



To Disclose or Not to Disclose: A Comprehensive Analysis Into the Article Transparency of News Websites

Roeland Dubèl¹ , Mark Boukes¹ ,
and Damian Trilling² 

Abstract

An increasing part of the public is distrustful toward journalism. Transparency has been advocated to counter this trend. Therefore, the question arises to what extent news outlets have implemented transparency. General content analyses of the implementation of transparency routines are scarce. We try to add to the literature by conducting a generic quantitative content analysis on the prevalence of a diverse set of transparency routines over multiple news outlets. We also assessed whether transparency not only differs between outlets but also within outlets; namely, between hard vs. soft news section items. We hypothesized that digital-native, public, and quality news outlets and hard news section items would be relatively more transparent. After analyzing 27,096 news items from six news outlets, we find that the outlets differed in the extent to which and areas in which they had implemented transparency: (a) The digital-native news outlet was more transparent than legacy news outlets, except for author transparency, (b) the public news outlet was more transparent in terms of news updates and source use compared to commercial news outlets, but less so in terms of authors and production processes, (c) no substantial systematic differences were found in the extent of transparency implementation between quality and popular news outlets, and (d) hard news section items were only more transparent in source use than in soft news section items. As being one of the only generic quantitative content analyses, this study has contributed to our understanding of the different patterns across news outlets in transparency implementation.

Keywords

transparency, online news, disclosure, routines

¹Universiteit van Amsterdam, The Netherlands

²Vrije Universiteit Amsterdam, The Netherlands

Corresponding Author:

Roeland Dubèl, Department of Communication Science, University of Amsterdam, Postbus 15791, 1001 NG Amsterdam, The Netherlands.

Email: r.dubel@uva.nl

Fueled by a perception that the quality of news is lacking, distrust is rising among segments of the public toward legacy journalism (Newman & Fletcher, 2017). This leads more readers to consume alternative news sources, which tend to contain more disinformation (Andersen et al., 2023) or drastically reduce their news intake (Hameleers, 2022), resulting in a less well-informed public. Although academic findings have been mixed, *transparency* is increasingly being advocated as a means to regain public trust (Karlsson, 2020). The rationale is that by becoming more open about how news is created, the knowledge gap between sender and recipient would be reduced, potentially enabling readers to better judge whether the news has been properly created (Craft & Heim, 2008). The digitization of news also contributed to an increased focus on transparency (Karlsson et al., 2016). Unlike offline news, online news can and does change after publication (Timmerman & Bronselaer, 2022). This fluidity serves as an argument for implementing more transparency mechanisms about what has changed in an article and why (Karlsson, 2010).

Due to these developments, news media are increasingly convinced of the necessity of more transparent journalism (Loosen et al., 2020). The digitization of news also provides opportunities to do so (Vos & Craft, 2017): The absence of page limits in the online context provides room for disclosures, and digital features such as hyperlinks can contribute to a transparent use of sources. Yet, obstacles hamper the implementation of transparency: Among others, increased time pressure and decreased revenues frustrate the implementation of transparency (Deprez & Raeymaeckers, 2012). Besides, some point out potentially negative consequences, depending on the stakeholder (Karlsson, 2021). For instance, research has shown that too much transparency about faulty information in earlier article versions could reflect negatively on the diligence of journalistic work (Fengler et al., 2015).

These trade-offs beg the question of how news media have balanced the advantages and disadvantages of routinely implementing transparency. Hence, we pose the overarching research question: *To what extent have online news outlets implemented transparency routines?*

Prior research that mapped which transparency routines news media have implemented generally consists of case studies: Often qualitative descriptions of single news outlets at a particular moment in time (Karlsson, 2010). These are valuable to get a detailed picture of specific implementations of transparency routines, but cannot speak to the implementation of such routines in daily reporting over a broader landscape and a longer time period. Such general quantitative analyses of the prevalence of transparency routines in news items are scarce (with notable exceptions Humprecht & Esser, 2016; Karlsson, 2010). To contribute to this literature, we therefore quantitatively analyze how the implementation of transparency has been realized for six outlets, over a period of 1 year, and over a diverse set of routines in the Dutch news media landscape.

The studies by Humprecht and Esser (2016) and Karlsson (2010) revealed that the implementation of transparency differs at the outlet-level, based among others, on their business model. Beyond this, one might also consider whether a difference exists at the item level (i.e., different news articles within the same outlets). For instance,

survey research suggests that readers perceive the level of transparency to vary from article to article (Uth et al., 2021). Moreover, experimental research indicates that a consistent implementation of transparency reinforces its beneficial impact in the form of reduced distrust (Masullo et al., 2021). Hence, this study will compare not only outlet-level features but also item-level features, namely whether a news item qualifies as hard or soft news.

Theoretical Background

From Norms to Routines

Transparency, in essence, comes down to the provision of information (Fengler & Speck, 2019). In the context of journalism, transparency involves providing greater insight into journalistic products and internal processes (Kovach & Rosenstiel, 2001). This may include offering background information about the authors, the rationale and motives behind reporting, and the sources and methods used (Craft & Heim, 2008; Ward, 2014). Transparency is not an end in itself but a means to realize a set of values. To begin with, openness about the choices behind reporting should generate more *understandability* for the public about how the news comes about (Fengler & Speck, 2019). This increases the *accountability* of news media, given that with more information, the public will be better able to evaluate and monitor how news media operates (Humprecht & Esser, 2016). Opening up internal processes to the public also provides more room for a *dialogue* between producer and consumer (Craft & Heim, 2008). By providing the reader with the *empowerment* to participate in the journalistic process, news media gain *legitimacy* (Karlsson, 2010). These values of transparency should lead to greater credibility and trust from the public (Phillips, 2010).

Whether transparency is an efficient means to achieve these goals remains to be seen, however. First, there are doubts about the effectiveness of transparency (Karlsson, 2021). Although readers are mostly positive about the idea of transparency (Karlsson et al., 2017), experiments to date tend to show no or small positive effects on credibility and trust in journalism (Curry & Stroud, 2021; Karlsson, 2020; Karlsson & Clerwall, 2018) and at times even negative effects (Karlsson & Clerwall, 2018; Tandoc & Thomas, 2017).

Second, there is uncertainty as to whether there are enough resources to implement transparency. For instance, news staff are increasingly experiencing time and work pressures (Deprez & Raeymaeckers, 2012). Journalists often argue that they simply do not have the time to add transparency to news coverage (Koliska & Chadha, 2018). Therefore, in light of efficiency, managers often direct journalists to limit themselves to low-effort methods of transparency (Chadha & Koliska, 2014; Gade et al., 2018) such as the heuristic use of timestamps to signal changes and indicating source usage via hyperlinks.

Finally, there are concerns regarding the desirability of the goals of transparency. One such question is whether transparency encourages “dis-accountability” rather than accountability: The demand for transparency stems, in part, from those who wish to discredit news media and gather as much intel to question the legitimacy of news

media (Craft & Heim, 2008). It is also far from obvious what the optimum amount of transparency would be (Tandoc & Thomas, 2017): Too much transparency could overload the public with information that distracts from the understandability of the news, making it counterproductive (Craft & Heim, 2008).

Notwithstanding these concerns, there is a growing conviction within the journalistic profession itself that transparency is part of how journalists ought to do their work (Loosen et al., 2020). The focus on transparency has led to new journalistic routines (Karlsson, 2020); that is, “patterns of behaviors that form the immediate structures of media work” (Reese & Shoemaker, 2018).

Transparency Routines

Transparency routines can roughly be divided into routines of *disclosure transparency* and *participatory transparency*. Disclosure transparency centers on communicating to the public how the news was selected and constructed (Karlsson, 2010). This includes routines that lead to more background information about authors, openness about source use, disclosure of changes, and breaking down the news production process. Participatory transparency goes beyond this by not only communicating with the public but also allowing the public to participate (Karlsson, 2010). This can, for instance, be achieved by facilitating feedback opportunities for readers through comment sections, forums, blogs or polls, and by providing chances for readers to co-produce the news.

The current manuscript concentrates on disclosure transparency for several reasons. First, initial optimism regarding interacting with the public has largely been replaced by a more pessimistic view (Reimer et al., 2023), visibly evidenced by the widespread shutdown of comment sections (Rossini, 2022). This is quite understandable: The lack of resources to moderate comment sections has resulted in users mostly adopting an uncivil tone (Bowd, 2016). We also saw this reflected in our data, where only one out of six outlets (the digital-native outlet) has an active comment section. Second, the public itself is less enthusiastic about participatory transparency compared to disclosure transparency (Karlsson & Clerwall, 2018), which holds true within the Dutch context (Van der Wurff & Schönbach, 2014). Hence, participatory transparency might be less efficient in achieving its goals of increased credibility and trust among the public.

This study, moreover, investigates transparency routines at the level of the news-item and disregards other tools such as the publication of editorial statues and the like. After all, news items are the final result of the news production process in which routines manifest themselves (Karlsson, 2010). In addition, it is mostly news items on which readers base their evaluation of journalism (Mellado, 2015), making it a vital space to build trust and credibility.

We examine the disclosure transparency routines as described by Karlsson (2010). We have categorized these into the following four categories: *Author* (author name, contact, and biographical information), *update* (timestamps and update disclosures), *source* (hyperlinks and sources in text), and *production* transparency (process

information). We discuss these in order of appearance in news items, working from the top to the bottom of the article. After describing how disclosure transparency can materialize at the news item level, we address why systematic differences in its implementation can be expected between and within outlets.

Author Transparency

The first category of disclosure transparency a reader may encounter is *author transparency*: Transparency about who wrote the article. Information about who wrote a news article allows the audience to understand and identify who produced the article rather than making it appear whether the article has been created by an anonymous voice (Reich, 2010). Hence, providing author information could contribute to the credibility of the author, the news item, and the outlet (Curry & Stroud, 2021; Johnson & St John, 2021). Nevertheless, greater author information could potentially evoke the impression that the news is only an individuals' account of the events that are being reported; that is, evoking the impression of subjectivity rather than objectivity (Tandoc & Thomas, 2017). Moreover, due to increasing online harassment against individual journalists, there is reluctance about exposing author information (Waisbord, 2020).

A starting point for author transparency is including the *author's name* in the byline. Disclosing the author names of news articles followed the introduction of the byline, which makes it one of the oldest forms of transparency (Koliska & Chadha, 2018). Giving explicit full names of the authors creates more openness about who wrote the story. This openness is less forthcoming when abstract author names are used, such as organizational or generic names (e.g., "by the economy newsroom"), or when a name is completely absent. Although giving an author's name would be a precondition for disclosing further author information, it has not yet been addressed in content analyses of transparency. Thus, little is known about the prevalence of author name transparency.

In addition, *author contact information* can be made available by supplementing the byline with either email addresses or socials. Disclosing contact information opens up the possibility for dialogue between consumer and producer (Chadha & Koliska, 2014). Previous research found this to be standard practice for major United States and European (United Kingdom and Sweden) news outlets (Chadha & Koliska, 2014; Karlsson, 2010).

Furthermore, *author biography information* can be provided (Chadha & Koliska, 2014). This is often limited to providing superficial information, such as purely linking to previous stories, education and/or employment of the author(s). Nonetheless, this can be a tool to communicate the competence and background of authors(s) (Uth et al., 2021).

