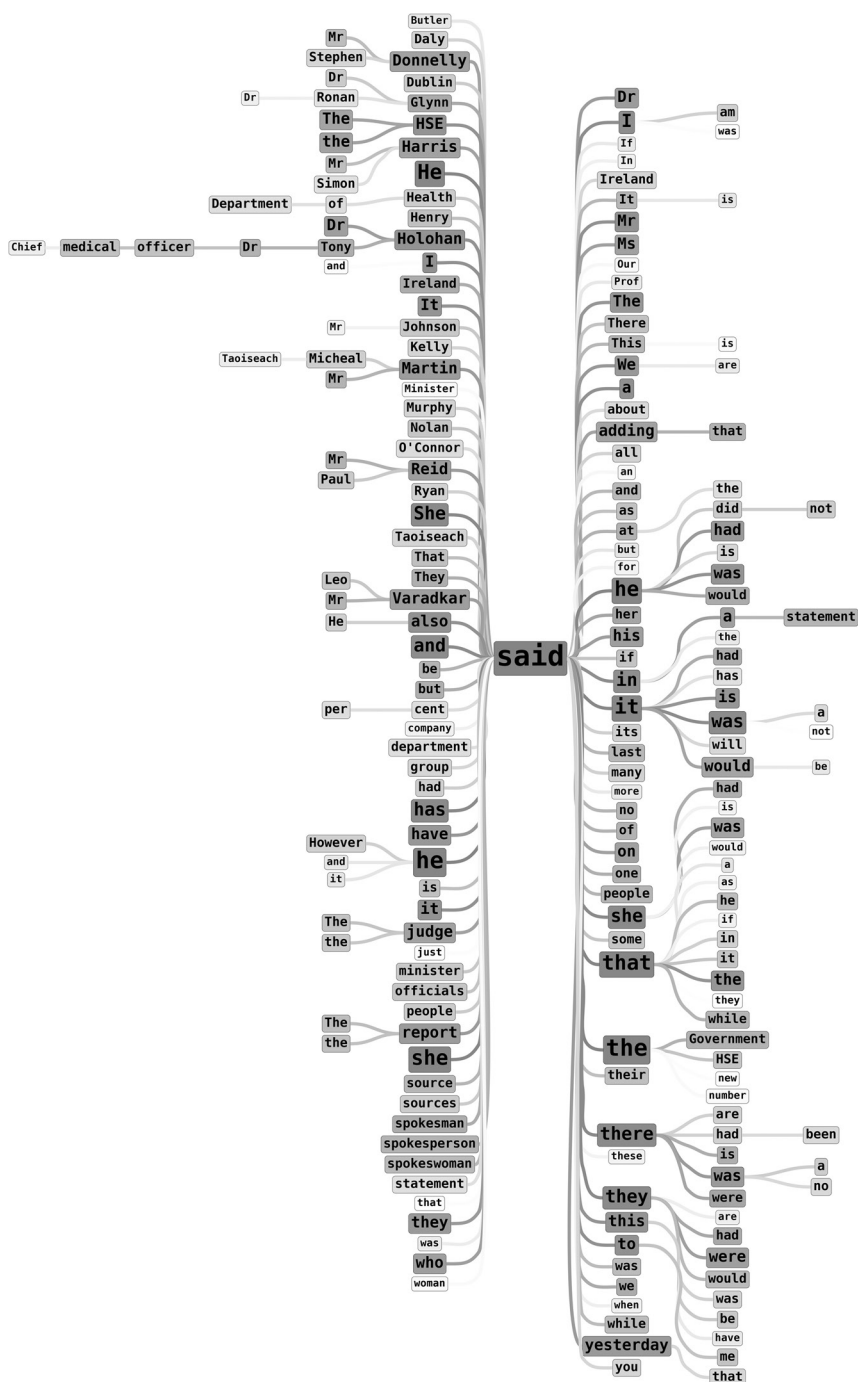
Case 2: Pandemic power structures

When we highlight specific *word embeddings*, we can highlight dominant themes (or interest groups) clustered around a theme of interest. Within our corpus of text we can identify who is permitted a platform to speak about, on behalf of, or from within our groups of interest. [Figure 3.2](#) is a visualisation of the word vector associated with the word ‘said’, displaying co-occurring words and their frequencies. These related terms are typically the identities of the voices represented in print media when discussing these topics. It is striking that of the most frequently identified people who ‘said’ words quoted in print media, only one of the top 50 is female, and not one is a Traveller, a migrant, a pupil with special educational needs, a patient with an underlying condition, or a representative advocating on behalf of these groups of people. Identified individuals, with their pandemic-era roles, include former Teachta Dála (TD) Jim Daly, Minister for Health Stephen Donnelly, Deputy Chief Medical Officer Dr Ronan Glynn, Minister for Health Simon Harris, Chief Medical Officer Tony Holohan, Boris Johnson PM, Alan Kelly TD, Tánaiste Micheál Martin, Paul Murphy TD, chair of the Irish Epidemiological Modelling Advisory Group Philip Nolan, Chief Operations Officer of the Health Service Executive (HSE) Anne O’Connor, Director-General of the HSE Paul Reid, neuroscientist Dr Tomás Ryan, Taoiseach Leo Varadkar, and a wide variety of politicians, officials, and spokespeople for related entities. The lived experience of vulnerable people is notable for its absence, falling far from the ideal representation of ‘nothing about us without us’ (Charlton, 1998).

Our visual analysis similarly exposed a notable change in the embedding of ‘mental health’ across time. Both pre- and post-crisis, mental *health* or *illness* co-occurred with psychiatric diagnoses, suicide, and support. During the pandemic crisis, mental *well-being* co-occurred with access to gymnasia, hospitality, and the imperative to return to the workplace.

We expose the word embeddings within weighted semantic networks, displaying the relative frequencies of each word node (in the font size) and the strength of the embedding (in the line thickness or colour density of connecting arcs). Word vectors visually represent the self-reinforcing frequency with which terms or individuals are associated with media discourses. It is easier to repeat a familiar concept than to synthesise a novel concept, and ease increases with

Figure 3.2: Most frequent 50 contexts (word vector) of the word 'said'



Note: The figure displays a dominance of male and official sources of statements, and a paucity of sources with lived experience of 'mental health', 'underlying conditions', and 'vulnerability'.

repetition. Like well-worn grooves in a stone surface, it is easy to repeat familiar word sequences, harder to say rare combinations, and impossible to speak a novel concept until a new formulation is created to express it.

Case 3: Subjectivity and narrative storytelling

