

Piasecki, Stanislaw; Morosoli, Sophie; Helberger, Natali; Naudts, Laurens

Article

AI-generated journalism: Do the transparency provisions in the AI Act give news readers what they hope for?

Internet Policy Review

Provided in Cooperation with:

Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: Piasecki, Stanislaw; Morosoli, Sophie; Helberger, Natali; Naudts, Laurens (2024) : AI-generated journalism: Do the transparency provisions in the AI Act give news readers what they hope for?, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 13, Iss. 4, pp. 1-28, <https://doi.org/10.14763/2024.4.1810>

This Version is available at:

<https://hdl.handle.net/10419/312555>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/de/deed.en>



RESEARCH
ARTICLE



OPEN
ACCESS



PEER
REVIEWED

AI-generated journalism: Do the transparency provisions in the AI Act give news readers what they hope for?

Stanislaw Piasecki *Hitotsubashi University*

Sophie Morosoli *University of Amsterdam*

Natali Helberger *University of Amsterdam*

Laurens Naudts *University of Amsterdam*

DOI: <https://doi.org/10.14763/2024.4.1810>

Published: 23 October 2024

Received: 5 February 2024 **Accepted:** 1 July 2024

Funding: This publication is part of the AI, Media & Democracy Lab (Dutch Research Council project number: NWA.1332.20.009). The authors were also supported by the VolkswagenStiftung Foundation under grant number C.2127.0331.

Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Licence: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. <https://creativecommons.org/licenses/by/3.0/de/deed.en>
Copyright remains with the author(s).

Citation: Piasecki, S. & Morosoli, S. & Helberger, N. & Naudts, L. (2024). AI-generated journalism: Do the transparency provisions in the AI Act give news readers what they hope for?. *Internet Policy Review*, 13(4). <https://doi.org/10.14763/2024.4.1810>

Keywords: Transparency, AI Act, Media, Generative AI, Journalism

Abstract: Issues linked to the increasing presence of AI-generated content in people's lives, and the importance of being able to effectively navigate and distinguish such content, are inherently linked to transparency, a notion that our study focuses on by evaluating Art. 50 of the AI Act. This article is a call for action to take the interests of end users into account when specifying AI Act's transparency requirements. It focuses on a specific use case – media organisations producing text with the help of generative AI. We argue that in its current form, Art. 50 leaves many uncertainties and risks doing too little to protect natural persons from manipulation or to empower them to take protective actions. The article combines documental and survey data analysis (based on a sample representative of the Dutch population) to propose concrete policy and regulatory recommendations on the operationalisation of the AI Act's transparency obligations. Its main objective is to respond to the following question: how to reconcile AI Act's transparency provisions applicable to digital news articles generated by AI with news readers' perceptions of manipulation and empowerment?

Introduction

The Merriam-Webster dictionary has chosen “authentic” as word of the year for 2023, which underlines the increasing presence of AI-generated content in people’s lives, and the importance of being able to effectively navigate and distinguish such content (Italie, 2023). Being able to do so is inherently linked to transparency, a notion that our study focuses on by evaluating Art. 50.1 and Art. 50.4 of the AI Act (AI Act, 2024). The regulation was published on the 12th of July 2024 in the Official Journal of the European Union and is in force since the 1st of August. Transparency is “one of the core values promoted by the EU for the development, deployment, and use of AI systems” (Kiseleva, 2021). The importance of identifying the source of information is also confirmed by various international initiatives such as the Adobe Content Initiative, the objective of which is to promote the “adoption of an open industry standard for content authenticity and provenance” (Adobe, n.d.). This article is a call for action to take the interests of end users into account when specifying AI Act’s Art. 50 transparency requirements. It focuses on a specific use case – media organisations producing text with the help of generative AI. The study argues that in its current form, Art. 50 still leaves many uncertainties and risks doing too little to protect news readers from manipulation or to empower them to take protective actions. Moreover, considering the sector’s particularities, including the value-driven approach of journalists (Bastian et al., 2021), further guidance is needed for the media and policymakers.

Before the AI Act, media professionals were unsure whether they should inform their readers about the use of AI in news production. This will soon become (in certain circumstances) subject to a legal requirement. This article combines documental and survey data analysis (based on a sample representative of the Dutch population) to propose concrete policy and regulatory recommendations on the operationalisation of the AI Act’s transparency obligations (which could be included in a code of conduct and guidelines as explained in Section 2). The findings and suggestions are grounded in empirical evidence and in the expectations of news readers. The main objective is to respond to the following question: how to reconcile the AI Act’s transparency provisions applicable to digital news articles generated by AI with news readers’ perceptions of manipulation and empowerment?

Firstly, this study explores the current legal landscape: how to interpret transparency provisions in Art. 50 of the AI Act in relation to digital news articles? (Section 1)

Secondly, this work evaluates through survey data people’s perceptions of manipu-

lation and empowerment in the context of news articles fully or partly generated by AI systems. In order to do so, it analyses the following topics: how much transparency and agency (as a result of the former) do people want when reading news produced by AI versus news produced by humans? How do they perceive manipulation and empowerment? (Section 2 – “design and sample”, “measurement” and “results” sub-sections.)