Update Transparency

Second, disclosure transparency can be given about changes in an article, which is termed *update transparency*. Disclosing these changes is mostly evaluated positively

by readers (Karlsson & Clerwall, 2018). There are two ways to communicate these changes: Providing *detailed timestamps* for both the initial publication date and the last update made. This is the most common, yet also a heuristic form, of update transparency (Chadha & Koliska, 2014): It signals transparency but does not provide what and why the article has changed. Complementary, *textual update disclosures* can be used: Explaining what and why it was changed in the text, to take responsibility for the content of a change (Karlsson, 2010). However, this type is less common (Chadha & Koliska, 2014).

Disclosing every update, for example, even typos, might backfire by causing skepticism 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 be applied mainly to major cases (Koliska & Chadha, 2018).

Source Transparency

Third, the body of the article can disclose source usage, known as *source transparency* (Fengler & Speck, 2019). Source transparency routines can, in part, enhance the credibility of the outlet (Johnson & St John, 2021), and such routines are likewise perceived favorably by the public (Karlsson & Clerwall, 2018). Source transparency can be reached by hyperlinking, adding original documents, and sourcing in text.

Through *hyperlinking*, the embedding of clickable links in text, the source usage of journalists becomes traceable (Karlsson & Clerwall, 2018; Vos & Craft, 2017). A distinction can be made between internal and external hyperlinking. Where *internal hyperlinking* guides people to items from the same website, *external hyperlinking* guides people to sources elsewhere. Internal hyperlinks are more prevalent than their external counterparts (Humphrecht & Esser, 2016; Karlsson et al., 2015) because they are easier to insert: Internal stories are often tagged and thus easy to retrieve (Chadha & Koliska, 2014). Moreover, internal hyperlinking increases visitor duration, and hence, ad revenue. External hyperlinking lacks those advantages. Conversely, internal hyperlinking can obscure source traceability by, instead of linking directly to the external source, doing it indirectly, through an internal hyperlink, for profit. Still, the binary distinction of hyperlinks based on externality can be challenged. Increasingly, news outlets are run by the same parent company (Doyle, 2002). This blurs the commercial disadvantage of external hyperlinking: Although external hyperlinking sends people away, at the expense of ad revenue, this is less so when linking to an outlet that reverts profits to the same parent company. We will term these external hyperlinking patterns *family hyperlinking*.

Beyond linking to websites, outlets can also add *original documents* in news content to increase openness and credibility (Karlsson, 2010). However, the practice of doing so is rather limited compared to hyperlinking (Chadha & Koliska, 2014; Karlsson, 2010).

Source transparency can also be achieved through (textual) *sourcing*: Mentioning sources in the text itself (Chadha & Koliska, 2014). Presenting identifiable sources,

termed *explicit sources*, is one of the core tasks of being open to the public (Kovach & Rosenstiel, 2001). It allows readers to infer the motives and thus independence of sources (Carlson, 2010). Nonetheless, unidentifiable *unnamed sources* (Chadha & Koliska, 2014) are also commonly used. The degree of unidentifiability can vary: There may be no denomination at all for the source (e.g., “insider”), as with *anonymous sources*. Due to the call for transparency, journalists are increasingly being criticized for using anonymous sources (Carlson, 2010). Journalists themselves argue that appeals to anonymity are sometimes too easily granted. However, anonymity, for instance of whistle-blowers, is sometimes a must to obtain crucial intel (Carlson, 2010). Moreover, convincing sources to go public would be too time-consuming (Kovach & Rosenstiel, 2001). Unnamed sources can however also be attributed with details or abstract identifiers (e.g., “highly-ranked government official”) (Carlson, 2010), henceforth referred to as *opaque sources*. In the latter form, this primarily adds credibility to the source (Stenvall, 2008).

Production Transparency

Finally, disclosure transparency can also be provided through openness about the production processes behind the news. We call this *production transparency*. Giving insights about the production of the news is appreciated by the audience and can foster understanding among them (Karlsson & Clerwall, 2018). Furthermore, the presence of production information increases the perceived credibility of the author, the news story and the outlet (Curry & Stroud, 2021; Johnson & St. John III), while a lack of production information can lead to accusations of bias (Gravesteijn et al., 2024). Nevertheless, such information is rarely shared (Chadha & Koliska, 2014; Kovach & Rosenstiel, 2001). Journalists argue that elaborating on production processes would create too much confusion among the public, as these processes would be too difficult to explain due to their messy nature (Chadha & Koliska, 2014). Using a fabricated article, Figure 1 offers an overview of all discussed disclosure transparency routines.

External and Internal Differences

The implementation of the discussed disclosure transparency routines can differ greatly between outlet types (Karlsson, 2010). Prior work demonstrates that the *origin* (Salaverria, 2020), *funding* (Ryfe, 2021) and *profile* (Esser, 1999) of an outlet shape the routines and/or standards within the outlet. Likewise, these outlet-level features can explain the use of disclosure transparency routines.

Origin of Outlet. Although *legacy news media*, media that operated through television, radio or print, nowadays also own digital channels (Langer & Gruber, 2021), *digital-native news media* lack this analog history (Salaverria, 2020). Therefore, these digitally born outlets were never bound by the conditions, routines, and traditions that emerged from analog media (Karlsson & Clerwall, 2018). For instance, there is a lack of space to fulfill the norm of transparency in print, TV, and radio (Vos & Craft, 2017).

Large firms startled by sudden stock market crash



By [James Smith](#)



🕒 Monday, 12:00; **Last update:** Monday, 12:15

Correction: in an earlier version ...

Large companies react with surprise to the sudden stock market drop. In an earlier article, this news website reported on [this stock market crash](#). The director of one of the companies called it “disastrous.” For this article, we chose to only contact companies that are listed on the stock market, because these are regarded most relevant.

Author transparency

- 1 (explicit) author name
- 2 biographic information hyperlink
- 3 (email) contact information

Update transparency

- 4 Last updated timestamp
- 5 textual update disclosure

Source transparency

- 6 (internal) hyperlink source
- 7 (opaque) textual sourcing

Production transparency

- 8 process information

Figure 1. Fabricated Article Example Illustrating the Routines of Each Transparency Category.

This historical independence leads them to be more innovative, developing alternative workflows tailored to digital opportunities (Humphrecht & Esser, 2016; Kashyap et al., 2022). Also, the survivability of new ventures is enhanced by focusing on distinctiveness from existing competition (Buschow, 2020), making them more likely to be at the forefront of making digital capabilities their own. For instance, through the use of hyperlinking and detailed timestamps. Although legacy media today also ventures

online, the dominant analog workflow tends to persist, slowing down their digital innovation (Buschow, 2020). Previous research, for instance, found that digital-native news media were more active in disclosing their methods and linking to sources compared to legacy news media (Kashyap et al., 2022; Robles et al., 2023). In line with these findings, we expect:

Hypothesis 1 (H1): Routines regarding (a) author transparency, (b) update transparency, (c) source transparency, and, (d) production transparency are more prevalent in the content of a digital-native news outlet than in the content of legacy news outlets.

Funding of Outlet. Second, the funding of a medium affects the content it produces. We distinguish between state-funded *public media* and *commercial media*. According to *practice theory*, “strips of activity [i.e., practices] serve as meso-level resources that mediate the impact of macro-level forces” (Ryfe, 2021, p. 61). Financial considerations are a macro-level force.

For commercial media, dependence on commerce can serve as a barrier to adopting time-consuming and thus costly transparency routines. Public media are deprived of this financial pressure. Less driven by commercial trends, they may put more emphasis on the future return of audiences that transparency can bring through trust repair. Moreover, Dutch public media, unlike commercial media, are even required by law to (a) be independent of commercial influences and (b) adhere to high journalistic standards (Van Es & Poell, 2020). Hence, given the state resources and legal obligations of public media, they can be expected to incorporate more disclosure transparency routines compared to commercial media (Humphrecht & Esser, 2016). For instance, hyperlinking tends to be more prevalent in the content of public as opposed to commercial news media (Humphrecht & Esser, 2016). We expect:

Hypothesis 2 (H2): Routines regarding (a) author transparency, (b) update transparency, (c) source transparency, and, (d) production transparency are more prevalent in the content of a public news outlet than in the content of commercial news outlets.

Profile of Outlet. Third, the implementation of disclosure transparency might similarly differ for *popular* and *quality* news media, as they differ in business goals and target audiences. Whereas popular news outlets aim to reach a mass-market audience, through entertaining and sensationalist reporting, quality outlets aim to reach a highly educated up-market, through informative and rationalistic reporting (Doyle, 2013). Journalists from popular outlets therefore experience more commercial pressure, making them focus more on personalization and sensationalism in news coverage and less on objectivity and relevance compared to other types of journalists (Skovsgaard, 2014). Likewise, we expect that due to the commercial drive of popular outlets, their journalists will pay less attention to disclosure transparency compared to those of quality outlets. To exemplify, hyperlinking is more

present among quality news outlets as compared to popular outlets (Karlsson et al., 2015). Thus, we expect that:

Hypothesis 3 (H3): Routines regarding (a) author transparency, (b) update transparency, (c) source transparency, and, (d) production transparency are more prevalent in the content of quality news outlets than in the content of popular news outlets.

Hard Versus Soft News. Previous comparative research on transparency routines mainly made between-outlet instead of within-outlet comparisons. Nonetheless, the implementation of transparency tends to differ even *within* outlets (Gade et al., 2018). This resonates with the perception of consumers that transparency varies from item to item (Uth et al., 2021).

Therefore, both outlet and article features need consideration. For example, the distinction between quality and popular media is blurring, a process termed *tabloidization* (Esser, 1999). With increasing dependence on advertisers, more focus is arising on attention-grabbing features, to generate more views and thus ad revenue. Consequently, *hard news*, news that seeks to inform through factual coverage, and *soft news*, more personalized and entertaining storytelling (Reinemann et al., 2012), come to exist side by side. Where hard news is mostly prevalent in highly societal relevant sections (e.g., politics, economy, science and technology), the soft-style format mostly occurs in sections such as sports, entertainment and regional (Curran et al., 2010).

Thus, the division between quality and popular can also be made at the article level: Even within quality outlets, some sections are more entertaining and mass-oriented than others, potentially at the cost of journalistic standards. As such, it is expected that hard-news-related sections pertain to higher news standards compared to soft-news-related sections, and thus implement more disclosure transparency routines:

Hypothesis 4 (H4): Routines regarding (a) author transparency, (b) update transparency, (c) source transparency, and, (d) production transparency are more prevalent in the content of hard-news-related sections than in the content of soft-news-related sections.

Methodology

Data Collection

To make outlet comparisons, one digital-native (*NU.nl*), one public (*NOS*), two quality (*NRC & Volkskrant*), and two popular outlets (*AD & Telegraaf*) were selected.

These news websites are among the most-read in the Netherlands (Bakker, 2021). The determination of the class of the outlets followed previously made classifications (e.g., Louwse & Van Dijk, 2022; Vliegenthart et al., 2011).

To collect the data, a two-step approach was employed. First, article links from 2023 were scraped from the sitemaps of the respective websites.¹ Sitemaps provide an

overview of the hierarchical structure of pages within a website (Schonfeld & Shivakumar, 2009). For an example, see Figure 11 in the Appendix. For the *NOS*, no sitemap was available.

Instead, article IDs present in URLs were used (e.g., <https://nos.nl/artikel/1234567>). For all possible IDs between the first and last findable article of 2023, we assessed whether it yielded an existing URL, derived by returning an HTTP 200 status code.