In much of our work, we have also paired statistical word embeddings with more exploratory forms of communication using video-photographic composite artwork. We have exhibited these techniques in analyses of pedestrian flow in public space (Neilson, 2023), tracing bird flight above the city when humans were in pandemic lockdown, and visualising the body language of public personalities. We have noted a significant under-representation of women quoted in print media during the pandemic. There is an under-reporting of the lived experience of single mothers (who greatly outnumber single fathers) juggling blended learning, sick children, and working from home, women in front-line jobs in health care and education, pregnant women faced with frequently changing COVID-19 maternity guidance, and the widespread expectations on women as carers. We processed a set of video sequences to humanise these multifaceted and complex issues in a persuasive image. We selected a convenience sample of three men and three women identified in our newspaper text analysis as often-quoted speakers during the pandemic, and for whom we could locate television interview recordings. All six were interviewed under similar circumstances on the same Irish television programme, RTÉ News. The motion composite exposed the much greater volume occupied by men gesticulating while speaking. The calculated motion intensity visualises spatial volumes symbolic of the political capital accorded to and assumed by men and women in society and in our text corpus.

Discussion: influences and messages

This chapter has outlined how creative approaches to both the analysis and representation of big data drawn from word patterns within the media narrative may have utility in supporting readers to see the hidden interplay between influences and messages in the media. The COVID-19 pandemic provided a unique opportunity as large tracts of the media showed a persistent and overwhelming focus on a socially important single topic. Our analysis targets the role degrees-of-freedom decisions play in both the methodological rigour and ensuing findings within quantitative research, using the analysis of big data to explore how such patterns of interest can be made observable within the massive media coverage of the pandemic. We draw from a critical realism framework to support the recognition of both the existing of causal relationships across scientific and public health research. We use this framework to analyse the social construction of our understanding of COVID-19, the seriousness of pandemic outcomes, and what manner of policy response is most appropriate. We have

1 presented the outcomes of our approach across a series of case studies to indicate
2 the value of our approach.

3 Case Study 1 indicates clear cultural influences regarding how the pandemic
4 was represented within the media data, and clear impacts of government policy on
5 word frequencies within media messaging across countries. This is indicative of a
6 looping effect whereby media narratives impact government policy or responses,
7 with subsequent policy responses, such as lockdowns, also then clearly heavily
8 influencing media narratives in response. Such processes could serve to amplify
9 differences across jurisdictions or media ecosystems over time, leading to increasing
10 polarisation or discord within the populace in response to serious social phenomena,
11 such as the pandemic. Greater awareness among the public of how such feedback
12 loops can develop over time and create a self-sustaining dynamic of their own
13 would be important. Also important would be knowledge of how factors can
14 contribute towards the development of narratives within the media at the outset.

