To disclose or not to disclose: an extensive analysis into the product transparency on Dutch newspaper websites using archive data

Roeland Dubèl[1]

[1]Universiteit van Amsterdam

To disclose or not to disclose: an extensive analysis into the product transparency on Dutch newspaper websites using archive data

Traditionally, *objectivity* was seen as the golden standard of quality journalism (Singer, 2008). By leaving value judgments and opinions out of the equation, giving opposing parties an equal say, and presenting 'cold hard facts', the credibility of journalism was to be ensured (Skovsgaard et al., 2013). However, it is being questioned whether the standard of objectivity still meets (all) the public's expectations (Koliska & Chadha, 2018). Some scholars and professionals propose that this traditional journalistic norm of objectivity should be complemented, if not replaced, by a norm of *transparency* (Hellmueller et al., 2013; Phillips, 2010a). The focus should come to be on giving readers more information and insights into journalistic products and the production processes leading up to them (Uth et al., 2021).

One of the causes of this paradigm shift is the digitisation of the news media landscape (Karlsson et al., 2016). The introduction of online news has given news content a more fluid character: even after publication, online news articles can and do change (Timmerman & Bronselaer, 2022). Given that online news content often does not provide a fixed end product (Saltzis, 2012), it increases the pressure on news websites to transparently communicate what changes have been made and why (Karlsson, 2010).

On the other hand, relocating the news to the web has also created new opportunities to increase transparency (Karlsson, 2010): for instance, it is easier than before to disclose sources by the practice of hyperlinking (Mor & Reich, 2017), likewise, the addition of comment sections offers the possibility to openly receive and address criticism (Prochazka & Obermaier, 2022), rituals that find their roots in news-blogging (Hayes et al., 2007).

The focus on transparency is not only driven by the digitisation of news; it is also driven by a growing distrust in journalism (Craft & Heim, 2008). The perception among parts of the population that the news contains political bias, is sensationalised, has too low standards and/or provides factually incorrect information can be identified as possible causes for this growing distrust (Newman & Fletcher, 2017). However, by being open about how the news is created, the knowledge gap between sender and recipient can be reduced, empowering readers to judge for themselves whether the news has been created in a proper way (Craft & Heim, 2008).

This openness can be provided at different levels of the news production process (Uth et al., 2021): at the *organisational level*, for instance, by making internal journalistic guidelines public or by making editorial meetings open to all online, on the *individual level*, by publishing profiles of journalists about their personal and professional backgrounds or by interacting with readers on forums and lastly at the *content level*, this is the level where transparency practices take place within and around the article, it is the level where hyperlinking takes place and a comment section can exist.

Whether transparency at the organisational and individual level is genuinely effective in restoring trust is questionable. Namely, experimental research seems to indicate that transparency practices are most effective when they are displayed prominently (Koliska, 2022a), used consistently (Masullo et al., 2021), used in conjunction (i.e. various transparency practices are used simultaneously) (Curry & Stroud, 2021) and when they require little effort on the part of the recipient (Uth et al., 2021). Conditions that transparency practices at the organisational and individual level often do not meet: for example, internal guidelines are often not prominently featured, and interacting on forums with journalists also takes commitment from the reader. The focus of this study will therefore be at the level of the product, the news article itself. After all, it is in the form of the consumption of news articles that readers most prominently come into contact with journalism, it is hence where transparency should be present.

Therefore the aim of this study will be to map the extent to which transparency practices are consistently implemented in article content on online news websites. Although previous research has made similar attempts to map the use of transparency practices in the news, this study attempts to add to the current body of work, in various ways.

First of all, previous work has mainly used manual methods such as (exploratory) content analysis or interviewing, resulting in relatively small samples (see for example Chadha & Koliska, 2014; Karlsson, 2010; Kashyap et al., 2022; Robles et al., 2023; Tandoc & Oh, 2015). In turn, attempts to analyse a larger corpus of data were often confined to studying one specific transparency element, mostly the use or non-use of hyperlinking (e.g. Humprecht & Esser, 2016; Karlsson et al., 2014). In contrast, this study will serve as a proof of concept to explore to what extent it is possible, through computational methods, thus the use of algorithmic solutions, to scale up the sample size without limiting itself to one specific transparency practice.

Although reaching a higher *N* is not an end in itself, scalability, however, offers more ease and opportunities to solve theoretical puzzles in relation to journalistic transparency. For instance, some transparency practices are quite uncommon, such as utilising polls in news articles, which allows readers to express themselves (Karlsson, 2010). By scaling up, these rare transparency practices can be more easily identified, and more statistical power can be obtained to explain exactly when these unique transparency practices occur.

Moreover, two specific strands of research on journalistic transparency may benefit from scalability: longitudinal research and cross-national research. For example, research is being done on how the implementation of transparency develops over time in journalism (see for instance Karlsson et al., 2015, on the trends of hyperlinking). Similarly, following the work of Hallin and Mancini (2015) on media systems, Humprecht and Esser (2016) made an attempt to cluster 48 news media websites from six different countries using a set of dimensions of which transparency was one of them. It is research like these studies, where theoretical analyses take place at the macro level, and thus involve a lot of data, that would benefit from automatic techniques to detect transparency.