URLs whose URL text showed that they referred to unconventional content were excluded (e.g., live, puzzle, podcast, video and newsletter pages). Of the remaining links, the content was scraped from a feasible random sample of 5,000 articles per outlet.² Articles that were too lengthy ($> \bar{X} + 3 \times \sigma = 10,206$ characters; indicative for live content) or contained too many hyperlinks ($>$ the 99.9th quantile = 14 hyperlinks; indicative for sports results overviews and travel tips) were excluded from further analysis.

Finally, the dataset was randomly downsampled to the least present news source ($n = 4,516$). Down-sampling was done to make balanced statements across the entire sample and led to a loss of only 9.68% of cases. This procedure eventually led to the inclusion of 27,096 news articles across six outlets for the main analyses.

Features

Article Features. Items from sections whose names correspond to conventional hard (domestic & foreign affairs, economy, politics and science) and soft news topics (entertainment, sports, celebrity and royalty) as described in the literature review by Reinemann et al. (2012) were automatically grouped into *hard* and *soft news* (62.42%). For items published on a section that cannot be clearly characterized as hard or soft news, rare sections and items published on multiple sections (37.58%), a procedure based on De León et al. (2023) was employed: A Naïve Bayes classifier was trained using the automatically grouped labels of a balanced sample in terms of outlet and hard versus soft news prevalence ($n = 1,800$). A Naïve Bayes classifier is a simple supervised machine learning algorithm, suitable to classify texts based on the occurrence of words (for an extensive introduction see Van Atteveldt et al., 2022). Here, the words occurring in the annotated sample were used to predict which classes the remaining items belonged to.

Compared to the automatically grouped labels, the classifier produced performance scores well above .7 (*soft*: $F_1 = .96$, *hard*: $F_1 = .97$), a common benchmark for good performance (Lombard et al., 2010).³ For all metrics of the classifier and remaining classifiers, see Table 8 in the Appendix. In the final dataset, 60.53% of the articles consisted of hard and 39.47% of soft news.

Author Features

Explicit Author(s). The presence of *explicit* author(s) in the byline was detected by a regular expression; “patterns for matching sequences of characters” (Van Atteveldt et al., 2022). The pattern assessed whether two capitalized words coexisted, indicative of giving full names. For the full translated pattern and other patterns, see Table 9 in

the Appendix. For *abstract* author(s) solely the word(s) editor(s), correspondent(s), or reporter(s) were present. Just mentioning the name of a press agency was also labeled as “abstract.” If no author description was present, it was automatically coded as *absent*. An *other* category existed for conceptually difficult-to-classify author(s) ($n = 197$), such as for disclosed AI-generated news content.

Author(s) Contact Information. For contact information, we examined whether a social media handle or email address was present (*yes*) or absent (*no*) in the byline.

Author(s) Biographic Information. Biographic information was present if the byline provided a URL containing the name(s) of the author(s). This was solely done for individual author(s).

Update Features. To detect changes, the “dateModified” tag present in the HTML file of web pages was used. This tag provides no timestamp or a timestamp equal to the publication date (*no change*) or a timestamp that exceeds the publication date, indicating that the page has been updated after publication (*changed*). To validate whether the dateModified tag correctly indicates changes, a second dataset was collected in which the front pages of the websites were routinely scraped. The dateModified tag correctly changed in tandem whenever textual changes occurred, except for *NOS*: Between iterations, textual changes occurred, while the dateModified tag remained empty.

For the *NOS* articles with a blank dateModified tag, a validation procedure was employed: An article version close to publication was retrieved via the *Wayback Machine* (a free web archive which routinely snapshots web pages) and automatically compared to the later retrieved version.⁴ This method revealed that 30.91% of the denoted “unchanged” articles, in reality, did change.

Update Timestamp Disclosure. Timestamp disclosures were absent if one timestamp appeared in the byline (*change undisclosed*), and present if more than one was present (*change disclosed*). We could not infer if the last-updated timestamp was equal to the dateModified tag, as sometimes modification was disclosed relative to access (e.g., “4 hours ago”). We could thus solely assess if an article ever changed and if a change was ever disclosed. Finally, an article could also not change (*no change*).

Update Textual Disclosure. To detect textual disclosures, a regular expression was used. Here, we searched whether the word “earlier” was near (within 20 characters) the word “version” (or synonyms). See Table 8 in the Appendix. This led to the detection of 221 items. From manually annotating a random sample from these hits ($n = 30$), 97% of cases proved to indeed contain an update textual disclosure.

Sourcing Features

Internal Hyperlinking. Internal hyperlinking was assessed by counting the number of hyperlinks containing the same base URL as the news website. As for every hyperlinking feature, this was later recoded to being present once (*yes*) or never (*no*).

Family Hyperlinking. For family hyperlinking the number of hyperlinks that referred to a website with the same parent company was counted. For an overview of the “family” of each outlet, the stems of related news websites are provided in Table 1.

External Hyperlinking. External hyperlinking was the count of links referring to a website outside of the own website and “family.”

Document Hyperlinking. Document hyperlinks were detected by assessing whether the word “document” or any common file type extension appeared in the link (e.g., .pdf, .docx, .txt, et cetera).

Textual Sourcing. First, we examined whether sentences contained a source at all via a Large Language Model (LLM) prompt. LLMs are trained on vast amounts of text and can solve annotation tasks (Liu et al., 2023). We used the LLM *google/flan-t5-xl* (see Chung et al., 2022) as it can be stored locally; thereby, making it reproducible. We prompted the following: *Does the text provide any form of sourcing?*⁵ This was asked about a randomly selected sentence of all sampled articles until each class (*absent & present*) contained 1,000 cases.⁶ These 2,000 hits were verified by manual annotation performed by the first author. For the full codebook see the Sourcing presence codebook in the Appendix. The LLM “pre-selected” the annotated sentences to ensure that classes were equally represented ($n_{absent} = 1051$, $n_{present} = 949$).

Table 1. Overview News Source Family.

NPO & RPO (Nederlandse en Regionale Publieke Omroepen)	DPG Media	Mediahuis
NOS (Nederlandse Omroep Stichting)	Algemeen Dagblad	De Telegraaf
I Limburg	De Volkskrant	NRC Handelsblad
NH Nieuws	NU.nl	Metro
Omroep Brabant	Trouw	De Gelderlander
Omroep Flevoland	BN DeStem	IJmuider Courant
Omrop Fryslân	Brabants Dagblad	Noordhollands Dagblad
Omroep Gelderland	De Gelderlander	Haarlems Dagblad
Omroep West	De Stentor	De Gooi- en Eemlander
Omroep Zeeland	Eindhovens Dagblad	Leidsch Dagblad
Radio en Televisie Rijnmond	Het Parool	Dagblad van het Noorden
Radio en Televisie Drenthe	Provinciale Zeeuwse Courant	Friesch Dagblad
Radio en Televisie Noord	Tubantia	Leeuwarder Courant
Radio en Televisie Oost		
Radio en Televisie Utrecht		

Note. The news outlets under study are printed in bold.

Second, the manual annotations were used to fine-tune a transformer model by supplementing it with annotated data. We used the *pdelobelle/robbert-v2-dutch-base* model for this (see Delobelle et al., 2020). These 2,000 annotated sentences are commonly enough to fine-tune a model (Van Atteveldt et al., 2022). The training of the transformer model boosted performance compared to the LLM prediction (*absent*: $F_1 = .83$, *present*: $F_1 = .81$), producing even better performance metrics (*absent*: $F_1 = .96$, *present*: $F_1 = .95$; see Table 8 in the Appendix). The model was consequently used to predict source presence on the sentence level.

Third, for the predicted sentences containing sourcing, 2,000 sentences were further manually analyzed, asking the question: *Which type of sourcing does the text provide?* Based partly on the descriptions of named and unnamed sources by Carlson (2010), the answer categories were the following: Sources that are not identifiable (*anonymous*), partly identifiable (*opaque*) or fully identifiable (*explicit*). Since the transformer model was not always precise, it could also be that—contrary to prediction—no sources were present (*absent*). See the Sourcing category codebook in the Appendix.

Fourth, the annotated sentences were again used to fine-tune a transformer model. Three datasets were used to train the model: (a) The sentences without sources present (*absent*) from the first manual annotation round, (b) the sentences of the second manual annotation round, and (c) given low presence, anonymous sources detected by a regular expression. This expression examined whether the abstract sourcing terms “insider(s),” “anonymous(ly),” or “source(s)” were used in a sentence (see Table 9 in the Appendix). The detected anonymous sources were manually verified. The model was able to classify the sourcing categories well (*absent*: $F_1 = .89$, *anonymous*: $F_1 = .99$, *opaque*: $F_1 = .76$, *explicit*: $F_1 = .86$). The model was then used to predict the type of sourcing in all the sentences that contained sourcing.

Finally, considering that opaque and anonymous sources can become explicit in other sentences, classification was collapsed to the highest transparency tier at the article level: Explicit if at least one sentence was explicit, opaque if there were only opaque sources, anonymous if there were only anonymous sources, and absent if the article lacked sources. All steps taken can be found in Figure 2.

Production Features

To retrieve process information, we took a four-step approach. First, “self-reference sentences” were extracted from articles. We assume that to release process information, an outlet or journalist must refer to itself. Self-referential sentences included those mentioning (a) “we,” “us,” “our,” “I,” “me,” “my,” or “mine” outside of quotes and opinion pieces; (b) their outlet name; or (c) individual author names of the article. This led to the inclusion of 9,405 sentences.

Second, a balanced random sample based on outlet amount was manually annotated ($n = 2,000$) for having process information (*yes* or *no*). Drawing on the descriptions of Chadha and Koliska (2014), Karlsson and Clerwall (2018), and Kovach and

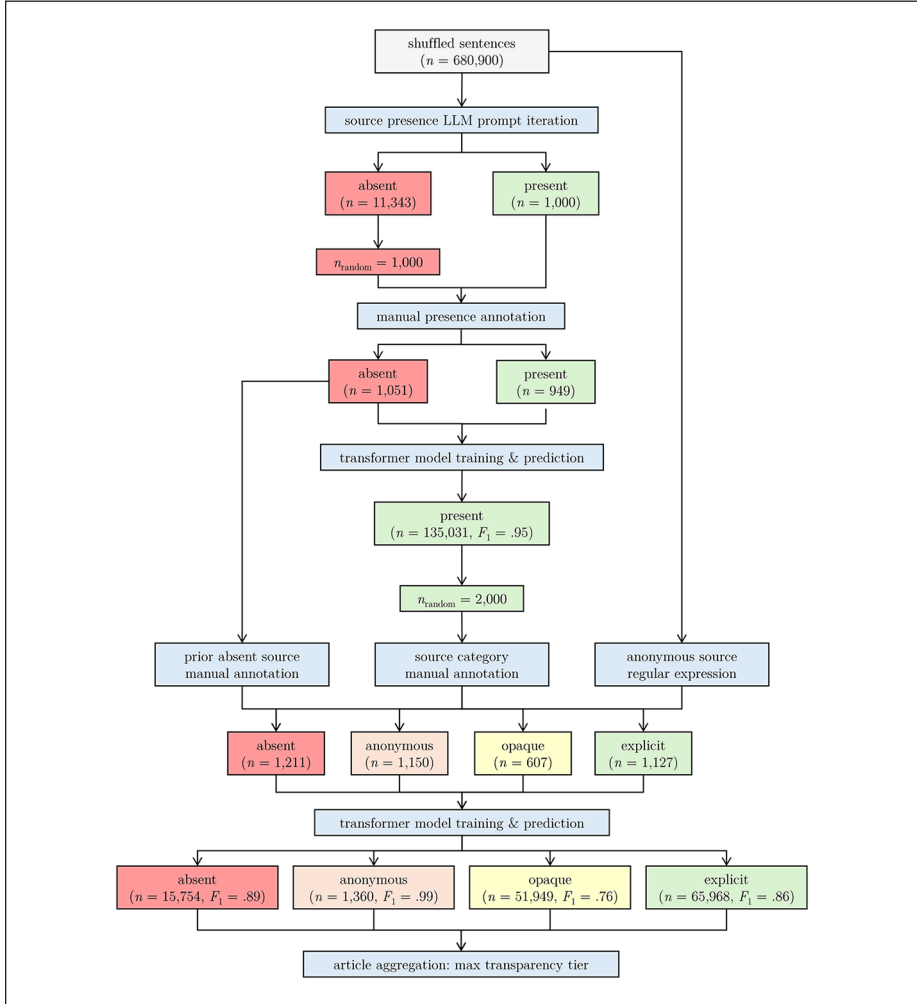


Figure 2. Textual Sourcing Detection Flowchart.