15 Case Study 2 explored *word embedding* as a visual representation of dominance
16 in media discourse. These embeddings also function as patterns facilitating word
17 choice and repetition, which are linked to particular roles or individuals. As
18 Geena Davis said ‘They can be what they see’ (Winick, 2013), and vulnerable
19 people can only see themselves as objects of concern, not authors of their own
20 concerns. It is notable, for example, that an initial predominance of ‘psychiatric’
21 pathologising terms were replaced in popular discourse by the progressively more
22 fluid (and wellness industry-friendly) phrases ‘mental health’ and ‘mental well-
23 being’, marginalising discussion of the scarcity of psychiatric care resources.
24 There may be a range of meanings behind such a choice of terminology or
25 association, particularly regarding communication impact and tractions across
26 as large a segment of the population as possible. Most lay understandings of
27 ‘psychiatric’ would see the term as denoting more serious psychological or
28 behavioural presentations, which require medical or institutional support in many
29 cases. In other words, they are experiences that impact a relatively smaller subset
30 of the population rather than larger segments of the citizenry. Given the serious
31 and pathological nature of conditions considered as falling under ‘psychiatric’
32 categories of psychological diagnosis, they often involve medical diagnosis and
33 specific conditions, presentations, or symptoms. In other words, support for
34 people with psychiatric conditions requires services, support staff, and budgetary
35 spending to put such responses in place.

36 By comparison, mental well-being, as a concept, has much broader applications
37 across the population, is much less specific or medical in association, and may
38 not require clinical involvement or consultation. As a concept, it is likely to have
39 much greater traction or relatability across a larger subset of the population. In
40 addition, mental well-being does not imply specialist health or psychological
41 services or interventions by the state are required, and can be discussed through
42 the lens of individual actions, choices, or experiences. Such individualistic foci
43 characterise the existing ‘well-being industry’, which aligns more closely with
44 leisure or lifestyle narratives than with medicalised ‘psychiatric’ domains.

1 The *effect* of the increased interest in or concern with ‘mental health’,
2 ‘underlying conditions’, and ‘vulnerability’ in Irish print media was evident in
3 the early reopening of gymnasiums, pubs, and other businesses, the continuous
4 access to alcohol with ‘essential retail’ throughout the pandemic, and ending both
5 lockdown and all visible restrictions on commerce. Concern has *not* translated
6 into increased funding for psychiatric services, or for social and health care of
7 vulnerable people. Indeed the health budget overall increased by just 4 per cent in
8 2022 (less than inflation) (Shine et al, 2022), waiting lists for medical procedures
9 have risen by 30 per cent to nearly 900,000 patients (Cullen, 2022), and there is
10 a pending ‘cancer timebomb’ (Doyle, 2021) of cases undiagnosed and untreated
11 during the pandemic (Cosgrave, 2021). In Ireland, capital spending on public
12 health and the number of staffed hospital beds remains the lowest in the EU
13 (IMO, 2022).

14 Misinformation and disinformation were widely shared on social media,
15 flooding into mainstream media as both inadvertent misinformation and as quite
16 consciously contentious topics. Marginalisation increases the attractiveness of
17 these alternative sources, further decreasing trust in official supports and access
18 to care, vaccination, and mitigations. All participants identified the media as a
19 catalyst for fuelling Traveller hostility and prejudice throughout the pandemic.
20 Negative representations increased stigma towards vulnerable groups including
21 Travellers (Friel, 2021) and prisoners (User Voice, 2022), who were consistently
22 excluded from media narratives about themselves.

23 Case Study 3 processed video to expose the difference in volumes occupied
24 by body motion of speakers representative of dominant and non-dominant
25 groups in high-profile media platforms. There was a notably higher frequency
26 of male pundits and experts invited onto media platforms and public discussion
27 programmes on the topic of pandemic-era health policy, leading to greater
28 familiarity and amplification of selected voices or perspectives. In contrast, when
29 women – or Travellers, migrants, or people with underlying conditions – gain
30 access to media platforms, the lower frequency of such opportunities, augmented
31 by their lack of familiarity to viewers and listeners, was associated with noticeable
32 differences in their presentation and behaviour while on air. Case Study 3
33 processed video into a visual image that contrasts the movement dynamics and
34 physical expressions of male and female panellists on TV discussion programmes.
35 The space the female panellists feel confident occupying is noticeably more
36 constrained than that of the male, despite their expertise, experience, or
37 knowledge being comparable. Frequent panellists develop a level of familiarity,
38 confidence, and experience that is visibly displayed in how they present their
39 views or messages – a profile that undoubtedly impacts how they are perceived
40 by those viewing their contributions. In contrast, less frequent panellists do
41 not have the opportunity to develop this bedrock of expertise, experience, and
42 communications competence, which is unintentionally on display in their non-
43 verbal communications while speaking. Our visualisation extends beyond the
44 quantifiable lack of representation of women, minorities, and vulnerable people