Thirdly, Section 3 proposes how to further specify the AI Act with relevant obligations and affordances based on the documental and empirical findings. What kind of regulatory and policy measures could help in reconciling AI Act’s transparency provisions with people’s expectations regarding news consumption? Why is it important for policy makers to meet those expectations? (Section 3)

Finally, the article is briefly concluded by summarising its main findings and calling for action to further specify AI Act’s transparency requirements.

1. The legal landscape related to transparency in the AI Act and digital news articles

Prior to the analysis of the empirical study where we asked survey participants about their transparency expectations and follow-up action preferences in the context of news articles produced by AI, it is important to understand what the law is and how media organisations need to navigate the new transparency requirements. To what extent are media professionals currently obliged to provide information to news readers and to what degree is disclosure left to their own decision? Art. 50 mandates transparency obligations for AI systems regardless of whether they are considered as high risk or not (Almada & Petit, 2023). As the media is not considered high risk in the AI Act, this is the main transparency-related provision applicable to this sector – “by mandating disclosure of the artificial character of the system, the AI Act seeks to close opportunities for impersonation and deception, which can be harmful even if the system itself is not used for a high-risk purpose” (Busuioc et al., 2023, p. 93). This is confirmed in Recital 70, which states that “certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception irrespective of whether they qualify as high-risk or not”. According to the first paragraph of Art. 50.1 AI Act:

Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons

concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. [...].

In terms of AI-generated text, the second paragraph of Art. 50.4 states that:

Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply [...] where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.

Before reflecting on the content of these provisions, it is worth clarifying some of the terminology used in this study and in the AI Act. Firstly, in our article we discuss providers (such as Open AI or Microsoft), natural persons, and deployers (while focusing on the obligations and challenges of the latter) in relation to Art. 50 of the regulation. Although it is still used in a few recitals, in the final version of Art. 50, the term “user” (adopted in all of the previously proposed drafts) has been replaced by “deployer” (the media organisation), which is a welcome development. The former has been criticised – for example, by the Ada Lovelace Institute (Circiumaru, 2022) – as it may lead to confusion on whether the user is actually the end user or the deployer.

Secondly, not all AI content is deceptive. In the context of our work, manipulation is not considered as due to the system just being AI but rather due to simply not knowing that it is AI. People trust and read some journalists more than others. Similarly, they may trust AI (also in relation to a particular AI provider over another) more than humans (or the reverse) and should be able to decide (if they want to decide) what sources they prefer. As a consequence, not disclosing that AI generated an article or that it has been written by a particular journalist would be equally misleading. In addition, although not always deceptive, AI functions differently than human journalists, which further confirms the importance of being informed about its use in news production.

Art. 50.1 imposes information obligations only on providers (contrarily to the European Parliament’s version, it does not mention deployers) and only when the “AI

systems are intended to directly interact with natural persons". The goal of this provision is to inform the "concerned natural persons" about their interaction with AI, that is, providers need to design AI systems in a way that makes it possible to do so. On the one hand, this provision certainly applies to generative AI chatbots (such as ChatGPT) when the content is directly presented to natural persons as a result of their own queries. According to the AI Act, the latter have the right to be informed that such an interaction occurs. On the other hand, it could be argued that in the media context there is an intermediary (the media organisation) before the content arrives to the natural person (news reader) and, as a result, there is indirect interaction causing Art. 50.1 not to apply. However, one could also contend that the AI system was still "intended" to directly provide information to the end user (natural person). Moreover, when the content provided by the deployer is the same as the one originally generated by AI, one could assert that direct interaction still occurs. Nothing was changed except the place where the content was accessed (for example, the media organisation's website instead of the generative AI provider's website). Which interpretation will prevail requires further clarification. If "direct" means direct interaction with the originally produced content, this would signify that the providers' obligation to inform also extends to situations where deployers publish any kind of original AI-generated material, and that the providers' marks, such as watermarks, should not be removed. If "direct" is interpreted to strictly mean that natural persons must generate the content themselves, then deployers (including media organisations) might be allowed to use the AI-generated content without providers' marks in place and not disclose to news readers that they are interacting with AI (or disclose it differently). A yet open question is how Art. 50.1 and Art. 50.4 relate to each other in case of a broad interpretation of "directly interact". Is disclosing the provider's mark enough to also satisfy the Art. 50.4's deployer information obligation? Or would the media organisation need to inform the news reader in addition to the provider's label?

Finally, when a person sees an AI-generated picture, video or text published by a deployer (for example, a media organisation), it will certainly not be "obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect" (as required by Art. 50.1) that they are interacting with AI. It is increasingly difficult for all people to distinguish AI and human content (unless perhaps, for example, a particular media organisation is known for always publishing AI material). This provision is an argument in favour of extending the applicability of the information obligation imposed on providers by Art. 50.1, to situations where the original content is deployed by a third party. Otherwise, the scope of Art 50.1 would be quite narrow. If one interacts directly on Bing's website with their

generative AI systems, it will be certainly “obvious” for most people that they are interacting with AI. However, if a person sees AI-generated content on a media organisation’s YouTube channel or, for example, an AI-generated response on an internet browser following a search query, it would be much harder to contend that they are aware of interacting with AI.