To demonstrate these computational capabilities, two specific transparency practices will be discussed in detail. Firstly, the use of 'last-edited' timestamps alongside publication timestamps, to be transparent about the volatility of the news article (Karlsson, 2010). On the one hand, previous research on the prevalence of transparency has only looked at the presence or absence of such timestamps but has not looked at what actually changed (e.g. Karlsson, 2010). On the other hand, research that does monitor the changes a live article undergoes then again does not link this to whether or not 'last-edited' timestamps are used (Saltzis, 2012). By using snapshots from the *Wayback Machine's* digital archive, this study will explicitly make that connection and map whether changes are always disclosed with a timestamp, and if not, which factors predict the use of a last-edited timestamp.

Secondly, the use or non-use of anonymous sources will be examined as a measure of how transparent the source usage is (Robles et al., 2023). Instead, much research tends to focus exclusively on digital-powered transparency practices (such as hyperlinking and last-edited timestamps, practices unthinkable without the web). However, as stated earlier, the demand for transparency is fuelled not only by digitisation but also by growing distrust. Therefore, in principle, it should not matter whether transparency is achieved through a technological method or a more traditional method, such as minimising anonymous sources. By using large language models and prompt design (writing commands to chatbots) (Mitrović et al., 2023), an attempt will then be made to scale up manual annotation of source disclosure at the paragraph level.

Finally, previous research mostly just made external comparisons between news websites when it comes to the implementation of transparency, mostly on the basis of public versus private (Humprecht & Esser, 2016). However, since consistency is an important condition for transparency to effectively influence trust, it is equally important, if not more so, to also look at internal differences, such as between sections, in the use of transparency. Moreover, survey research shows that there is a perception among readers that transparency is sometimes present in articles and sometimes not (Uth et al., 2021). This study will therefore also identify which internal predictors explain transparency at the article level.

All in all, this research can further contribute to mapping the use of transparency in news websites, and therefore possibly expose shortcomings.

Theoretical background

From norms to routines

Like in every other market, the news industry is driven by a tango of supply and demand. On the consumer side, the public can have certain expectations about how journalism should operate (Loosen et al., 2020) and thus how they ought the end product, the news content itself, to be. Similarly, on the producer side, journalists have their "own formulation of how they ought to do their work" (Mellado & Van Dalen, 2014, p. 861). These formulations are often referred to as *role perceptions*. Although some continuity exists, the expectations of the public and the role perceptions of journalists are not stable across time (Loosen et al., 2020).

As for the consumers, there is an increasing demand for transparency in journalism (Loosen et al., 2020). In generic terms, transparency refers to providing availability of information. In the context of journalism, it refers to giving insights into how the news actually comes about: from disclosing the motives behind the story, the research process, the primary sources used, the author of the story and whether or not the author had competing interests in writing it (Craft & Heim, 2008; Ward, 2014). For the news consumers transparency can be a way to take away the uncertainties they have about the factuality, independence and quality of journalistic writing.

Namely journalistic transparency can heighten *checkability*: by being open about the journalistic process, it is easier for the public to assess and monitor the trustworthiness of journalistic work (Craft & Heim, 2008; Fengler & Speck, 2019; Vos & Craft, 2017). Moreover, transparency can also induce *empowerment* and *autonomy*: by opening up the workflow behind a story the public is given the tools to reproduce the journalistic research process and thus verify the information from the news, heightening the public's autonomy. This more informed, empowered and autonomous public can consequently lead to more *correctiveness*: through providing the public with the same information as the producers of the news, the public is enabled to criticise the work of the news industry more thoroughly and thereby serve as a corrective mechanism (Phillips, 2010b). This new role of the public can foster *dialogue* between producer and consumer, in which the public no longer solely consumes the news but also starts to actively contribute to it (Craft & Heim, 2008).

The increasing call for transparency is reflected in the growing perception among journalists that being transparent is part of how they should do their work (Loosen et al., 2020). For the producers of news, the journalists, transparency can serve as a substitute for the traditional norm of objectivity, which is often regarded as being unachievable. Instead of striving for objectivity, one would recognize that one can never wholly distance oneself from one's own perspectives, giving way to the recognition of and openness about one's subjectivity in journalistic writing (Koliska, 2022b; Vos & Craft, 2017; Ward, 2014). Moreover, by providing openness and bringing about the complexity of truth-telling, transparency can foster understanding and trust among the public (Craft & Heim, 2008; Fengler & Speck, 2019; Phillips, 2010b).

Even though the notion of transparency is increasingly receiving resonance within journalism, a difference can be made between considering it to be important and the actual implementation of it. The latter can be referred to as *role performance*: "the collective outcome of concrete news decisions that put in practice certain roles — what journalists actually do and how that materializes in professional practice — which are under constant and dynamic negotiation" (p. 254, Mellado et al., 2018). The concrete outcomes of the notion of transparency could materialize in the form of (media) *routines* (Karlsson, 2010). Routines are "patterns of behaviours that form the immediate structures of media work" (p. 399, Reese & Shoemaker, 2018). Whether or not the notion of transparency transcends to visible patterns of transparency does not solely depend on the routines of the journalist.