Rosenstiel (2001), we used the following question: *Does the text give any insight into when, why, how, or against which standards the article was created?* Examples include disclosing information about the origin, selection process, internal standards, motives, decisions, source gathering and production processes behind the news. The full codebook, including examples, is in the Appendix (see the Process information codebook).

Third, once again the transformer *pdelobelle/robert-v2-dutch-base* model was fine-tuned.⁷ The model reached a good performance in both classes compared to the manually annotated data (*no*: $F_1 = .79$, *yes*: $F_1 = .94$; see Table 8 in the Appendix).

Finally, we predicted the presence of process information among all self-referential sentences. Consequently, the number of hits was divided by the entire number of sentences of an article (thus, including non-self-referential sentences.) This produces a **process transparency factor** from 0, none of the sentences contain process information, to 1, all sentences contain process information. For a step-by-step overview see Figure 3.

Indices & Scales. To analyze the impact of the category author, update and source transparency, weighted transparency indices were calculated. The proportional score of the presence of a transparency routine over the entire sample was subtracted from 2, and taking the logarithm of this score to produce (roughly) normally distributed weights, $\log(2 - p_{d=1})$.⁸ The weights therefore now range between 0 and 1. By weighing we distinguish more common and thus, likely peripheral routines, from rare and thus, presumably costly routines. This theoretical distinction is often made in journalistic transparency literature (Chadha & Koliska, 2014). For an overview, see Table 2.

Consequently, the transparency routine dummies ($d_1 \dots d_i$) were multiplied by their weights ($w_1 \dots w_i$) and added together within each category ($\sum_{i=1}^{i=n} d_i \times w_i$). Finally, the transparency indices were normalized by applying a min-max scaling. This produces a **weighted author, update & source transparency index**. The exact formulae are offered below:

$$1. i_a = \sum_{d=1}^{n_d} (\log(2 - p_{d=1}) \times d)$$

$$2. \tilde{i}_a = \frac{i_a - \min(i_a)}{\max(i_a) - \min(i_a)}$$

Finally, for production transparency, a **process transparency scale** was made in which a min-max normalization was applied to the process information factor, given that this category consists of one indicator.

Analytical Strategy. Pairwise independent samples *t*-tests are used to compare mean indices and scales between outlets. Adjusted *p*-values are presented in concordance with the strict Bonferroni correction: *p*-values have been multiplied by 15 since for each transparency category, 15 *t*-tests have been run, $C(6, 2) = 15$.

To test the effect of hard (vs. soft) news items on the (weighted) transparency indices and scale (y_{ij}), a multilevel model is run in which news items are clustered among outlets. The intercept is broken down into a fixed part (γ_{00}), the constant between all news items, and a random part between outlets (u_{0j}), given that the (weighted) transparency indices and scales vary between outlets. Likewise, the effect of hard (vs. soft) news is broken down into a fixed ($\gamma_{10}x_{1ij}$) and random part ($u_{1j}x_{1ij}$), as the effect may vary between outlets. By factoring in the intercept and effect variance, the effect of hard (vs. soft) news can be better isolated from the other factors in the model:

$$3. y_{ij} = \gamma_{00} + u_{0j} + \gamma_{10}x_{1ij} + u_{1j}x_{1ij} + \int_{ij}$$

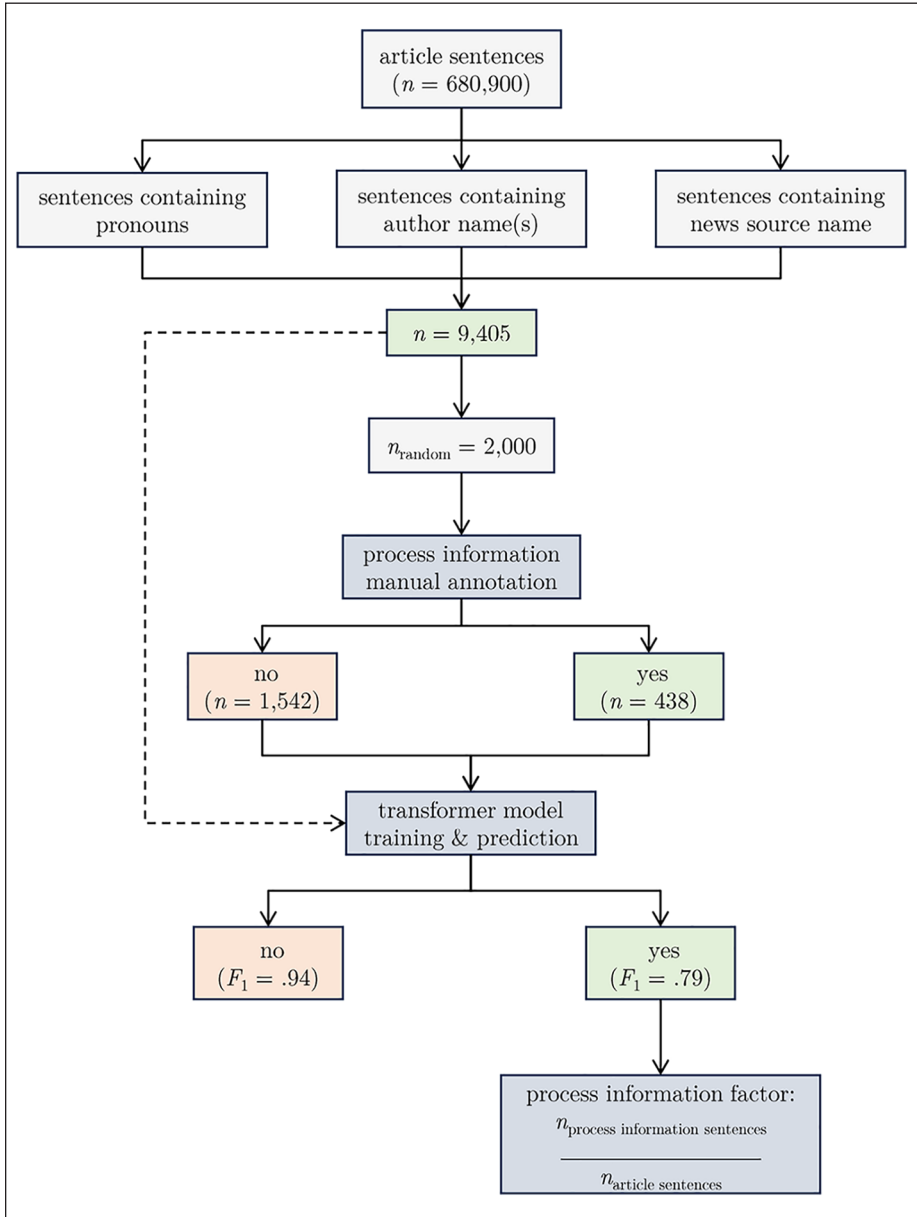
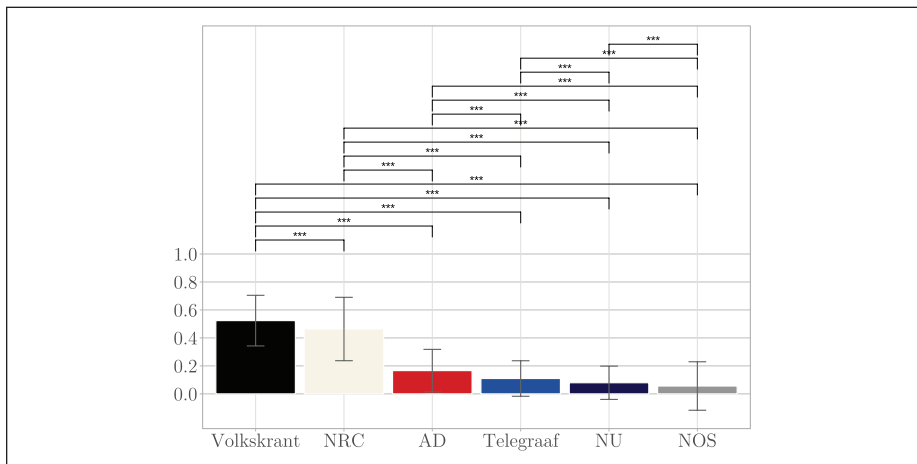


Figure 3. Process Information Detection Flowchart.

Table 2. Overview Transparency Rituals Weights.

Category	Routine	Weight
Author	Explicit author(s)	0.39
Author	Contact information	0.68
Author	Biographic information	0.55
Update	Timestamp disclosure	0.54
Update	Textual disclosure	0.69
Source	Document hyperlinking	0.68
Source	External hyperlinking	0.60
Source	Internal & family hyperlinking	0.56
Source	Explicit sourcing	0.27

**Figure 4.** Mean Author Transparency Index Per Outlet.

Note. Error bars showing mean scores and standard deviations within outlets. Significance indicators indicate the significance of mean differences between outlets after the Bonferroni correction: $p \geq .05 = ns$; $*p < .05$. $**p < .01$. $***p < .001$.

Results

Author Transparency Analysis

Against the expectations of Hypothesis 1a and 2a, digital-native outlet *NU.nl* ($M = .08$, $SD = .12$) and public outlet *NOS* ($M = .06$, $SD = .17$) have the lowest author transparency indices on average. In line with Hypothesis 3a, the quality outlets *Volkskrant* ($M = .52$, $SD = .18$) and *NRC* ($M = .46$, $SD = .23$) have higher mean transparency indices than the popular outlets *Telegraaf* ($M = .11$, $SD = .13$) and *AD* ($M = .17$, $SD = .15$). All mean differences at the outlet level are significant as shown in Figure 4. Regarding Hypothesis 4a we did not find an effect of hard (vs.) soft news on author transparency across outlets ($\beta = -.018$, $p = .367$, $CI [-.018, .050]$).