Art. 50.4 explicitly mentions deployers. Its latest version does not only concern deployers of deepfakes (Art. 50.4 paragraph 1) but now also deployers of AI-generated text (Art. 50.4 paragraph 2). For textual content, the media will need to inform natural persons that they are interacting with AI only under specific conditions. Deployers must do so if the article is written “with the purpose of informing the public on matters of public interest” and if it has not “undergone a process of human review or editorial control and a natural or legal person” does not hold “editorial responsibility for the publication of the content” (Art. 50.4). In addition, similarly to Art. 50.1, Art. 50.4 indicates that the content needs to be “artificially generated or manipulated” for the information obligation to apply.

The Art. 50.4 provision leads to many questions. Firstly, what does “artificially generated or manipulated” signify in the context of Art. 50? When is AI used “enough” to mean that information has been manipulated? This article argues that whenever AI writes part of a news article’s substance (fully generated) or is paraphrased by a journalist (manipulated), information about this should be provided. While both texts generated by AI and by journalists may contain mistakes, their causes and prevalence diverge and, as mentioned above, the authors are simply not the same. For this reason, news readers should have a choice in terms of which type of author they prefer. To the contrary, if AI is only used for research purposes, then information provision should not be necessary.

Secondly, it is uncertain how to interpret the “matters of public interest” condition. A fixed definition of the latter is difficult to provide as it differs depending on the context (further legal clarification is needed). If the criteria would be the potential reach of the content, any article may become popular and gain people’s interest, especially through social media. Always accurately predicting which type of news will do so is not possible. This may result in important practical difficulties to effectively implement this provision. If the criteria would be the type of content (for example, sport results versus political news), then this differentiation may be possible. However, it should be based on relevant empirical research rather than assumptions. Our survey data showed (as will be discussed later in this study) that people want information about the source of news regardless of whether the news is controversial or not. Moreover, differentiating between different types of topics

might in practice result in more work for media organisations. The “public interest” factor of a particular type of content has always been difficult to determine and there is not one method to evaluate it (Caple, 2018, p. 10). Always labelling AI-generated news could be a more effective solution (and could lead to more trust from natural persons).

Thirdly, assuming the “matters of public interest” condition is satisfied, for the information provision obligation to apply, the content must also not undergo “a process of human review or editorial control”. What does this signify? If a human fact-checks an article written by AI, is the disclosure obligation lifted? As argued above, the decision on whether to inform or not should not be taken based on the content being deceptive (as the legislator seems to suggest) but rather because news readers should have the right to be informed who produced the content (even if a human or the editor fact-checks it, the source would still be AI). As will be discussed in the next section, this article’s empirical findings confirm that the public expects this information. In addition, is not all content published by media organisations to a certain extent under editorial control? A broad interpretation of this condition would leave only a very narrow scope of application for the provision. For this reason, this work considers that the condition of “human review or editorial control” should be interpreted as meaning that the AI-generated text must be sufficiently transformed (not just paraphrased) by the media professional (as a result, it would not be “artificially generated or manipulated” anymore). All texts written on “matters of public interest” should disclose the use of generative AI systems unless sufficiently transformed under the editor’s control.

Fourthly, the main objective of Art. 50 is to inform natural persons about their interaction with AI (Hacker, 2023). This provision could be interpreted strictly, as signifying that simply informing about the existence of an interaction is sufficient. However, it could also imply that additional information should be disclosed (such as the name of the AI provider’s company). Moreover, transparency can be “an important means to improve procedural rights” if such rights were to be given to news readers (Varošanec, 2022, p. 95). As mentioned in Recital 14 of the AI Act:

Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while *making humans aware that they communicate or interact with an AI system, as well as duly informing* deployers of the capabilities and limitations of that AI system and *affected persons about their rights* (emphasis added).

The AI Act explicitly states that “affected persons” should not only be informed about the fact that they “communicate or interact” with AI but also about rights that they can exercise as a direct consequence of information provision. Information is given to empower people and provide them with more control (not only to inform) (*The Amsterdam Paper*, 2024). A valid question is therefore to ask what information about the interaction should be provided to the natural person when the latter reads news articles generated by AI (to reduce manipulation)? What kind of rights should news readers possess in this context? How to effectively empower news readers to exercise those rights?

There is still much to be discussed in terms of how the AI Act should be implemented and what kind of policy and implementation measures need to be adopted to do so. These questions will be explored in Section 2. This is especially crucial for the media sector where values and building natural persons’ trust in an ethical manner are an essential part of journalistic codes of conduct and work processes – “transparency in the view of the law is not a goal in itself, but a means that is needed to promote a range of very different values” (Gyevnar et al., 2023).

2. Differences between human and AI-generated news – Empirical analysis of news readers’ expectations and their reconciliation with the AI Act

Having identified that transparency plays an important role in the AI Act for the media sector, and thus influences how content will be presented to natural persons in the near future, it is essential to uncover how people react to transparency cues in the context of news content. Central questions connected to this arise: would news readers feel manipulated if they were not informed about the fact that an article was produced by AI? What do people do with news content once they have been confronted with a transparency cue? How would their interest in being able to exercise control over news content differ for human and AI written content? For instance, would they like to have additional information about news production or have the option to filter news articles from a certain source? The following section empirically tackles these challenges. As mentioned by Haresamudram, “user-centred research on AI transparency remains limited” (Haresamudram et al., 2023, p. 99). This contribution should be seen as only one element of a larger and needed discussion on how to operationalise and specify AI Act’s Art. 50 in the media sector. The empowerment-related issues we tackled are not exhaustive and there might be other relevant topics requiring empirical research.