First and foremost, as an extra-media technological force, the arrival of the Web which consequently led to the digitization of the news media landscape has been a key influential development for the notion of transparency. The digital evolution of the news has both been a blessing and a burden for the notion of transparency. To begin with, compared to traditional channels such as print and television, the web is not limited by respective pages and airtime: digital space is practically endless (Vos & Craft, 2017). Therefore the argument that there is no room for disclosure in journalism has lost its validity: the digitisation has made room for lengthy disclosures of author backgrounds, previous errors and process information. Moreover, digitisation has also expanded the toolkit to implement transparency (Karlsson & Clerwall, 2018): linking to sources by hyperlinking or providing downloadable content and openly receiving criticism through comment sections, are a direct consequence of the newly found technological possibilities.

Paradoxically, the digitization has also been a burden for transparency. Digitization has paved the way for an instant publishing model: due to the low distribution costs, as compared to for example printing costs, news can be published rather instantaneously (Saltzis, 2012). This renewed focus on immediacy has increased the presence of errors, and therefore the need to disclose those too. Moreover, the instant publishing model has led to the routine of publishing drafts right away (most certainly for 'breaking' news), making it that often multiple updates are needed to come to a final product (Timmerman et al., 2023). This higher volatility of the news requires more disclosure towards the public in regard to possible additional, nuanced or contrasting information in relation to previous versions of the article.

Furthermore on the organizational level journalism is facing severe financial challenges, in the form of greatly shrunken revenues (Olsen et al., 2021). Consequently, news staff have been reduced causing increased time pressures on individuals within newsrooms. Journalists therefore often argue that they do not have time to implement transparency (Koliska & Chadha, 2018). Managers therefore often direct journalists to stick with low-effort computer-mediated transparency features (Chadha & Koliska, 2014; Gade et al., 2018): features such as the use of detailed timestamps to heuristically disclose changes and (internal) hyperlinking.

Transparency rituals

In the following section the transparency routines, or as Karlsson, 2010 labels them *transparency rituals*, will be discussed in further detail. It must be noted that the rituals discussed are not exhaustive. The routines that will be elaborated upon are chosen on the basis that they occur 1) on the content level given that transparency rituals are most influential when they occur at this level (Uth et al., 2021). Therefore transparency practices that appear on the organizational level, such as publishing ethical guidelines and opening up editorial meetings, are excluded. Furthermore the mentioned routines 2) must be executed internally: they are carried out by journalists or the news organizations, as the goal of the study is to assess to what extent newspaper websites themselves apply transparency practices. External transparency forms, such as fact-checking, which often is performed by third parties, are therefore excluded from further theoretical and statistical analysis. Lastly given that this paper acts as a proof-of-concept to what extent the implementation of transparency practices can be automatically detected using computational techniques, highly unstandardized transparency practices are not taken into account. For instance, user-generated news content tends to not have a standardized appearance on newspaper websites and is therefore troublesome to automatically detect.

The routines will be discussed in order of appearance, working down from the top of the article to the bottom of the article. Additionally, the discussed transparency rituals are clustered among the categories *actor transparency*, *error transparency*, *source transparency*, *disclosure transparency* and *participatory transparency*. The overarching research question of this paper would then be the following:

RQ1 To what extent do these transparency rituals occur in news article content?

Actor transparency

One of the first elements of transparency a reader could encounter is the disclosure of information about journalists and/or the news organizations themselves. This can be referred to as actor or producer transparency. An elementary way to provide actor transparency is by including the *author name* in the byline. The routinization of disclosing the name of the author(s) of news articles followed the institutionalization of the byline (Koliska & Chadha, 2018). As such it can be regarded as one of the oldest forms of transparency as it gave openness about who wrote the story and therefore empowered the public to assess the independence and competence of the author(s).

Although disclosing one's name should be a precondition before being able to disclose any additional information, it is something that is rarely addressed in studying the prevalence of transparency practices. Since it is often overlooked, little is known about the extent the author(s) get disclosed.

In the ideal scenario, an explicit name would be mentioned (fore- and surname). Only then it would fully empower the reader to assess the credibility of the writer. Logically the total absence of an author does not provide that opportunity, but also abstract in-between forms of author disclosure can hinder transparency. The use of public affairs (PA) and relations (PR) sources, a source type often used (Lewis et al., 2008), remains limited to stating the names of the organizations, but not specific individuals. Similarly, the disclosure of the authors of articles by generic terms (e.g. 'by the economy department') provides little empowerment to the reader.

Additionally *author contact information* can be made available by supplementing the byline with either the email address or socials of the author(s). Disclosing contact information opens up the possibility for dialogue between consumer and producer (Chadha & Koliska, 2014). Previous research found this to be common practice (Chadha & Koliska, 2014; Karlsson, 2010).

What is also common practice is to disclose *author biography information* about the author(s) (Chadha & Koliska, 2014). This practice is often limited to providing superficial information such as purely linking to previous stories, education and/or employment of the author. Nonetheless, this simple form of disclosure can be a tool to communicate the competence of authors(s) (Uth et al., 2021).

Error transparency

The second element of transparency relates to the disclosure of errors or changes in the article. We will simply refer to this as ‘error transparency’. Previous research found that most online news stories receive a couple of updates after publication (Saltzis, 2012). These updates tend to occur within the first few hours after publication (Timmerman et al., 2023). Examples of these article changes are correcting typos, adding background information, clarifications and responses of relevant actors as well as removing faulty information (Kutz & Herring, 2005).