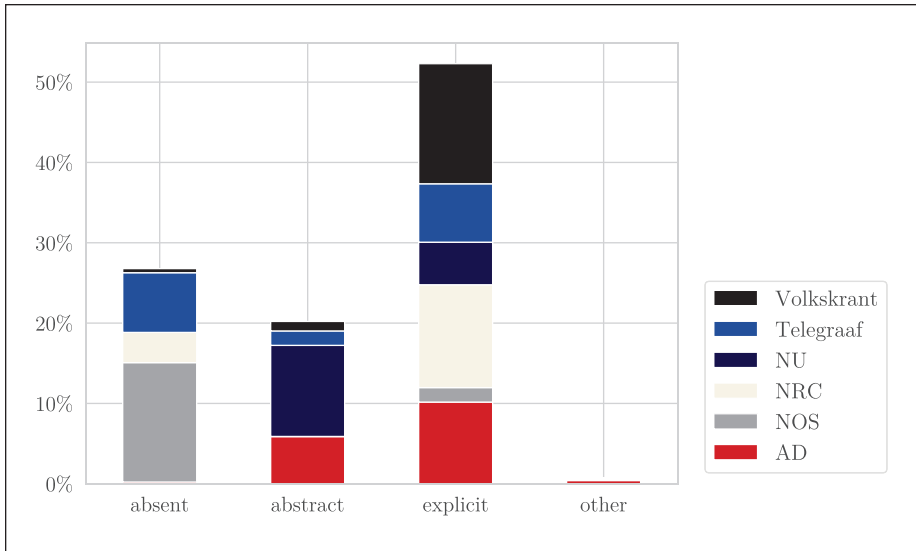


Figure 5. Author Transparency Class Per Outlet.

Quality outlets *Volkskrant* and *NRC* score relatively high on the author transparency indices; this is especially evident in the explicit mentioning of author names ($\%_{Volkskrant} = 89.59$, $\%_{NRC} = 73.91$), see Figure 5. In other words, quality outlets most frequently give the first and last names of authors in the byline. The public outlet *NOS*, with the lowest index, does this the least out of all outlets (10.70%).

In contrast, the public outlet delivers contact information in 7.13% of all its articles, the highest percentage out of all outlets ($\%_{AD} = 4.45$, $\%_{NOS} = 7.13$, $\%_{NRC} = 6.53$, $\%_{NU} = 0.53$, $\%_{Telegraaf} = 1.00$, $\%_{Volkskrant} = 0.75$). This contradicts the relatively low percentage of providing explicit author names. Put differently, in the rare case in which the *NOS* provides an explicit author name, it is often complimented with contact information ($\rho_{Pearson} = .79$). As for providing biographic information, solely the high-ranked quality news outlets do so and do so for most of their articles ($\%_{NRC} = 73.91$, $\%_{Volkskrant} = 89.59$).

Update Transparency Analysis

Excluding the unchanged items, in line with Hypothesis 1b and 2b, the digital-native outlet *NU.nl* ($M = .34$, $SD = .19$) and public outlet *NOS* ($M = .20$, $SD = .22$) score the second and third highest update transparency indices on average (see Figure 6). Contrary to Hypothesis 3b, quality outlets *NRC* ($M = .02$, $SD = .10$) and *Volkskrant* ($M = .01$, $SD = .07$) have the lowest average update transparency indices. Whereas popular outlet *AD* ($M = .41$, $SD = .12$) has the highest mean update transparency

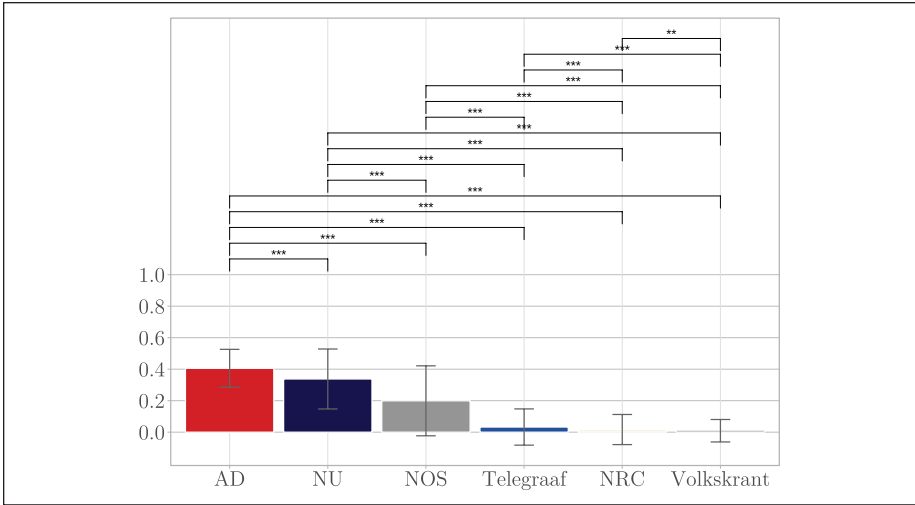


Figure 6. Mean Update Transparency Index Per Outlet.
 Note. Error bars showing mean scores and standard deviations within outlets. Significance indicators indicate the significance of mean differences between outlets after the Bonferroni correction: $p \geq .05 = ns$. * $p < .05$. ** $p < .01$. *** $p < .001$.

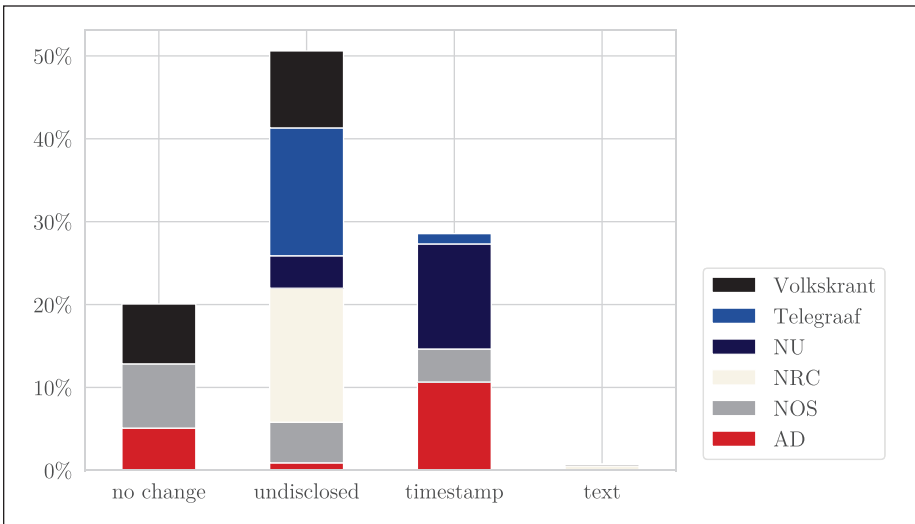


Figure 7. Update Transparency Class Per Outlet.

index, *Telegraaf* ($M = .03, SD = .12$) has one of the lowest. Regarding Hypothesis 4b, we did not find an effect of hard (vs.) soft news on update transparency across outlets ($\beta = .007, p = .650, CI [-.025, .040]$).

Still, the high mean transparency indices of popular outlet *AD*, digital-native outlet *NU*, and public outlet *NOS* are mostly due to disclosing updates via time-stamps, the more peripheral route ($\%_{AD} = 63.91$, $\%_{NOS} = 23.38$, $\%_{NU} = 75.84$, $\%_{Telegraaf} = 7.44$), see Figure 7. In the few cases that the low-ranked quality outlets *Volkskrant* and *NRC* reveal an update, they do so relatively most often via textual revelation ($\%_{NRC} = 2.99$, $\%_{Volkskrant} = 0.93$, $\%_{AD} = 0.13$, $\%_{NOS} = 0.18$, $\%_{NU} = 0.44$, $\%_{Telegraaf} = 0.00$).

Source Transparency Analysis

Although with small margins, contrary to Hypothesis 1c, digital native outlet *NU* ranks fourth in mean source transparency index ($M = .20$, $SD = .18$), see Figure 8. In line with Hypothesis 2c, public outlet *NOS* ($M = .30$, $SD = .22$) scores among the highest mean source transparency indices. As for Hypothesis 3c, quality outlet *NRC* has the highest mean source transparency index ($M = .32$, $SD = .24$). Yet, fellow quality outlet *Volkskrant* has the second lowest index ($M = .19$, $SD = .19$), falling between popularity outlets *AD* ($M = .18$, $SD = .17$) and *Telegraaf* ($M = .08$, $SD = .10$). In line with Hypothesis 4c, there is an effect of hard (vs.) soft news on source transparency across outlets ($\beta = .080$, $p = .003$, $CI [.027, .132]$).

The high mean transparency indices of quality outlet *NRC* and public outlet *NOS* is partly due to the highest presence of external hyperlinks in articles ($\%_{NRC} = 40.48$, $\%_{NOS} = 21.86$), see Figure 9. The *NOS*, then again, also uses the more peripheral internal hyperlinks (41.28%) and family hyperlinks (11.09%) most often. Linking to documents is a rarity, but again *NRC* and *NOS* score highest in this regard ($\%_{NRC} = 5.91$, $\%_{NOS} = 2.41$), complemented by fellow quality outlet *Volkskrant* (2.48%). Also, in terms of source use in text, *NRC* and *NOS* articles most often include an explicit source ($\%_{NRC} = 81.86$, $\%_{NOS} = 77.02$). The low source transparency of popular outlet *Telegraaf* is striking: It scores lowest on almost every indicator, whereas on the use of anonymous sources they score highest (2.02%).

Production Transparency Analysis

Also concerning production transparency, in line with Hypothesis 1d, digital-native outlet *NU.nl* has one of the highest scores ($M = .010$, $SD = .047$), see Figure 10. In contrast, as for Hypothesis 2d, the public outlet *NOS* has the lowest average process transparency scale ($M = .002$, $SD = .018$). Again, regarding Hypothesis 3d, the quality outlets show a divergent pattern: *NRC* has the highest average process transparency scale ($M = .015$, $SD = .052$), while *Volkskrant* ($M = .002$, $SD = .022$) scores similar to popularity outlets *AD* ($M = .004$, $SD = .032$) and *Telegraaf* ($M = .004$, $SD = .038$). As for Hypothesis 4d, we did not find evidence of a hard (vs.) soft news on the process transparency scale ($\beta = .026$, $p = .265$, $CI [-.020, .073]$).

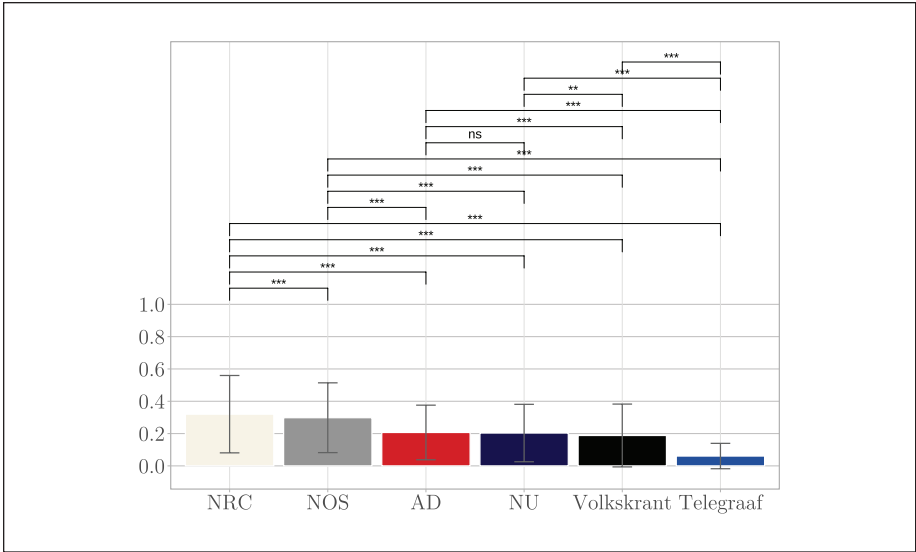


Figure 8. Mean Source Transparency Index Per Outlet.

Note. Error bars showing mean scores and standard deviations within outlets. Significance indicators indicate the significance of mean differences between outlets after the Bonferroni correction: $p \geq .05 = ns$. * $p < .05$. ** $p < .01$. *** $p < .001$.