Design and sample

In this study, we focus on two transparency cues (written by a human journalist vs written by AI) and two news topics (neutral vs politicised). By the means of a 2 x 2 survey experiment, we presented our participants with different news headlines which were created for the purpose of this study. One concerning a private donation made to the national museum (neutral) and one concerning an increase of immigrants in the Netherlands, especially because of the war in Ukraine (politicised) (see Appendix A for the used stimuli). We are focusing on two different types of issues to control whether the type influences how news readers react to transparency labels or if they react consistently across topics. We chose two recent events with which participants might be familiar and the layout resembles a generic news website without any source indication to avoid priming.

Each of the headlines received a transparency cue. The first one stated that the article was written by a human journalist and the second one stated that the article was written by artificial intelligence. In total 227 respondents participated in the experiment and were recruited by a panel company based in the Netherlands. The experiment was part of a larger survey ($N = 1448$), where 227 of the 1448 participants took part in the additional experiment after completing the survey. The polling company Bilendi recruited the sample based on country-specific census data and specific quotas on age, gender, and education. This resulted in a representative sample of the Dutch population regarding age ($M = 50.46$, $SD = 17.38$), gender (female = 52.9%, male = 47.1%), and education (lower = 22.1%, moderate = 50.2%, higher = 27.7%). In the following, we explore how news readers would react to news headlines with different transparency labels and we focus on two central concepts, which are key when exploring the effects of transparency cues connected to the AI Act: perceived *manipulation* and individual *empowerment*.

All in all, this empirical analysis aims to shine a light on the effects of transparency labels (news articles marked as written by a human versus produced by AI) on (1) people's follow up actions regarding news headlines as well as on (2) their perceptions regarding manipulation and (3) empowerment. We are interested in whether natural persons would feel manipulated if they were not informed about the use of AI in news production and whether people wish to have agency over news content produced by AI. Additionally, this study aims to identify possible group differences between the two different transparency cues and whether the topic of the news article plays a role in this relationship.

Measurement

In a first step we asked our respondents what they would normally do with the news headline they have been exposed to. We measured this follow-up behaviour through six items on a seven-point scale (1 = completely disagree, 7 = completely agree). The scale included actions like sharing the article online ($M = 2.08$, $SD = 1.71$), sharing the article with friends ($M = 2.29$, $SD = 1.79$), talking about the article with peers ($M = 3.11$, $SD = 1.92$), the willingness to pay for this article ($M = 1.91$, $SD = 1.59$), reporting the article as misleading ($M = 2.59$, $SD = 1.88$), and no action ($M = 5.08$, $SD = 2.04$).

Secondly, two central concepts are of interest in this section: manipulation and empowerment behaviours connected to the use of AI in journalism. It is important to note that within the larger survey we provided the participants with a definition of AI.¹ Hence, all the respondents had the same level of understanding when answering the questions connected to the experiment. Manipulation was measured on a 7-point scale (1 = completely disagree, 7 = completely agree). Participants were asked to what extent they agree with the following statements: (1) “I would feel manipulated if the news I read was written by AI instead of a human journalist without me being informed about it”, $M = 4.92$, $SD = 1.91$. And (2), “I think the article written by this source is more likely to deceive and mislead people”, $M = 4.07$, $SD = 1.70$. Within this study, the notion of “empowerment” broadly captures participants’ agency over their interactions with and consumption of (AI-generated) news content. To measure empowerment, we asked participants to what extent they agree or disagree with seven different behaviours that reflect, in a non-exhaustive fashion, an exercise of control over one’s news environment on a seven-point Likert scale (1 = completely disagree, 7 = completely agree). Examples of such actions include, “I want to be able to filter news content that has been written by this source” ($M = 3.85$, $SD = 1.85$), “I want additional information about the news production and distribution” ($M = 3.48$, $SD = 1.93$), or “I want to be able to report the article” ($M = 3.85$, $SD = 1.98$). The full scale can be found in Appendix B. As previously mentioned, within this context, transparency provisions often hold an instrumental function as the information they offer can facilitate news readers in their exercise of agency.

1. Definition of AI: “artificial intelligence” means a software or computer that is developed based on, for example, machine learning. It has the ability to generate outputs such as content, predictions, recommendations, or decisions, which is normally linked to human intelligence.

Results

To explore group differences regarding respondents' follow-up behaviour, we performed multiple ANOVAs (Analysis of Variance) to test whether the means of the four different experimental groups significantly differ from each other. For the majority of the follow-up actions we do not find any significant group differences, meaning that news readers behave similarly across experimental groups (see Table 1). However, we do find that the groups are significantly different from each other regarding the action "talking to friends and family about the article" ($F(3,203) = 3.56, p = 0.02$). To analyse where exactly the significant group differences lie, we performed a post-hoc comparison using the Tukey HSD test. The test indicates that the mean score for the neutral article written by a human journalist ($M = 2.50, SD = 1.78$) was significantly different from the score for the polarised article written by a human ($M = 3.59, SD = 1.97$). People would be more likely to talk with their friends about the article on immigration than about the article about a private donation made to the museum. This indicates that the topic of the article drives this behaviour and not the transparency label.