These changes can be disclosed in two distinct ways: through the use of *detailed timestamps* and *error corrections*. In the case of detailed timestamps timestamps can be provided of both the initial publication date as well as when the story was last updated. Although the use of detailed timestamps is quite prevalent it is often denoted as being heuristic (Chadha & Koliska, 2014): it signals transparency but does not provide actual openness on why and how the article has changed. In the case of error corrections the why and how do get disclosed. Using error corrections is therefore more of a profound way of taking responsibility by clearly disclosing the correction to the public (Karlsson, 2010). This form of error disclosure is however less common (Chadha & Koliska, 2014; Karlsson, 2010).

Although there is (micro-) longitudinal research on the changes in news content as well as content analysis on the use of error disclosure, these two strands of research often remain disconnected. It therefore remains unclear whether changes are always disclosed, in the case of detailed timestamps perhaps automatically, or if manual decisions are made when to disclose changes.

Arguments can be given for the latter option. Namely disclosing every error, e.g. even typos, might not serve the purpose of transparency and backfire by causing scepticism about the quality of the news (Fengler et al., 2015; Vos & Craft, 2017). This resonates with the general perception journalists have of transparency: something that should not be applied to every case, but mostly to major cases (Koliska & Chadha, 2018). It is therefore expected that error discloses will mostly be present when the errors are large, prominent and omissive (which is indicative of error removal):

H1 Articles which include 'last-edited timestamps' and/or 'error corrections' have had a) lengthier, b) more prominent, c) and omissive changes as compared to articles that do not include these timestamps and/or corrections.

Source transparency

Within the body of the text, transparency can be given on the sources of the story. This can be termed *source transparency* (Fengler & Speck, 2019). Transparency in sources can be reached by the practice of *hyperlinking*, adding *original documents* and/or *sourcing* in text.

Hyperlinking refers to embedding clickable links in text (Karlsson et al., 2014) and makes it possible for the public to track down sources (Karlsson & Clerwall, 2018). It therefore also provides a way to make journalists accountable by directly giving access to the source, through links, on which they built (Chadha & Koliska, 2014; Karlsson et al., 2014; Vos & Craft, 2017). A differentiation can be made between *internal hyperlinking* and *external hyperlinking.* Internal hyperlinking refers to the practice of linking to the same domain, whereas external hyperlinking refers to linking to sources outside the realm of the publisher (Karlsson, 2010).

Internal hyperlinking is regarded as a form of heuristic transparency (Chadha & Koliska, 2014); it takes little effort (internal stories tend to be tagged and can thus easily be added) and does not always truly provide transparency of sources. For instance, if a news article instead of naming the original source, hyperlinks internally to an earlier article in which the original source is named, it is making it more difficult for the reader to retrace the sources, actually reducing transparency. Technically external hyperlinking can cause a similar fog. One of the most frequent external linking patterns is linking to other newspapers (Humprecht & Esser, 2016; Vos & Craft, 2017), which can create a similar pattern in which the original source is not named but the indirect source, the 'competitor', is linked instead.

The use of hyperlinking is a relatively mainstream transparency practice (Chadha & Koliska, 2014; Karlsson, 2010; Karlsson et al., 2015). More specifically it is internal hyperlinking that is mostly common practice, whereas external hyperlinking is less common (Humprecht & Esser, 2016; Karlsson et al., 2015). An argument to explain this gap is the commercial incentive of internal hyperlinking: whereas external hyperlinking lures readers away from the publisher site, internal hyperlinking extends the presence of the reader, causing more views and thus advertisement revenue.

Similarly to hyperlinking one could also directly embed original documents in news content to increase openness and credibility (Karlsson, 2010). The practice of doing so is however less common as compared to hyperlinking (Chadha & Koliska, 2014; Karlsson, 2010).

Lastly, source transparency could also be achieved by sourcing: disclosing the (direct) sources used within the article, as well as why sources are or are not disclosed (Chadha & Koliska, 2014). The act of having identifiable sources is a core element of being open towards the public (Kovach & Rosenstiel, 2001). The use of anonymous sources however tends to be prevalent (Chadha & Koliska, 2014). Journalists themselves even argue that the request to stay anonymous is sometimes too easily accepted. They state that convincing the source to be public, is a rather time-consuming task (Kovach & Rosenstiel, 2001).

Disclosure transparency

Openness can also be given regarding the production process and internal standards of news (Karlsson, 2010). This form of transparency can be called disclosure or simply production transparency. By doing so a journalist makes oneself accountable (Craft & Heim, 2008) and can create understanding among the public (Karlsson & Clerwall, 2018). Disclosing process information is however rarely happening (Chadha & Koliska, 2014; Kovach & Rosenstiel, 2001). The argument presented by journalists to not engage in this behaviour is the fear that it will create confusion among the public as the journalistic process would be too difficult to explain due to its messy nature (Chadha & Koliska, 2014).

Participatory transparency

Lastly and often at the bottom of the news article, transparency can be given by aiming to include the audience in news production (Karlsson, 2010). Participatory transparency as it is called can among others be implemented in the form of *polls* and *comment sections*. Both polls and comment sections can enable the audience to express their thoughts, and can therefore be a starting point of dialogue between consumer and producer. Although the implementation of polls is rare, comment sections tend to be quite prevalent (Karlsson, 2010). Whether or not comment sections are truly transparent is doubtful as these sections tend to be highly moderated (Uth et al., 2021)

External and internal differences

The implementation of transparency rituals can differ greatly between newspaper types (Karlsson, 2010). Three dimensions are key in explaining these diverging patterns: web-only as opposed to multi-modal, quality against popularity and public versus private.