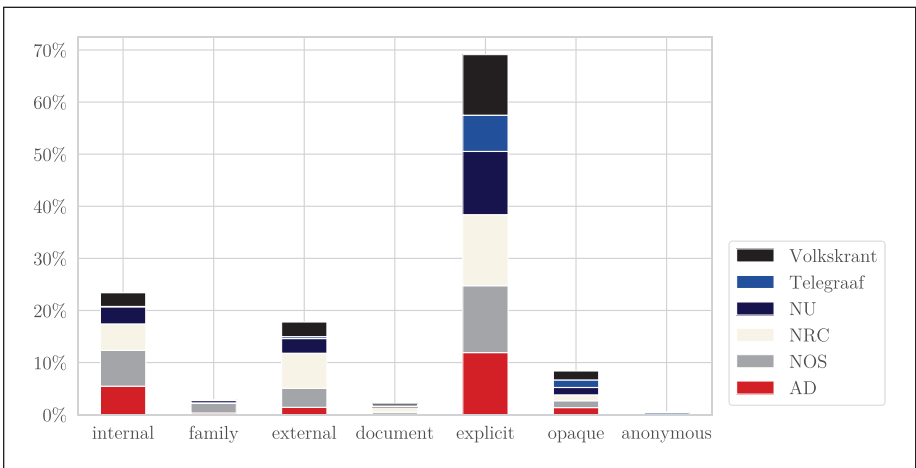


Figure 9. Source Transparency Class Per Outlet.

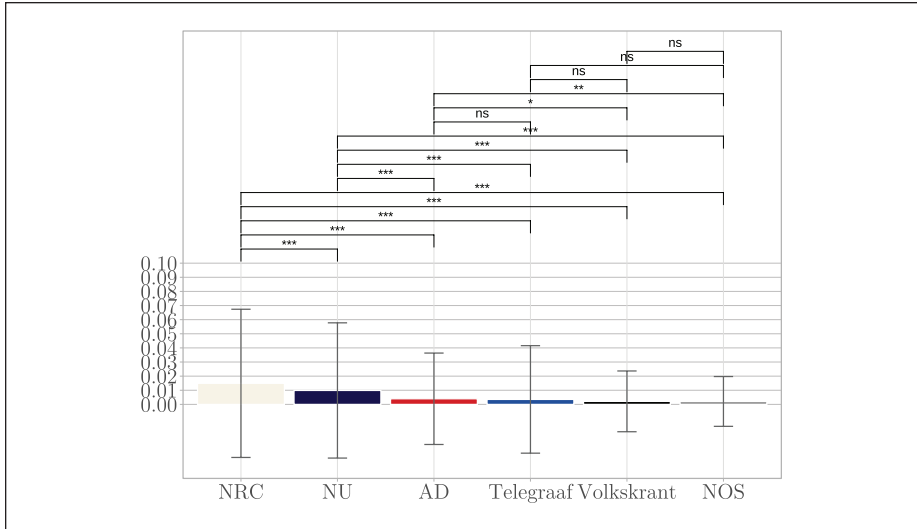


Figure 10. Mean Process Transparency Scale Per Outlet.

Note. Error bars showing mean scores and standard deviations within outlets. Significance indicators indicate the significance of mean differences between outlets after the Bonferroni correction: $p \geq .05 = ns$. $*p < .05$. $**p < .01$. $***p < .001$.

See Table 3 for an overview of the support of all hypotheses across author, update, source and production transparency. See Tables 4, 5, 6 and 7 for the mean differences for the indices and scale between each outlet combination.

Table 3. Overview Hypotheses Support.

Hypotheses	Author	Update	Source	Production
H1: Digital-native media contain more transparency than legacy media	No	Yes	Yes	Yes
H2: Public media contain more transparency than commercial media	No	Yes	Yes	No
H3: Quality media contain more transparency than popular media	Yes	No	No	No
H4: Hard news contains more transparency than soft news	No	No	Yes	No

Table 4. Mean Differences Weighted Author Transparency.

Pair	Δ_M	<i>p</i>
AD, NOS	0.11	<.001
AD, NRC	-0.30	<.001
AD, NU	0.09	<.001
AD, Telegraaf	0.06	<.001
AD, Volkskrant	-0.36	<.001
NOS, NRC	-0.41	<.001
NOS, NU	-0.02	<.001
NOS, Telegraaf	-0.05	<.001
NOS, Volkskrant	-0.47	<.001
NRC, NU	0.38	<.001
NRC, Telegraaf	0.35	<.001
NRC, Volkskrant	-0.06	<.001
NU, Telegraaf	-0.03	<.001
NU, Volkskrant	-0.44	<.001
Telegraaf, Volkskrant	-0.41	<.001

Table 5. Mean Differences Weighted Update Transparency.

Pair	Δ_M	<i>p</i>
AD, NOS	0.21	<.001
AD, NRC	0.39	<.001
AD, NU	0.07	<.001
AD, Telegraaf	0.37	<.001
AD, Volkskrant	0.40	<.001
NOS, NRC	0.18	<.001
NOS, NU	-0.14	<.001
NOS, Telegraaf	0.16	<.001
NOS, Volkskrant	0.19	<.001
NRC, NU	-0.32	<.001
NRC, Telegraaf	-0.02	<.001
NRC, Volkskrant	0.01	.002
NU, Telegraaf	0.30	<.001
NU, Volkskrant	0.33	<.001
Telegraaf, Volkskrant	0.02	<.001

Table 6. Mean Differences Weighted Source Transparency.

Pair	Δ_M	<i>p</i>
AD, NOS	-0.09	<.001
AD, NRC	-0.11	<.001

(continued)

Table 6. (continued)

Pair	Δ_M	<i>p</i>
AD, NU	0.00	1.000
AD, Telegraaf	0.15	<.001
AD, Volkskrant	0.02	<.001
NOS, NRC	-0.02	<.001
NOS, NU	0.09	<.001
NOS, Telegraaf	0.24	<.001
NOS, Volkskrant	0.11	<.001
NRC, NU	0.12	<.001
NRC, Telegraaf	0.26	<.001
NRC, Volkskrant	0.13	<.001
NU, Telegraaf	0.14	<.001
NU, Volkskrant	0.02	.002
Telegraaf, Volkskrant	-0.13	<.001

Table 7. Mean Differences Process Transparency Scale.

Pair	Δ_M	<i>p</i>
AD, NOS	0.002	.008
AD, NRC	-0.011	<.001
AD, NU	-0.006	<.001
AD, Telegraaf	0.000	1.000
AD, Volkskrant	0.002	.024
NOS, NRC	-0.013	<.001
NOS, NU	-0.008	<.001
NOS, Telegraaf	-0.001	.302
NOS, Volkskrant	0.000	1.000
NRC, NU	0.005	<.001
NRC, Telegraaf	0.011	<.001
NRC, Volkskrant	0.013	<.001
NU, Telegraaf	0.006	<.001
NU, Volkskrant	0.008	<.001
Telegraaf, Volkskrant	0.001	.528

Discussion

Given the increasing embrace of transparency in journalism as a means of building trust (Karlsson, 2020), we studied how six Dutch news outlets implemented transparency. Altogether, every news outlet tried to implement transparency in some way in the categories studied. However, through which routines and to what extent this was done differed from outlet to outlet.

First, the only digital-native outlet was more transparent than legacy news outlets, except for author transparency. This relatively high degree of transparency of the digital-native outlet can be explained, in part, by the need to distinguish oneself with the use of digital features (Humprecht & Esser, 2016; Kashyap et al., 2022). Indeed, in terms of update transparency, the digital news outlet *NU* stands out by using update timestamps most often. However, for virtually every hyperlink routine, *NU* is among those who use them least. Possibly the digital news outlet's transparency is driven not so much by drawing from digital capabilities, but more by having a different value structure in which transparency is more central compared to the legacy news outlets. That legacy news outlets are leaders in hyperlinking may suggest that the analog workflow does not persist as dominantly on their digital platforms as much as previously assumed (Buschow, 2020). As for author transparency, it is a familiar pattern that digital-native outlets fall short in this regard (Carlson, 2010). Possibly digitization processes have called into question the definition of what constitutes a journalist and thus to whom freedom of the press belonged. Given the debate about whether this new generation was entitled to the same legal protections (Oster, 2013), author anonymity was, perhaps, at first, preferred. This behavior may have persisted.

Second, the public news outlet is more transparent than commercial news outlets in terms of updates and source use but is the least transparent in terms of authors and production processes. It was hypothesized that the public news outlet *NOS* would implement more transparency given that financial pressure is less of a consideration due to receiving state funding (Humprecht & Esser, 2016). This does not rule out the possibility that the *NOS* may well experience time pressure. Indeed, even in the areas where *NOS* is transparent, there seems to be a preference for less time-consuming transparency routines: The more superficial display of updates via timestamps rather than textual disclosures and hyperlinking the more easily retrievable internal and family hyperlinks rather than the more traceable external hyperlinks. That some commercial news outlets sometimes operate more transparently than *NOS* may demonstrate that they are less affected by short-term financial incentives than assumed. Thus, they might also pin their hopes on the future return of audiences by restoring trust through transparency. The low transparency of *NOS* in authors may be due to balancing it with their safety. Namely, during the pandemic, journalists of *NOS* experienced much harassment (NOS, 2020). Yet, this can potentially set in motion an undesirable cycle: Author anonymity does not allow the public to judge authors' independence (Koliska & Chadha, 2018), fueling distrust and thereby, possibly the incidence of attacks.

Third, both quality news outlets are more transparent about their authors than the popular news outlets. For update transparency, the quality news outlets score lowest, while for source and production transparency, the quality news outlets show opposing findings: Whereas *NRC* is among the most transparent in these areas, *Volkskrant* is among the least. This inconsistency might be explained by further distinguishing the profiles of both outlets: Looking at how much hard news (vs. soft news) the outlets produce, the quality news outlets record the highest percentages, but *NRC* still produces substantially more hard than soft news compared to *Volkskrant* ($\Delta_{\%} = 11.79$). That no unequivocal conclusion can be drawn about the degree of transparency

implementation of quality vs. popular news outlets may prove the tabloidization process in which the standards between quality and popular news outlets are blurring (Esser, 1999). At a more refined level, systematic differences do emerge: If transparent, quality news outlets tend to choose more informative routines compared to popular news outlets, such as providing authors their biographical and contact info and textual update disclosures.

Within outlets, only source use is systematically more transparent in hard news sections than in soft news sections. Again, the lack of systematic differences in transparency between hard and soft news sections signals a tabloidization process: Standards blur not only between but also within outlets, namely between hard and soft news sections. That hard-news sections are systematically more transparent about source use than soft-news sections may be due to the more human-interest orientation of soft-news, where sources are less pivotal compared to the more factual hard-news (Reinemann et al., 2012).

As one of the few generic content analyses into transparency routines, this study contributed to our understanding of how different outlets implement transparency. That outlets differ in the implementation of transparency, suggests that, in line with prior research (Humprecht & Esser, 2016), it stems from different value structures and policy choices between outlets. Given that the differences in transparency implementation between outlets do not always align with theoretical expectations, the question is raised to what extent they do align with the expectations of outlets' own audiences. If these are not met, it can lead to disappointment among their audiences (Steindl et al., 2024), resulting in a decline of media trust. A linkage analysis—the pairing of survey and content analysis data (De Vreese et al., 2017)—could potentially reveal the extent to which audience expectations align with transparency implementation.