1 quoted in media text or broadcast time to a visceral portrayal of the difference
2 in embodied occupation of media space.
3

4 **Conclusions: missing voices**
5

6 Our approach to data analysis creatively visualises the quantitative relationships
7 within language, exploring the cyclic interaction of social media, lobbying,
8 expertise, and media reporting that itself operates like a large language model,
9 with all its attendant oracle-like powers and limitations. We make visible
10 the constant retrenchment of existing biases, amplified by negative and false
11 portrayals of marginalised and vulnerable groups, whose images are appropriated
12 by dominant voices, evident in the visualisation of newspaper reporting
13 about vulnerability.

14 The people who are the ostensible subjects of concern – mental health patients,
15 Travellers, migrants in Direct Provision, neurodivergent and disabled people, and
16 other vulnerable groups – are visibly missing voices in our text corpus. These are
17 also the groups most affected by the pandemic in measures of illness, mortality,
18 and financial hardship. Identifying these groups is not a remedy, but is the first
19 step on a longer road to expand and rebalance the objective envelope of media
20 representation, to include the first-person contributions of those missing voices,
21 with a lived experience claiming the central role of ‘Nothing about us without
22 us’ (Charlton, 1998).

23 Our analysis is limited by our ability to correctly identify relevant vulnerable
24 and powerful voices, and by our own participation in the culture we are analysing.
25 Even seemingly trivial aesthetic choices such as font or colour may carry –
26 perhaps subconscious – cultural meaning.

27 More broadly, these techniques can be applied to text or video collections that
28 are supposed to represent culture, in which the relative dominance of voices is
29 of interest. The visualisation of word embeddings is relevant to both using large
30 language models as a research tool and as a diagnostic aid to so-called artificial
31 intelligence (AI) generative outputs. There is a paucity of techniques to expose
32 how their output is generated. Visual representations of word embeddings offer
33 a view inside these black boxes, and a measure of their value as analytic tools.
34

35 **References**

36 Arnall, T. (2014) ‘Exploring “immaterials”: mediating design’s invisible materials’,
37 *International Journal of Design*, 8(2): 101–17.
38 Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., and Bao, M. (2022) ‘The
39 values encoded in machine learning research’. In *Proceedings of the 2022 ACM*
40 *Conference on Fairness, Accountability, and Transparency*, New York: Association
41 for Computing Machinery, pp 173–84.
42 Botha, M. (2021) ‘Academic, activist, or advocate? Angry, entangled, and
43 emerging: a critical reflection on autism knowledge production’, *Frontiers in*
44 *Psychology*, 12:727542.