Firstly, previous research found that web-only, digital-native, news media were more active in disclosing their methods and linking to sources, as compared to multi-modal, legacy, news media (Humprecht & Esser, 2016; Kashyap et al., 2022; Robles et al., 2023). A possible explanation for this opposing pattern can be derived from *neo-institutional* *theory* (Karlsson & Clerwall, 2018). One of the tenets of neo-institutional theory is that when institutions are faced with a period of change, these institutions tend to fall back on traditions. The digitisation of the news-media landscape can be seen as such a period of change, causing legacy media to fall back on their traditional reporting techniques. Given that digital native media lack those traditions, they are more open to creating new ones (Karlsson & Clerwall, 2018). Therefore, web-only news media tend to bring in new work approaches, making use of new technological capabilities and can be seen as an alternative to legacy media (Humprecht & Esser, 2016; Kashyap et al., 2022). Since the use of transparency is largely induced by technological capabilities it is therefore to be expected that they make more use of transparency practices (Robles et al., 2023).

Secondly, transparency practices, in the form of hyperlinking, tend to be more prevalent in the content of public as opposed to private news media (Humprecht & Esser, 2016). According to *practice theory* "strips of activity [i.e. practices] serve as meso-level resources that mediate the impact of macro-level forces" (p. 61, Ryfe, 2021). These macro-level forces in the form of economic considerations differ for public media, which are largely subsidized, compared to private media, which are profit-driven. Consequently, it can be expected that for public media the benefit of transparency practices, namely trust repair, is deemed to be more important than the costs of implementing it, making it more likely that across the board public media apply more transparency practices as compared to private media (Humprecht & Esser, 2016).

Thirdly it has also been found that the transparent practice of hyperlinking, both internally and externally, is consistently (over time) more present among broadsheet (quality) newspapers as compared to tabloid (popular) newspapers (Karlsson et al., 2015). This could be explained by the different business goals and target audiences of tabloid and broadsheet newspapers. Whereas the aim of tabloid newspapers is to reach a mass-market audience, through the use of entertaining and sensationalist reporting, the aim of broadsheet newspapers is to reach an up-market, higher educated, audience through informative and rationalistic reporting (Doyle, 2013). The more entertaining, sensational and populist drive of tabloid newspapers often comes at the cost of journalistic standards (Esser, 1999), making it more likely that it is quality outlets instead of popular outlets to adhere to the call for transparency and thus to implement it.

Taken together it is expected that transparency rituals are mostly present in the article content of digital native, public and broadsheet newspapers:

H2 Transparency rituals are more prevalent among the news article content of a) digital native, b) public and c) quality news media, as compared to legacy, private and popular media.

Previous comparative research on transparency practices mainly focused on external differences, thus between outlets, instead of internal differences, within outlets. Nonetheless, the implementation of transparency practices is poorly coordinated from the organizational level to the journalist level, even causing inconsistent use of it within outlets (Gade et al., 2018). This resonates with the perception of consumers that transparency varies from item to item (Uth et al., 2021).

A possible dimension to explain the inconsistent implementation of transparency practices is that of hard versus soft news. The difference between these two news types can be defined across several dimensions, one of those being their presentation format (Reinemann et al., 2012). Whereas hard news tends to provide stories in a more informative and factual manner, the focus of soft news tends to be more personalized and entertaining (Patterson, 2000). Moreover, hard news tends to be mostly prevalent in sections with high societal relevance, such as politics, economy, science and technology, whereas the soft-style format, is mostly occurring in sections such as sports, entertainment and regional (Curran et al., 2010).

In principle an argument could thus be made that a distinction between quality and popular news can not only be made on the newspaper level, but also on the content level: there are some sections that, even in broadsheet outlets, tend to be more entertaining and mass-oriented than others, at the cost of journalistic standards. As such it is similarly expected that the hard-news-related sections pertain to higher news standards as compared to the soft-news-related sections, and thus implement more transparency practices:

H3 Transparency rituals are more prevalent in hard-news-related sections as compared to soft-news-related sections.

Methodology

Data collection

As for data collection, a diverse set of outlets will be scraped which include digital native (NU.nl), public (NOS.nl) and both broadsheet (De Volkskrant, Het NRC) and tabloid news websites (De Telegraaf, Het AD). These outlets will be scraped until the beginning of the calendar year. Since corporations typically change policies, particularly, per calendar year, it is expected that the implementation of transparency is fairly standardised over the calendar year. If there is a large variation in the number of cases for each outlet then the number of cases per outlet will be randomly sampled down to the size of the smallest output.

In the second step, the initial version of the news stories will be collected using the *Wayback Machine* API. Items that have been online for less than 24 hours will be removed from the data set as news content updates usually occur up to a maximum of 24 hours after publication.

Measurements

Author name The author name can either be absent, abstract (PA, PR or generic departments) or explicit in the form of a first- and surname. The presence or absence of an explicit author will be detected by using a regular expression (a method in which patterns of text can be found) on the byline. Namely, if an explicit name is mentioned one would expect at least two capitalized words after each other (or a multitude of them). The performance of this method will be manually verified.