The implementation of transparency may also depend on the country context. As for Dutch news outlets, trust is relatively high compared to other countries (voor de Media, 2023). This may put less pressure on Dutch news outlets to make their current practices more transparent. Yet, as the Netherlands has many news outlets (Hallin & Mancini, 2004), this high degree of competition may contribute to the need to distinguish oneself through being more transparent than one's competitors. Future research is needed to explore whether the transparency implementation patterns found in this study generalize, among others, for countries with other degrees of media trust and competition.

Appendix

```
▼<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_general.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_sections.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W48.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W47.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W46.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W45.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W44.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W43.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W42.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W41.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W40.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W39.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W38.xml</loc>
  </sitemap>
  ▼<sitemap>
    <loc>https://www.nu.nl/sitemap_2023_W37.xml</loc>
```

Figure 11. Sitemap Example NU.nl.

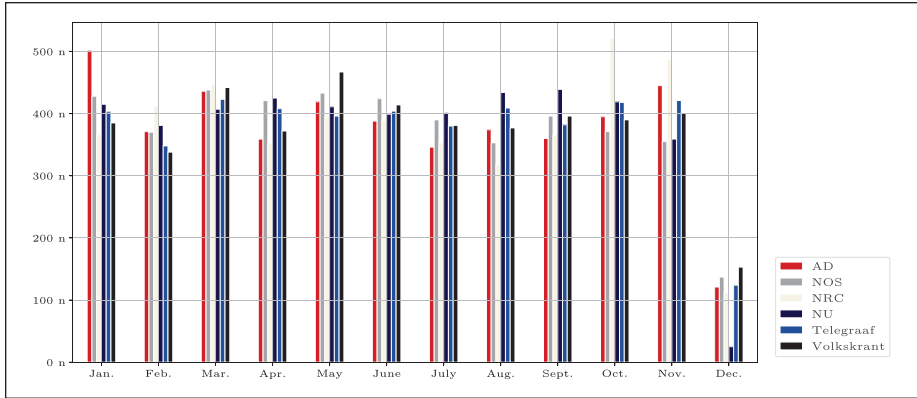


Figure 12. Sample Size Per Month Per News Outlet.

Table 8. Performance Metrics Overview.

Feature	Model	Class	Precision	Recall	FI-score	Accuracy
Article type	Naive Bayes	No	0.99	0.94	0.96	0.97
Article type	Naive Bayes	Yes	0.95	0.99	0.97	
Sourcing presence	Naive Bayes	No	0.84	0.76	0.80	0.79
Sourcing presence	Naive Bayes	Yes	0.75	0.83	0.78	
Sourcing presence	LLM	No	0.86	0.81	0.83	0.82
Sourcing presence	LLM	Yes	0.79	0.84	0.81	
Sourcing presence	Roberta	No	0.98	0.91	0.94	0.94
Sourcing presence	Roberta	Yes	0.90	0.98	0.94	
Sourcing category	Logit Regression	Absent	0.72	0.82	0.77	0.77
Sourcing category	Logit Regression	Anonymous	1.00	0.99	0.99	
Sourcing category	Logit Regression	Opaque	0.60	0.43	0.50	
Sourcing category	Logit Regression	Explicit	0.67	0.68	0.68	
Sourcing category	Roberta	Absent	0.92	0.87	0.89	0.89
Sourcing category	Roberta	Anonymous	1.00	0.99	0.99	
Sourcing category	Roberta	Opaque	0.79	0.73	0.76	
Sourcing category	Roberta	Explicit	0.82	0.91	0.86	
Process information	Naive Bayes	No	0.88	0.98	0.93	0.88
Process information	Naive Bayes	Yes	0.88	0.52	0.66	
Process information	Roberta	No	0.94	0.93	0.94	0.90
Process information	Roberta	Yes	0.77	0.81	0.79	

Table 9. Overview Regular Expressions.

Routine	Expression
Explicit author(s)	[A-Z ÇÖ][a-z öéüñá].*[A-Z ÇÖ][a-z öéüñá]
Abstract author(s)	redact correspondent verslaggever ANP
Contact information	@ https://twitter.com/\w+?
Biographic information	/auteur/[A-ZÇÖa-z].*(-)[A-ZÇÖa-z] ombudsman tag/[A-ZÇÖa-z].*%
Update textual disclosure	\b(eerder vorig).{0,20}\b(versie publicatie bericht\b) \b(versie publicatie).{0,20}\b(eerder vorig)
Document linking	document .pdf .epdf .doc .xls .csv .txt .jpeg .jpg .png .gif .mp4 .avi .ppt .m4a .mp3 .wav .zip
Anonymous sourcing	\b(anoniem(e)? bronn(en)?(?! van\b) ingewijde(n)?)\b

Sourcing Presence Codebook

Does the text provide any form of sourcing?

- *Yes*
- *No*

Examples of sources are . . .

- Anonymous sources: sources are not identifiable by withholding full names and disclosing little to no descriptive features:
- “Sources,” “insiders,” etc.
- Opaque sources: sources are only partly identifiable by withholding full names and solely providing abstract or overarching features:
- “The media,” “several press agencies,” “experts,” etc.
- Explicit sources; sources are directly identifiable by providing full names:
- Specific news and/or public organizations, governmental bodies, corporate entities, identifiable statistics/content/reports/individuals.

Sourcing Category Codebook

Which type of sourcing does the text provide?

- *absent*
- *anonymous*
- *opaque*
- *explicit*

Examples of sources are . . .

- Anonymous sources: sources are not identifiable by withholding full names and disclosing little to no descriptive features:

- “Sources,” “insiders,” etc.
- Opaque sources: sources are only partly identifiable by withholding full names and solely providing abstract or overarching features:
- “The media,” “several press agencies,” “experts,” etc.
- Explicit sources; sources are directly identifiable by providing full names:
- Specific news and/or public organizations, governmental bodies, corporate entities, identifiable statistics/content/reports/individuals.

Process Information Codebook

Does the text give any insight into when, why, how, or against which standards the article was created?

- *Yes*
- *No*

Examples are providing explanations of . . .

- The origin of the news and/or news selection processes:
- This article is a new version of a piece we published late last month. Since the judge is ruling today, we are bringing it to your attention again.
- A version of this article also appeared in the newspaper of 11 October 2023.
- Internal news standards and/or motives:
- With negative news often dominating NU.nl, the positive news sometimes snows under. That is why we list cheerful news every week.
- Internal news decisions:
- The NOS chose not to approach companies falling under electricity, cement and hydrogen because little of that is currently imported from outside the EU to the Netherlands.
- News sourcing and production processes:
- The Volkskrant contacted King Saud University by email but did not receive a response.
- This summary was created using AI and checked by NU.nl.

Note: The role of the media company and or its workers must be proactively disclosed. For instance, whereas simply stating which sources are used does not constitute process information, indicating how sources were contacted by the media company and or its workers and/or why sources were used does.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work used the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-7673. This work was supported by the Dutch Research Council (NWO) with a Vidi grant under project number: VI.Vidi.211.101.

ORCID iD

Roeland Dubèl  <https://orcid.org/0009-0005-6685-2323>

Mark Boukes  <https://orcid.org/0000-0002-3377-6281>

Damian Trilling  <https://orcid.org/0000-0002-2586-0352>

Notes

1. The article links were scraped on December 11, 2023.
2. The article contents were scraped on December 12, 2023.
3. The (micro) F_1 -score represents the harmonious mean between precision and recall (Van Atteveldt et al., 2022): The ratio between the percentage of cases with the correctly predicted labels (*precision*) and the percentage of cases found in the entire (annotated) sample (*recall*).
4. The reasons for not employing this method for the remaining news outlets are threefold: First, it is a time-consuming procedure to retrieve archived web pages through the *Wayback Machine* API. Second, the availability of the public news outlet *NOS* was much higher and consistent compared to its commercial counterparts: Often, no archived data were available in the *Wayback Machine* and if so, it was long past publication. Last, whereas the content of the public news outlet is freely available, most of the contents of the commercial media are behind a paywall and hence not archived by the *Wayback Machine*.
5. For the exact prompt, see script 11 on GitHub: https://github.com/Roeland-Dubel/news_transparency.
6. Only the first five sentences were surveyed because LLMs are demanding to run ($n = 128, 689$). These constitute 19% of all sentences ($n = 680, 900$).
7. In the training data if the name of the news source was present in a sentence, it was censored and replaced by the abstract term “source” to reduce bias toward specific news outlets.
8. For author and sourcing transparency, respectively, the author’s name and textual sourcing feature were dichotomized to being explicit or not. The dummies of internal and family hyperlinking were conjoined to produce a smaller weight for these more peripheral hyperlinking patterns as compared to external hyperlinking (see Chadha & Koliska, 2014).

References

- Andersen, K., Shehata, A., & Andersson, D. (2023). Alternative news orientation and trust in mainstream media: A longitudinal audience perspective. *Digital Journalism*, 11(5), 833–852. <https://doi.org/10.1080/21670811.2021.1986412>
- Bakker, P. (2021, August). *NOS profiteert het meest van online nieuwshonger* [NOS benefits the most from online news hunger]. <https://www.svdj.nl/nieuws/nos-profiteert-het-meest-van-online-nieuwshonger/>
- Bowd, K. (2016). Social media and news media: Building new publics or fragmenting audiences. In M. Griffiths, & K. Barbour (Eds.), *Making publics, making places* (pp. 129–144). Cambridge University Press.

- Buschow, C. (2020). Why do digital native news media fail? An investigation of failure in the early start-up phase. *Media and Communication*, 8(2), 51–61. <https://doi.org/10.17645/mac.v8i2.2677>
- Carlson, M. (2010). Whither anonymity? Journalism and unnamed sources in a changing media environment. In B. Franklin, & M. Carlson (Eds.), *Journalists, sources, and credibility* (pp. 49–60). Routledge.
- 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>
- Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, E., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Narang, S., Mishra, G., Yu, A., . . . Wei, J. (2022). Scaling instruction-finetuned language models. *Arxiv*. <https://doi.org/10.48550/arxiv.2210.11416>
- Craft, S., & Heim, K. (2008). Transparency in journalism: Meanings, merits, and risks. In L. Wilkins, & C. G. Christians (Eds.), *The Routledge 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. <https://doi.org/10.1177/146488490935064>
- Curry, A. L., & Stroud, N. J. (2021). The effects of journalistic transparency on credibility assessments and engagement intentions. *Journalism*, 22(4), 901–918. <https://doi.org/10.1177/1464884919850387>
- De León, E., Vermeer, S., & Trilling, D. (2023). URLs can facilitate machine learning classification of news stories across languages and contexts. *Computational Communication Research*, 5(2), Article 1. <https://doi.org/10.5117/CCR2023.2.4.DELE>
- Delobelle, P., Winters, T., & Berendt, B. (2020). RobBERT: A Dutch RoBERTa-based language model. *Findings of the Association for Computational Linguistics: EMNLP*, 2020, 3255–3265. <https://doi.org/10.18653/v1/2020.findings-emnlp.292>
- Deprez, A., & Raeymaeckers, K. (2012). A longitudinal study of job satisfaction among Flemish professional journalists. *Journalism and Mass Communication*, 2(1), 1–15.
- De Vreese, C. H., Boukes, M., Schuck, A., Vliegthart, R., Bos, L., & Lelkes, Y. (2017). Linking survey and media content data: Opportunities, considerations, and pitfalls. *Communication Methods and Measures*, 11(4), 221–244. <https://doi.org/10.1080/19312458.2017.1380175>
- Doyle, G. (2002). *Media ownership: The economics and politics of convergence and concentration in the UK and European media*. SAGE. <https://doi.org/10.4135/9781446219942>
- Doyle, G. (2013). *Understanding media economics*. SAGE. <https://doi.org/10.4135/9781446279960>
- Esser, F. (1999). Tabloidization of news: A comparative analysis of Anglo-American and German press journalism. *European Journal of Communication*, 14(3), 291–324. <https://doi.org/10.1177/0267323199014003001>
- Fengler, S., Eberwein, T., Alsius, S., Baisnée, O., Bichler, K., Dobek-Ostrowska, B., Evers, H., Glowacki, M., Groenhard, H., Harro-Loit, H., Heikkilä, H., Jempson, M., Karmasin, M., Lauk, E., Lönnendonker, J., Mauri, M., Mazzoleni, G., Pies, J., Porlezza, C., . . . Zambrano, S. V. (2015). How effective is media self-regulation? Results from a comparative survey of European journalists. *European Journal of Communication*, 30(3), 249–266. <https://doi.org/10.1177/0267323114561009>
- Fengler, S., & Speck, D. (2019). Journalism and transparency: A mass communications perspective. In S. Berger, & D. Owetschkin (Eds.), *Contested transparencies, social movements*