TABLE 1: One-way ANOVA for individual follow-up behaviour

	DF	SUM SQ	MEAN SQ	F-VALUE	P-VALUE
<i>SHARE SOCIAL MEDIA</i>	3	17.74	5.91	2.06	0.11
RESIDUALS	203	581.87	2.87	-	-
<i>SHARE FRIENDS</i>	3	12.92	4.31	1.35	0.26
RESIDUALS	203	649.37	3.12	-	-
<i>TALK TO PEERS</i>	3	34.62	11.54	3.56	0.02
RESIDUALS	203	723.04	3.56	-	-
<i>WILLINGNESS TO PAY</i>	3	7.73	2.57	1.02	0.39
RESIDUALS	203	513.53	2.53	-	-
<i>REPORT ARTICLE</i>	3	1.02	0.34	0.09	0.96
RESIDUALS	203	726.89	3.58	-	-
<i>NO ACTION</i>	3	1.87	0.62	0.15	0.93
RESIDUALS	203	857.75	4.22	-	-

Considering that we did not find meaningful group differences, it is fruitful to investigate overall trends among the respondents. Even though the majority of respondents did disagree with the six follow-up actions, we do find that some actions were indicated by over a quarter of the participants. Overall, 16%² of our

sample indicated that they would share the headline they just saw on social media and 16.8% would share it with friends. With the highest amount of agreement, 27.6% of the participants would talk about the articles with their peers whereas only 11.1% would be willing to pay for this type of content. Next to talking about the headlines, the willingness to report the headline as misleading was the second most popular follow-up behaviour (17.8% agreement). Lastly, the majority of news readers would not do anything with the presented headlines (66.7%).

Equally, by means of an ANOVA, we aim to identify differences between the four experimental groups regarding perceived manipulation and empowerment behaviours. Interestingly, the findings reveal no significant group differences for both core concepts (see Table 2). Thus, the two different transparency labels (AI written vs human written) and different news issues (neutral vs politicised) did not seem to have an impact on how news readers perceive personal manipulation or how they would act towards the presented headline. This result offers a new avenue of looking at our data as a bigger picture.

TABLE 2: One-way ANOVA for manipulation and empowerment

	DF	SUM SQ	MEAN SQ	F-VALUE	P-VALUE
<i>MANIPULATION 1</i>	3	18.06	6.02	1.67	0.17
RESIDUALS	203	730.70	3.60	-	-
<i>MANIPULATION 2</i>	3	11.03	3.68	1.33	0.27
RESIDUALS	203	562.02	2.77	-	-
<i>EMPOWERMENT 1</i>	3	2.12	0.71	0.23	0.87
RESIDUALS	203	621.26	3.06	-	-
<i>EMPOWERMENT 2</i>	3	1.73	0.58	0.17	0.92
RESIDUALS	203	705.92	3.48	-	-
<i>EMPOWERMENT 3</i>	3	12.00	4.00	1.185	0.32
RESIDUALS	203	689.74	3.40	-	-
<i>EMPOWERMENT 4</i>	3	9.51	3.17	0.87	0.46
RESIDUALS	203	737.60	3.63	-	-
<i>EMPOWERMENT 5</i>	3	9.79	3.26	0.82	0.48
RESIDUALS	203	804.28	3.96	-	-
<i>EMPOWERMENT 6</i>	3	4.62	1.54	0.36	0.78

2. Percentages of agreement represent participants who answered 5 or higher on a seven-point Likert scale. Percentages of disagreement represent participants who answered 3 or lower.

	DF	SUM SQ	MEAN SQ	F-VALUE	P-VALUE
RESIDUALS	203	860.59	4.24	-	-
<i>EMPOWERMENT 7</i>	3	12.78	4.26	1.16	0.32
RESIDUALS	203	743.67	3.66	-	-

Overall, we find that 60.3%³ of all participants would feel manipulated if they were not informed about the fact that the article they read was written by artificial intelligence instead of a human journalist. Only 21.5% disagreed with this statement. This suggests a striking need for transparency no matter the type of issue presented in the article. Additionally, 37% of all respondents combined indicated that the article written by the respective source is more likely to deceive and mislead people. News readers seem sceptical and suspicious of a human journalist and AI as a news source regardless if they are confronted with a neutral or politicised topic. Thus, it does not seem to matter whether they received the human or AI transparency cue – they distrust the source in any case. This result points towards findings of previous studies, where trust in news has been declining over the past years and media cynicism increased (Newman et al., 2023; Quiring et al., 2021). They are also in line with this article’s argument in Section 1 that both AI and human content can be deceptive and that information obligations regarding news content generated by AI should not be simply based on the contention that AI may lead to misinformation. Or put differently, simply informing people about the fact that content has been produced by AI does not give them any hints whether it is deceptive or trustworthy.⁴

Altogether, we have observed that citizens would feel manipulated if they were not informed about the fact that news content has been created by AI and that many of them believe that the source – human journalist or artificial intelligence – of the article is deceiving or misleading. These results imply the unmistakable desire of news readers to be informed about the fact that content has been generated by AI. Further, we found that the most common immediate follow-up behav-