Author contact information Contact information can either be 1) not disclosed, 2) disclosed via email, 3) disclosed via social media handles or 4) via both email and social media handles. By applying regular expressions to the underlying HTML structure of the website it can be detected if there is being linked to email addresses and/or social media handles.

Author bio information Author bio information can either be present or not. The presence of bio information can be detected by assessing whether a hyperlink exists in the article byline that links to a sub-page in which the name of the author is present in the URL or on the sub-page itself.

Detailed timestamps The use of detailed timestamps can be divided into 1) neither a time of publication nor update are provided, 2) only the time of publication is provided, 3) only the time of update is provided, both the time of publication and update are provided (Karlsson, 2010). The presence (or absence) of timestamps can be easily extracted by locating where such timestamps would be stored in the underlying HTML structure.

Article changes An article has either changed or not. The manifestation of article changes will be measured in the form of the size of change, prominence of change and direction of change. To assess the ***size of change*** the *Levenshtein distance* between the initial version and the later version will be calculated. The Levenshtein distance is a metric that calculates how many changes need to be made to come to the same text (Zhang et al., 2017). The ***direction of change*** will be measured in a similar fashion as by Timmerman et al., 2023 by subtracting the length of the updated article from the length of the initial article and dividing it by the length of the initial article. In this way, the measure highlights the factor of added (a positive factor) or removed (a negative factor) information relative to the initial size of the article. In addition to asses the ***prominence of change***, the Levenshtein distance will be calculated between the first paragraph of the initial publication and all paragraphs of the updated version. It will then be assessed with which paragraph the initial first paragraph is most similar in the updated version. This new position will be divided by the total amount of paragraphs in the updated version. Thus the higher the score the more the initial first paragraph moved down in the article, indicating that information was mainly changed at the top of the article.

Error corrections Error corrections are either present or not. Detecting error corrections in text is likely to be difficult as it is a non-heuristic unstandardized method. Therefore after data collection changed articles will be randomly inspected to assess if and how error corrections in text are disclosed. The goal of this endeavour would be to find linguistic patterns in error corrections; e.g. "an earlier version stated ...". On the basis of this deductive approach, a regular expression will be made to detect error corrections. To verify the workings of this method the detected cases (if present) will be manually verified for a maximum of 30 randomly selected cases (as is recommended by Lombard et al., 2010). Till at least, 70 % of cases are correct (i.e. a precision of .7, which is the recommended acceptance level for more exploratory tasks, see Lombard et al., 2010). If this condition is not met the code will be adapted and if not successful it will be excluded from further analyses.

Hyperlinking The amount of internal and external links will be counted, as being distinct numeric features. The presence of hyperlinks can easily be extracted by inspecting the underlying HTML structure of the article, in which hyperlinks are often tagged as an 'href' element. Consequently, for the detected links it can be assessed whether the name of the original publisher is present in the link, which would make it internal, or not, making it external.

Original documents The amount of original documents will be counted as a numeric feature. This will be detected by looking for links to document files (.doc, .pdf, et cetera) in the HTML structure of the body of the text.

Sourcing Sourcing in the text will be divided into three categories: the absence of a source, the use of an anonymous source without reasoning, the use of an anonymous source with reasoning, and the use of an explicit source (i.e. a specific organization or person). As for assessing sourcing, this will be limited to the first paragraph. Both theoretical and practical arguments can be given. Theoretically, it is expected that the most important source, the primary source, is mentioned at the beginning of the article. Secondly, linguistically, it can be that the source is explicitly mentioned in the opening paragraph, whereas in later sections the source can be indirectly referred to. For example, if in the first paragraph, the United Nations is given as the source, it could be that later on in the article they simply refer to the UN as the “organizations stated ...”. In principle, the latter paragraph would be coded as anonymous sourcing since the organization is not explicitly named, but in reality, it is named, but simply not in the same paragraph. It will thus be limited to the first paragraph. The procedure will be to first attempt to do this via a large language model prompt, till a precision of.7 is reached on 90 manually coded cases. If not feasible we will resort to supervised machine learning and 1000 cases per class will be coded (thus 3000 in total) as is custom in this procedure (Van Atteveldt et al., 2022).

Process information Disclosing process information will be stored as a dummy variable, being either present or not. Trying to detect the presence of process information is a rather difficult task. Nonetheless, an attempt will be made by first looking at the usage of first-person pronouns outside of quotations (I, me, we, us, our, my, mine). As it would be outside quotations it would indicate that the journalist is reflecting on his or her own work. Additionally, there will also be a look if an explicit name in the byline is reiterated in the text to also capture third-person commentary: “Journalist X has chosen this method because of Y". To verify the workings of this method the detected cases (if present) will be manually verified for a maximum of 30 randomly selected cases till at least 70% of cases are correct.

Poll This will be a dummy variable with either having a poll or not. Partly due to the rarity of poll usage, it can be somewhat unpredictable how a poll would be embedded in the HTML structure. An attempt will be made by trying to detect whether a class is present within the HTML structure that contains the term 'poll'. Furthermore, it is also to be expected that within this element a question must be asked and that thus a question mark must be present in the text within this poll component. To verify the workings of this method the detected cases (if present) will be manually verified for a maximum of 30 randomly selected cases till at least 70% of cases are correct.