- 1 Bouchard, G. (2017) *Social Myths and Collective Imaginaries*, Toronto: University
- 2 of Toronto Press.
- 3 boyd, D. and Crawford, K. (2012) ‘Critical questions for big data: provocations for
- 4 a cultural, technological, and scholarly phenomenon’, *Information, Communication*
- 5 *& Society*, 15(5): 662–79.
- 6 Brown, A., Fleetwood, S., and Roberts, J.M. (2002) ‘The marriage of critical realism
- 7 and Marxism: happy, unhappy or on the rocks?’, in A. Brown, S. Fleetwood and
- 8 J.M. Roberts (eds.), *Critical Realism and Marxism*, London: Routledge, pp 1–22.
- 9 Bryan, C.J., Yeager, D.S., and O’Brien, J.M. (2019) ‘Replicator degrees of freedom
- 10 allow publication of misleading failures to replicate’, *Proceedings of the National*
- 11 *Academy of Sciences*, 116(51): 25535–45.
- 12 Caballero, A., Leath, K., and Watson, J. (2020) ‘COVID-19 consumer health
- 13 information needs improvement to be readable and actionable by high-risk
- 14 populations’, *Frontiers in Communication*, 5, 56.
- 15 CDC (Centers for Disease Control and Prevention) (2021) ‘Understanding
- 16 health literacy’. Available from: [https://www.cdc.gov/healthliteracy/learn/](https://www.cdc.gov/healthliteracy/learn/Understanding.html)
- 17 [Understanding.html](https://www.cdc.gov/healthliteracy/learn/Understanding.html)
- 18 Charlton, J.I. (1998) *‘Nothing About Us Without Us’: Disability Oppression and*
- 19 *Empowerment*, Berkeley: University of California Press.
- 20 Cosgrave, T. (2021) ‘After Covexit, what then?’, *Irish Medical Times*, 28 May.
- 21 Cullen, P. (2022) ‘One-quarter of state’s population on health waiting lists’, *Irish*
- 22 *Times*, 6 June.
- 23 Doyle, P. (2021) ‘Pandemic has created “cancer timebomb”’, *Irish Medical Times*,
- 24 24 May.
- 25 Fletcher, A.J. (2017) ‘Applying critical realism in qualitative research:
- 26 methodology meets method’, *International Journal of Social Research Methodology*,
- 27 20(2): 181–94.
- 28 Friel, D. (2021) ‘Life on the margins: an exploration on the impact of COVID-
- 29 19 on Irish Travellers’ lives in the North West Ireland’, MA thesis, Institute of
- 30 Technology, Sligo.
- 31 Fuller, S. (2020) ‘A post-truth proactionary look at the pandemic’, *Postdigital*
- 32 *Science and Education*, 2(3): 551–5.
- 33 Gravino, P., Prevedello, G., Galletti, M., and Loreto, V. (2022) ‘The supply and
- 34 demand of news during COVID-19 and assessment of questionable sources
- 35 production’, *Nature Human Behaviour*, 6: 1069–78.
- 36 IMO (Irish Medical Organisation) (2022) *Pre-Budget Submission, 2023*,
- 37 Dublin: Irish Medical Organisation.
- 38 Jarreau, P.B., Su, L.Y.F., Chiang, E.C.L., Bennett, S.M., Zhang, J.S., Ferguson,
- 39 M., and Algarra, D. (2021) ‘COVID issue: visual narratives about COVID-19
- 40 improve message accessibility, self-efficacy, and health precautions’, *Frontiers in*
- 41 *Communication*, 6: 712658.
- 42 Junk, W.M., Crepaz, M., Hanegraaff, M., Berkhout, J., and Aizenberg, E. (2021)
- 43 ‘Changes in interest group access in times of crisis: no pain, no (lobby) gain’,
- 44 *Journal of European Public Policy*, 29(9): 1374–94.

Jurafsky, D. and Martin, J.H. (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Upper Saddle River: Prentice Hall.

Kickbusch, I., Pelikan, J.M., Apfel, F., and Tsouros, A. (2013) *Health Literacy: the solid facts*, Copenhagen: WHO Regional Office for Europe. Available from: <https://www.who.int/europe/publications/i/item/9789289000154>

Lima, M. (2011) *Visual Complexity: Mapping Patterns of Information*, New York: Princeton Architectural Press.

Lockie, S. (2016) ‘Post-truth politics and the social sciences’, *Environmental Sociology*, 3(1): 1–5.

Lubchenco, J. (2017) ‘Environmental science in a post-truth world’, *Frontiers in Ecology and the Environment*, 15(1): 3.

Neilson, S. (2023) ‘The power of an autistic lens: visualising activity in shared public space’, in M. Mostafa, R. Baumeister, M. Ramsgaard, and M. Tamke (eds), *Design for Inclusivity: Proceedings of the World Congress of Architects 2023*, Cham: Springer Nature Switzerland, pp 335–46.

Shine, C., Hennessy, M., and Walker, E. (2022) *Spending Review 2022 Health Capital Investment in Ireland*, Dublin: Department of Health.

User Voice and QUB (Queen’s University Belfast) (2022) *Coping with COVID in Prison: The Impact of the Prisoner Lockdown*, Belfast: Queen’s University Belfast.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.S., Chen, M., Glaese, M., Kasirzadeh, A., and Gabriel, I. (2021) ‘Ethical and social risks of harm from language models’, *arXiv preprint*, arXiv:2112.04359. Available from: <https://arxiv.org/pdf/2112.04359.pdf>

Winick, K. (2013) ‘Geena Davis solves sexism in Hollywood in five minutes flat’, *Elle*, 13 December.

Yucel, R. (2018) ‘Scientists’ ontological and epistemological views about science from the perspective of critical realism’, *Science & Education*, 27(5): 407–33.