- and the public sphere: Multi-disciplinary perspectives* (pp. 119–149). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-23949-7_6
- 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. <https://doi.org/10.1080/14241277.2018.1488257>
- Gravesteijn, E., van Elsas, E., & Gattermann, K. (2024). Biased, not balanced broadcaster! Deconstructing bias accusations toward public service media. *Journalism & Mass Communication Quarterly*. Advance online publication. <https://doi.org/10.1177/10776990241284587>
- Hallin, D. C., & Mancini, P. (2004, April). *Comparing media systems*. <https://doi.org/10.1017/cbo9780511790867>
- Hameleers, M. (2022). “I don’t believe anything they say anymore!” Explaining unanticipated media effects among distrusting citizens. *Media and Communication*, 10(3), 158–168. <https://doi.org/10.17645/mac.v10i3.5307>
- 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>
- Johnson, K. A., & St John, I. I. B. (2021). Transparency in the news: The impact of self-disclosure and process disclosure on the perceived credibility of the journalist, the story, and the organization. *Journalism Studies*, 22(7), 953–970. <https://doi.org/10.1080/1461670X.2021.1910542>
- Karlsson, M. (2010). Rituals of transparency. *Journalism Studies*, 11(4), 535–545. <https://doi.org/10.1080/14616701003638400>
- Karlsson, M. (2020). Dispersing the opacity of transparency in journalism on the appeal of different forms of transparency to the general public. *Journalism Studies*, 21(13), 1795–1814. <https://doi.org/10.1080/1461670X.2020.1790028>
- Karlsson, M. (2021). *Transparency and journalism: A critical appraisal of a disruptive norm*. Routledge.
- Karlsson, M., & Clerwall, C. (2018). Transparency to the rescue? Evaluating citizens' views on transparency tools in journalism. *Journalism Studies*, 19(13), 1923–1933. <https://doi.org/10.1080/1461670X.2018.1492882>
- Karlsson, M., Clerwall, C., & Nord, L. (2016). Do not stand corrected. *Journalism & Mass Communication Quarterly*, 94(1), 148–167. <https://doi.org/10.1177/1077699016654680>
- Karlsson, M., Clerwall, C., & Nord, L. (2017). You ain't seen nothing yet: Transparency's (lack of) effect on source and message credibility. In B. Franklin (Ed.), *The future of journalism: In an age of digital media and economic uncertainty* (pp. 456–466). Routledge.
- 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. <https://doi.org/10.1080/1369118X.2014.984743>
- 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., & Chadha, K. (2018). Transparency in German newsrooms: Diffusion of a new journalistic norm? *Journalism Studies*, 19(16), 2400–2416. <https://doi.org/10.1080/1461670X.2017.1349549>
- 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

- Langer, A. I., & Gruber, J. B. (2021). Political agenda setting in the hybrid media system: Why legacy media still matter a great deal. *The International Journal of Press/Politics*, 26(2), 313–340. <https://doi.org/10.1177/1940161220925023>
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 1, Article 100017. <https://doi.org/10.1016/j.metrad.2023.100017>
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2010). *Practical resources for assessing and reporting intercoder reliability in content analysis research projects*. [PDF]. <https://www.researchgate.net/publication/242785900>
- 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. <https://doi.org/10.1080/1461670X.2020.1790026>
- Louwerse, T., & Van Dijk, R. E. (2022). Reporting the polls: The quality of media reporting of vote intention polls in the Netherlands. *Acta Politica*, 57(3), 548–570. <https://doi.org/10.1057/s41269-021-00208-5>
- 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. (2015). Professional roles in news content: Six dimensions of journalistic role performance. *Journalism Studies*, 16(4), 596–614. <https://doi.org/10.1080/1461670X.2014.922276>
- Newman, N., & Fletcher, R. (2017). *Bias, bullshit and lies: Audience perspectives on low trust in the media*. Social Science Research Network. <https://doi.org/10.2139/ssrn.3173579>
- Nederlandse Omroep Stichting (NOS). (2020). *NOS haalt na aanhoudende bedreigingen logo van satellietwagens* [NOS removes logo from satellite vans after persistent threats]. <https://nos.nl/artikel/2352452-nos-haalt-na-aanhoudende-bedreigingen-logo-van-satellietwagens>
- Oster, J. (2013). Theory and doctrine of media freedom as a legal concept. *Journal of Media Law*, 5, Article 57. <https://doi.org/10.5235/17577632.5.1.57>
- Phillips, A. (2010). Transparency and the new ethics of journalism. *Journalism Practice*, 4(3), 373–382. <https://doi.org/10.1080/17512781003642972>
- Reese, S. D., & Shoemaker, P. J. (2018). A media sociology for the networked public sphere: The hierarchy of influences model. In R. Wei (Ed.), *Advances in foundational mass communication theories* (pp. 96–117). Routledge. <https://doi.org/10.1080/15205436.2016.1174268>
- Reich, Z. (2010). Constrained authors: Bylines and authorship in news reporting. *Journalism*, 11(6), 707–725. <https://doi.org/10.1177/1464884910379708>
- Reimer, J., Häring, M., Loosen, W., Maalej, W., & Merten, L. (2023). Content analyses of user comments in journalism: A systematic literature review spanning communication studies and computer science. *Digital Journalism*, 11(7), 1328–1352. <https://doi.org/10.1080/21670811.2021.1882868>
- 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. <https://doi.org/10.1177/1464884911427803>
- 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) [Transparency in news content: An analysis of methods in Spanish data journalism (2019–2022)]. *Anàlisi*, 68, 97–116. <https://doi.org/10.5565/rev/analisi.3548>

- Rossini, P. (2022). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*, 49(3), 399–425. <https://doi.org/10.1177/0093650220921314>
- Ryfe, D. (2021). The economics of news and the practice of news production. *Journalism Studies*, 22(1), 60–76. <https://doi.org/10.1080/1461670X.2020.1854619>
- Salaverria, R. (2020). Exploring digital native news media. *Media and Communication*, 8(2), 1–4. <https://doi.org/10.17645/mac.v8i2.3044>
- Schonfeld, U., & Shivakumar, N. (2009). Sitemaps: Above and beyond the crawl of duty. In *Proceedings of the 18th international conference on world wide web* (pp. 991–1000). Association for Computing Machinery. <https://doi.org/10.1145/1526709.1526842>
- Skovsgaard, M. (2014). A tabloid mind? Professional values and organizational pressures as explanations of tabloid journalism. *Media, Culture & Society*, 36(2), 200–218. <https://doi.org/10.1177/0163443713515740>
- Steindl, N., Obermaier, M., Fawzi, N., & Lauerer, C. (2024). Explaining media trust among journalists and recipients: Different experiences, different predictors? *Journalism*, 25(8), 1657–1676. <https://doi.org/10.1177/14648849231190698>
- Stenvall, M. (2008). Unnamed sources as rhetorical constructs in news agency reports. *Journalism Studies*, 9(2), 229–243. <https://doi.org/10.1080/14616700701848279>
- Tandoc, E. C., & Thomas, R. J. (2017). Readers value objectivity over transparency. *Newspaper Research Journal*, 38(1), 32–45. <https://doi.org/10.1177/0739532917698446>
- Timmerman, Y., & Bronselaer, A. (2022). Automated monitoring of online news accuracy with change classification models. *Information Processing & Management*, 59(6), Article 103105. <https://doi.org/10.1016/j.ipm.2022.103105>
- Uth, B., Badura, L., & Blöbaum, B. (2021). Perceptions of trustworthiness and risk: How transparency can influence trust in journalism. In B. Blöbaum (Eds.), *Trust and communication: Findings and implications of trust research* (pp. 61–81). 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.
- Van der Wurff, R., & Schönbach, K. (2014). Audience expectations of media accountability in the Netherlands. *Journalism Studies*, 15(2), 121–137. <https://doi.org/10.1080/1461670X.2013.801679>
- Van Es, K., & Poell, T. (2020). Platform imaginaries and Dutch public service media. *Social Media + Society*, 6(2), Article 933289. <https://doi.org/10.1177/205630512093328>
- Vliegienthart, R., Boomgaarden, H. G., & Boumans, J. W. (2011). Changes in political news coverage: Personalization, conflict and negativity in British and Dutch newspapers. In K. Voltmer, & K. Brants (Eds.), *Political communication in postmodern democracy: Challenging the primacy of politics* (pp. 92–110). Springer.
- voor de Media, C. (2023, June). *Digital news report Nederland 2023 (tech. rep.)*. [PDF]. <https://www.cvdM.nl/wp-content/uploads/2023/06/CvdM-DigitalNewsReport-2023.pdf>
- Vos, T. P., & Craft, S. (2017). The discursive construction of journalistic transparency. *Journalism Studies*, 18(12), 1505–1522. <https://doi.org/10.1080/1461670X.2015.1135754>
- Waisbord, S. (2020). Mob censorship: Online harassment of us journalists in times of digital hate and populism. *Digital Journalism*, 8(8), 1030–1046. <https://doi.org/10.1080/21670811.2020.1818111>
- Ward, S. J. (2014). The magical concept of transparency. In S. J. Ward (Ed.), *Ethics for digital journalists* (pp. 45–58). Routledge.

Author Biographies

Roeland Dubèl is a PhD candidate at the Amsterdam School of Communication Research, University of Amsterdam. His research primarily focuses on how journalists and citizens deal with and understand the issue of trustworthiness, as well as the impact of strategies aiming to increase media trust. Furthermore, he studies the effects of news consumption and the workings of news consumption behavior patterns.

Mark Boukes is an Associate Professor in the Department of Communication Science at the University of Amsterdam. His research interests include journalism, media effects, infotainment, and (digital) research methods. In 2022, he received the “Early Career Scholar Award” from the International Communication Association. His research has been recognized with multiple awards (e.g., 5 ICA Top Paper awards) and grants (e.g., NWO Veni, NWO Vidi).

Damian Trilling is a full professor at Vrije Universiteit Amsterdam and holds the Chair for Journalism Studies. Currently, he spends a lot of his time on his ERC-funded project NEWSFLOWS (“Modelling news flows: How feedback loops influence citizens beliefs and shape democracy”). Among other things, he is interested in questions of news use and news exposure in a changing media environment. Methodologically, Damian is interested in automated content analysis and the use of computer science in communication science.