Comment section This will be a dummy variable with either having a comment section present or not. An attempt will be made to look at wether a paragraph title is present under the body of the text (often denoted with an 'h' tag in the HTML structure) in which the term comment(s) or "reactie(s)" (Dutch for responses) is present. This would serve as an indication of the presence of a comment section and will be manually verified.

References

Chadha, K., & Koliska, M. (2014). Newsrooms and transparency in the digital age. *Journalism Practice*, *9*(2), 215–229. https://doi.org/10.1080/17512786.2014.924737

Craft, S., & Heim, K. (2008). Transparency in journalism: Meanings, merits, and risks. In *The handbook of mass media ethics* (pp. 231–242). Routledge.

Curran, J., Salovaara-Moring, I., Coen, S., & Iyengar, S. (2010). Crime, foreigners and hard news: A cross-national comparison of reporting and public perception. *Journalism*, *11*(1), 3–19.

Curry, A. L., & Stroud, N. J. (2021). The effects of journalistic transparency on credibility assessments and engagement intentions. *Journalism*, *22*(4), 901–918.

Doyle, G. (2013). Understanding media economics. *Understanding Media Economics*, 1–232.

Esser, F. (1999). Tabloidization'of news: A comparative analysis of anglo-american and german press journalism. *European journal of communication*, *14*(3), 291–324.

Fengler, S., Eberwein, T., Alsius, S., Baisnée, O., Bichler, K., Dobek-Ostrowska, B., Evers, H., Glowacki, M., Groenhart, H., Harro-Loit, H., et al. (2015). How effective is media self-regulation? results from a comparative survey of european journalists. *European journal of communication*, *30*(3), 249–266.

Fengler, S., & Speck, D. (2019). Journalism and transparency: A mass communications perspective. *Contested transparencies, social movements and the public sphere: Multi-disciplinary perspectives*, 119–149.

Gade, P. J., Dastgeer, S., DeWalt, C. C., Nduka, E.-L., Kim, S., Hill, D., & Curran, K. (2018). Management of journalism transparency: Journalists' perceptions of organizational leaders' management of an emerging professional norm. *International journal on media management*, *20*(3), 157–173.

Hallin, D., & Mancini, P. (2015). Media systems. *The international encyclopedia of political communication. New York: John Wiley & Sons.*

Hayes, A. S., Singer, J. B., & Ceppos, J. (2007). Shifting roles, enduring values: The credible journalist in a digital age. *Journal of Mass Media Ethics*, *22*(4), 262–279. https://doi.org/10.1080/08900520701583545

Hellmueller, L., Vos, T. P., & Poepsel, M. A. (2013). SHIFTING JOURNALISTIC CAPITAL? *Journalism Studies*, *14*(3), 287–304. https://doi.org/10.1080/1461670x.2012.697686

Humprecht, E., & Esser, F. (2016). Mapping digital journalism: Comparing 48 news websites from six countries. *Journalism*, *19*(4), 500–518. https://doi.org/10.1177/1464884916667872

Karlsson, M. (2010). RITUALS OF TRANSPARENCY. *Journalism Studies*, *11*(4), 535–545. https://doi.org/10.1080/14616701003638400

Karlsson, M., & Clerwall, C. (2018). Transparency to the rescue? evaluating citizens' views on transparency tools in journalism. *Journalism Studies*, *19*(13), 1923–1933.

Karlsson, M., Clerwall, C., & Nord, L. (2016). Do not stand corrected. *Journalism &amp Mass Communication Quarterly*, *94*(1), 148–167. https://doi.org/10.1177/1077699016654680

Karlsson, M., Clerwall, C., & Örnebring, H. (2014). Hyperlinking practices in swedish online news 2007–2013: The rise, fall, and stagnation of hyperlinking as a journalistic tool. *Information, Communication &amp Society*, *18*(7), 847–863. https://doi.org/10.1080/1369118x.2014.984743

Karlsson, M., Clerwall, C., & Örnebring, H. (2015). Hyperlinking practices in swedish online news 2007–2013: The rise, fall, and stagnation of hyperlinking as a journalistic tool. *Information, communication & society*, *18*(7), 847–863.

Kashyap, G., Mishra, H., & Bhaskaran, H. (2022). Data journalists' perception and practice of transparency and interactivity in indian newsrooms. *Media Asia*, *50*(1), 24–42. https://doi.org/10.1080/01296612.2022.2076324

Koliska, M. (2022a). Trust and journalistic transparency online. *Journalism Studies*, *23*(12), 1488–1509. https://doi.org/10.1080/1461670x.2022.2102532

Koliska, M. (2022b). Trust and journalistic transparency online. *Journalism Studies*, *23*(12), 1488–1509.

Koliska, M., & Chadha, K. (2018). Transparency in german newsrooms: Diffusion of a new journalistic norm? *Journalism studies*, *19*(16), 2400–2416.

Kovach, B., & Rosenstiel, T. (2001, January). *The elements of journalism: what newspeople should know and the public should expect*. https://www.lasalle.edu/~beatty/310/ACES_CD/reference/books/Elementsofjournalism.pdf

Kutz, D. O., & Herring, S. C. (2005). Micro-longitudinal analysis of web news updates. *Proceedings of the 38th annual Hawaii international conference on system sciences*, 102a–102a.

Lewis, J., Williams, A., & Franklin, B. (2008). Four rumours and an explanation: A political economic account of journalists' changing newsgathering and reporting practices. *Journalism practice*, *2*(1), 27–45.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2010). Practical resources for assessing and reporting intercoder reliability in content analysis research projects.