3. Percentages of agreement represent participants who answered 5 or higher on a seven-point Likert scale. Percentages of disagreement represent participants who answered 3 or lower.
4. To capture the concept of mistrust more encompassing, we performed two additional ANOVAs (see Appendix C). The results show that two experimental groups differ from each other when looking at the trust in the source and trust in the information of the article. For both trust measurements, there is a significant difference between the politicised topic written by AI and the politicised topic written by a human journalist. News readers seem to distrust the AI written headline more than the human written one – but only if the headline is politicised. Hence, overall only two groups differ from each other, meaning that the remaining groups display an equal amount of distrust no matter the source and topic.

your after being exposed to the headlines was talking about the articles with peers and the willingness to report the article as misleading. Knowing this, there might be certain needs of natural persons to feel empowered to counter this feeling of manipulation or deception. As previously described, we did not find statistically significant differences between group means regarding empowerment. However, we can identify patterns across groups, which are of high relevance for specifying transparency labels based on empowerment behaviours.

First and foremost, our data shows that news readers reported on average moderately high agreement (i.e. above 3 – the midpoint of the scale) with two actions. Respondents want to be able to filter news that has been written by the respective source ($M = 3.85$, $SD = 1.85$), and they want to have the option to report the article ($M = 3.90$, $SD = 1.98$). Next, 38.1% of our participants would like to be able to complain to the news organisation, followed by the option to inform about biases that they see in the news produced by this source (29.9%). Less popular were the options to get additional information about news production and distribution (45.4% disagreed) and to continue consuming news from that source (47.4% disagreed). However, still 29.9% of the participants would like more information and 27.5% indicated that they would continue reading the news. Lastly, only 24.3% would like to have the option to talk to the editor, which was the least wanted empowerment behaviour. This might be the case because of the connected effort that comes with this option. Hence, we conclude that filtering out content from certain sources, reporting articles, and complaining to the news company are the three most desired actions when it comes to empowerment behaviours. However, even for the least desired behaviour, almost one-quarter of respondents remain a significant number of news readers.

It is evident that this empirical study does not come without certain limitations. For instance, can the results not be generalised beyond the Dutch context, and we encourage scholars to study manipulation and empowerment behaviours in a comparative manner beyond the Western and democratic context. Furthermore, we rely on self-reporting measures even though we have an experimental setting. This means that the results could be skewed towards social desirability. We recommend future research study this issue under more realistic circumstances by for example recreating a news website and tracking participants follow up behaviour.

3. Discussion

Thus far, we have discussed the legal landscape of transparency in the context of the AI Act and digital news articles. We have also analysed empirical findings re-

garding people's follow up actions when they read news headlines, as well as their perceptions of manipulation and empowerment. Moving forward, we put the findings into a bigger picture.

Based on the results of this study, we identified the need for transparency to counteract the feeling of manipulation. In other words: news readers want to know whether an article has been written by AI or not, irrespective of whether that content is liable to influence the broader public opinion and irrespective of their beliefs of whether AI-generated content is more or less accurate, trustworthy, value-oriented, or has their best interests in mind. Insofar, Art. 50 AI Act is important from the perspective of news readers. However, our research also shows the current limitation in Art. 50.4, that exempts content that has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication, is contrary to the interests of the audience. People would feel manipulated if not informed about the synthetic origin of the content, even if it has undergone editorial review (unless, as suggested in Section 1, the editorial review condition would be interpreted as meaning that the text needs to be modified to the extent that it is not possible to consider its content as "artificially generated or manipulated" anymore). Moreover, if no information is provided, people would not be able to exercise any of the follow-up actions in relation to the empowerment behaviours measured through this study's survey data (as mentioned in the previous section, an important number of news readers would want to do so).

Importantly, our research also finds that simply informing people about the fact that a text has been automatically generated or manipulated by AI gives them little cues about how to interpret and assess the article. This may also explain why we saw no significant differences in their intended follow-up behaviour. Indeed, our results suggest that being informed about the fact that a headline was generated by AI or written by a human did not significantly affect their willingness to share, continue reading or even pay for the content. Contrary to our findings, Altay and Gilardi (2023) observed that people were less willing to share news headlines labelled as AI-generated. The authors explained this outcome with a decrease in accuracy perception. AI-generated headlines were perceived less accurate and thus news readers were less willing to share them. Furthermore, the AI-generated transparency labels did not reduce the trust in news or journalists and the authors could not find any significant group differences regarding this relationship (Altay & Gilardi, 2023).

These inconclusive findings raise a more fundamental question about the trans-

parency obligations in Art. 50.4 paragraph 2 of the AI Act: what exactly is and can be the goal of this provision? If the goal is to empower people to make informed decisions on how trustworthy or qualitative synthetic content is, Art. 50 of the AI Act is likely to fail that goal. For the same reason, the transparency provisions are unlikely to solve the problem of popular misconceptions, misleading imaginaries, and folk stories about AI (Jasnoff, 2015; Cave & Dihal, 2019). Simply informing natural persons that content has been AI generated or manipulated does not convey enough information to decide whether synthetic content is trustworthy and leaves it to newsreaders to draw their own conclusions based on whatever their ideas or imaginaries of AI are, or their level of knowledge. An interesting question for further research could be whether simple labels can not even reinforce persistent imaginaries of machine autonomy and AI taking over ever larger parts of society. For the same reason, the information obligation in Art. 50 will not be particularly useful as a tool to fight disinformation (as hinted at in Recital 70 AI Act). In that regard, this second text-based condition of Art. 50.4 differs from the deepfake condition in the first part of the paragraph. Deepfakes are defined as “image, audio or video content that appreciably resembles existing persons, places or events and would falsely appear to a person to be authentic” (Recital 70b). Here the goal of transparency is to warn end users that the content mimics real persons or places but is not authentic. In contrast, the second paragraph of Art. 50.4 AI Act only requires that text must be intended to inform the public, which only says something about the function of the text, not whether the events or facts it describes are authentic or inauthentic. The objective of Art. 50.4 remains unclear.

Is the goal of Art. 50.4 paragraph 2 AI Act then to empower and enable natural persons to exercise their rights against synthetically generated or manipulated content? Again, the simple fact that a piece of text is synthetic does not give enough information to enable natural persons to assess if their rights are affected, what those rights are, and how they could be exercised. Simply producing synthetic content is not against the law, neither does the AI Act in its current form give people any rights to intervene. Art. 85 AI Act foresees a right to lodge a complaint with a market surveillance authority but only if a user has reasons to believe that the provisions of the regulation, such as the transparency obligations, have been infringed. Further below we will discuss what rights or entitlements news readers *would* like to see.

This leaves dignitarian arguments as the main goal that Art. 50.4 paragraph 2 may serve, namely that natural persons have a moral right to be informed if they are subjected to AI-generated or manipulated content. The question is then: why

would the transparency obligations only apply to textual content that is intended to inform on public matters, or is excluded for content under editorial control? In conclusion: the transparency provisions in Art. 50.1 and Art. 50.4 of the AI Act are important and necessary, also from the perspective of natural persons, but simply not yet sufficiently thought through.

Worrying is the finding that people mistrust⁵ news sources in general – regardless of whether an article was written by an AI or a human journalist. The media is generally believed to have an important role in providing trustworthy and quality information as an antidote to human or automated disinformation. The fact that a significant portion of respondents were not convinced that media content is trustworthy, will give them the information they need, and have their interest in mind must be food for thought for the media sector.

Next to wanting to be informed about the fact that a piece of content has been generated by, or manipulated with AI, our study also shows that people want more than transparency: they want a choice and be able to exercise a certain level of agency. By agency, this article means “the exercise or manifestation of one’s capacity to take actions, or ‘do things’” (Andrada et al., 2023, p. 1327), which we also view as a component of news readers’ empowerment. The study by Andrada and colleagues aligns with our findings that simply informing end users does not increase their agency and that relevant mechanisms might need to be implemented to do so: “while we assessed only whether being exposed to a notice increased individuals’ agency and not the reasons behind, the lack of agency could be explained by the fact that the participants could not influence the interaction with AI (e.g. opting out to AI powered interactions)” (De Andrade et al., 2023, p. 23).

Transparency holds both intrinsic and instrumental value. To realise its instrumental function however, we argue, disclosure obligations should inform and direct news readers on how to regain and exercise agency over their news environment. The information that should be provided then, depends upon the forms of agency we wish to secure (through the law), as well as the risks we want to guard and empower people against. As digital landscapes have become characterised by great asymmetries in power and knowledge over technology (Helberger et al., 2021, 2022), a rich body of literature exists on (the limitations of) citizen empowerment, including the role transparency, as well as technical solutions and the law, can

5. In the context of this study, we consider and measure mistrust based on the additional analysis regarding trust in the information and trust in source (see Appendix C), the low willingness to pay for news content, and the strong feeling of manipulation across the stimuli. We are aware that different measurements for mistrust exist; however, we believe that we capture the concept extensively.

play therein (see among others: Micklitz et al., 2017; van Ooijen & Vrabec, 2019; Jablonowska & Palka, 2019; Felzmann et al., 2019, 2020; Lippi et al., 2020; El Ali et al., 2024). To further substantiate transparency's agency-related function, we asked survey participants what forms of agency they would find most desirable. Three actions stood out: being able to filter news that has been written by an AI or a human; being able to report an article, respectively to complain to a news organisation; and being able to flag biases that they see in news produced by a particular source. Filtering articles from a specific source could be seen as the choice to omit articles and curate people's news feeds so that they do not display articles from this source, and that persons have the power to personalise their news website or application. The other two interventions are more targeted at directly interacting with the news organisation, whereby it is not clear whether to contest the use of AI specifically, or more generally to express a wish for more interaction and responsiveness on the side of the media. The ability to complain recalls earlier demands from the European Parliament to give people the possibility to object against the application of AI systems. Within the latter, Art. 50.1 second paragraph (of the EP's original version) stated that information "shall also include which functions are AI enabled, if there is human oversight, and who is responsible for the decision-making process, as well as the existing rights and processes" that allow to object against the application of AI systems, "to seek judicial redress against decisions taken by or harm caused by AI systems", including the right to seek an explanation.