Loosen, W., Reimer, J., & Hölig, S. (2020). What journalists want and what they ought to do (in) congruences between journalists' role conceptions and audiences' expectations. *Journalism studies*, *21*(12), 1744–1774.

Masullo, G. M., Curry, A. L., Whipple, K. N., & Murray, C. (2021). The story behind the story: Examining transparency about the journalistic process and news outlet credibility. *Journalism Practice*, *16*(7), 1287–1305. https://doi.org/10.1080/17512786.2020.1870529

Mellado, C., Humanes, M. L., & Márquez-Ramírez, M. (2018). The influence of journalistic role performance on objective reporting: A comparative study of chilean, mexican, and spanish news. *International Communication Gazette*, *80*(3), 250–272.

Mellado, C., & Van Dalen, A. (2014). Between rhetoric and practice: Explaining the gap between role conception and performance in journalism. *Journalism Studies*, *15*(6), 859–878.

Mitrović, S., Andreoletti, D., & Ayoub, O. (2023). Chatgpt or human? detect and explain. explaining decisions of machine learning model for detecting short chatgpt-generated text. *arXiv preprint arXiv:2301.13852*.

Mor, N., & Reich, Z. (2017). From "trust me" to "show me" journalism. *Journalism Practice*, *12*(9), 1091–1108. https://doi.org/10.1080/17512786.2017.1376593

Newman, N., & Fletcher, R. (2017). Bias, bullshit and lies: Audience perspectives on low trust in the media. *Available at SSRN 3173579*.

Olsen, R. K., Kalsnes, B., & Barland, J. (2021). Do small streams make a big river? detailing the diversification of revenue streams in newspapers' transition to digital journalism businesses. *Digital Journalism*, 1–22.

Patterson, T. E. (2000). Doing well and doing good. *Available at SSRN 257395*.

Phillips, A. (2010a). TRANSPARENCY AND THE NEW ETHICS OF JOURNALISM. *Journalism Practice*, *4*(3), 373–382. https://doi.org/10.1080/17512781003642972

Phillips, A. (2010b). Transparency and the new ethics of journalism. *Journalism Practice*, *4*(3), 373–382.

Prochazka, F., & Obermaier, M. (2022). Trust through transparency? how journalistic reactions to media-critical user comments affect quality perceptions and behavior intentions. *Digital Journalism*, *10*(3), 452–472. https://doi.org/10.1080/21670811.2021.2017316

Reese, S. D., & Shoemaker, P. J. (2018). A media sociology for the networked public sphere: The hierarchy of influences model. In *Advances in foundational mass communication theories* (pp. 96–117). Routledge.

Reinemann, C., Stanyer, J., Scherr, S., & Legnante, G. (2012). Hard and soft news: A review of concepts, operationalizations and key findings. *Journalism*, *13*(2), 221–239.

Robles, F. A., Marín-Sanchiz, C. R., Abellán-Mancheño, A., & García-Avilés, J. A. (2023). Transparencia en los contenidos informativos. un análisis de métodos en el periodismo de datos español (2019-2022). *Anàlisi*, *68*, 97–116. https://doi.org/10.5565/rev/analisi.3548

Ryfe, D. (2021). The economics of news and the practice of news production. *Journalism Studies*, *22*(1), 60–76.

Saltzis, K. (2012). Breaking news online: How news stories are updated and maintained around-the-clock. *Journalism practice*, *6*(5-6), 702–710.

Singer, J. B. (2008). The journalist in the network: A shifting rationale for the gatekeeping role and the objectivity norm. *Blanquerna School of Communication and International Relations*, (23), 61–76.

Skovsgaard, M., Albæk, E., Bro, P., & De Vreese, C. (2013). A reality check: How journalists' role perceptions impact their implementation of the objectivity norm. *Journalism*, *14*(1), 22–42.

Tandoc, E. C., & Oh, S.-K. (2015). Small departures, big continuities? *Journalism Studies*, *18*(8), 997–1015. https://doi.org/10.1080/1461670x.2015.1104260

Timmerman, Y., & Bronselaer, A. (2022). Automated monitoring of online news accuracy with change classification models. *Information Processing &amp Management*, *59*(6), 103105. https://doi.org/10.1016/j.ipm.2022.103105

Timmerman, Y., Van Leuven, S., & Bronselaer, A. (2023). News in motion: A quantitative analysis of incremental news updates by flemish online news outlets. *Journalism Practice*, 1–27.

Uth, B., Badura, L., & Blöbaum, B. (2021). Perceptions of trustworthiness and risk: How transparency can influence trust in journalism. In *Trust and communication: Findings and implications of trust research*. Springer. https://doi.org/10.1007/978-3-030-72945-5

Van Atteveldt, W., Trilling, D., & Calderón, C. A. (2022). *Computational analysis of communication*. John Wiley & Sons.

Vos, T. P., & Craft, S. (2017). The discursive construction of journalistic transparency. *Journalism studies*, *18*(12), 1505–1522.

Ward, S. J. (2014). The magical concept of transparency. In *Ethics for digital journalists* (pp. 45–58). Routledge.

Zhang, S., Hu, Y., & Bian, G. (2017). Research on string similarity algorithm based on levenshtein distance. *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2247–2251.