Arguably, to give natural persons sufficient cues to be able to assess the quality and trustworthiness of an AI-generated or manipulated text, they would need additional information, such as the capabilities or limitations of the artificial text generator, or whether the text has been subject to editorial control.⁶ Providing detailed information about where and how AI has been applied in the news production and who is responsible for the decision-making process can help end users to better understand the abilities but also the limits of AI. For instance, informing readers if artificial intelligence has been used to create the headline or the teaser of an article (where) and if the media company uses their own AI system or if they rely on an existing one (how). Even though more extensive information obligations were not on the top of the list of the most preferred empowerment actions, still almost a third of our respondents wanted to have more information than a simple

6. See for example Open AI's Publication and Sharing policy, suggesting to include a statement along the lines of: "the author generated this text in part with GPT-3, OpenAI's large-scale language-generation model. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication" (OpenAI, 2022).

cue whether or not a text has been produced by AI.

Implications

The findings from this study are relevant for both policymakers and the media. For media organisations they suggest that being transparent to news readers about the fact that a piece of content has been AI generated or manipulated is paramount, even if the AI Act might exempt the (editorial) media from that obligation. Telling people that content has been AI generated can have consequences: some might stop reading the content, filter it out (if offered the option), or be less willing to pay for it, but the consequences of not telling could even be worse and threaten the already fragile trust relationship: people would feel manipulated. To some extent, the findings of this study are also encouraging: informing end users that a piece of content has been AI generated seems to have no direct effect on the trustworthiness or perceptions of quality of the content itself, or their willingness to continue reading or sharing the content. The study also confirms earlier findings that natural persons would value more choice and ability to exercise agency and voice (Monzer et al., 2020). While there can be clear economic, strategic, pragmatic, and organisational reasons against offering people more choice, doing so could also present media organisations with an opportunity to mend their relationship with the audience and profile themselves in a role as stewards or moral compass in an increasingly complicated and hostile digital environment. Lewis and colleagues, for example, introduced the concept of reciprocal journalism and argued that “by more readily acknowledging and reciprocating the input of audiences, and by fostering spaces for audiences to reciprocate with each other, journalists can begin to fulfil their normative purpose as stewards of the communities they serve” (Lewis et al., 2014, pp. 236-237). Put differently, media organisations do have an opportunity here to do more than playing by the rules. Instead, they can use transparency and explainability as a means to differentiate themselves from large technology corporations, whose main interest is to get as many people as possible “hooked” on their services, and explain how they make sure end users can trust their content.

For policymakers, this article’s findings are food for thought too. From the point of view of natural persons, the actual effectiveness of the current transparency obligation is questionable. As suggested earlier already by the European Parliament, transparency without agency is not much more than a label. And in the case of Art. 50 AI Act in its present form, that label does not convey enough meaningful information.

The AI Act does not discuss delegated acts or implementing acts in relation to Title IV on the “Transparency Obligations for Providers and Deployers of Certain AI Systems”. As a result, additional mandatory requirements specifying Art. 50 cannot be adopted based on the regulation’s provisions. However, Art. 95 mentions the possibility of developing codes of conduct:

The Commission and the Member States shall facilitate the drawing up of codes of conduct concerning the voluntary application, including by deployers, of specific requirements to all AI systems, on the basis of clear objectives and key performance indicators to measure the achievement of those objectives.

In addition, Art. 96 AI Act indicates that the Commission shall develop guidelines on the practical implementation of the regulation, including on “the practical implementation of transparency obligations laid down in Art. 50”. To do so, the Commission will need more clarity of what the goal of Art. 50 AI Act is: to inform for the sake of informing (dignitarian arguments), to warn, to empower? Depending on what the answer is, informing natural persons will need to take different forms and should be accompanied by different empowerment measures – the ability to flag content, to filter content out, to complain, but also an obligation on the side of developers and deployers to explain what the actual implications are if a piece of content has been AI generated or manipulated and why it should be trusted.

Directions for future research

Future legal-empirical research can help to design AI transparency labels that respond to the information needs of an audience with different levels of AI literacy and that convey the information that is relevant to empower newsreaders. Other potential avenues for future research that our study raises concern the possible need to differentiate between AI transparency labels for the news media and social media, the way transparency labels affect actual user behaviour, and could potentially result in less desirable side-effects, like re-enforcing socio-technical imaginaries and folk stories around AI.

Another issue that requires further research (and that should be considered in guidelines or a code of conduct) is how relevant information should be communicated. A discussion emerged on watermarking as a potential tool to achieve compliance with the Art. 50.1 information obligations of providers – “implementation of these obligations [transparency requirements of the AI Act] will likely require use of watermarking techniques” (Madiaga, 2023). In terms of both provider and

deployer-related transparency, Art. 50.5 states that “the information referred to in paragraphs 1 to 4 shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure”. It also adds that “the information shall respect the applicable accessibility requirements”. This paper argues that to comply with these requirements, inspiration could be drawn from other fields, such as data protection law, also by understanding the latter’s limitations (Prifti et al., 2023). Indeed, transparency-related conditions – in relation to information provision – have been widely debated in the context of the General Data Protection Regulation (GDPR) (Busuioc et al., 2023; Naudts et al., 2022). For example, intelligibility is one of the GDPR transparency obligations (and was also a requirement in the European Parliament’s version of Art. 50). As the Court of Justice of the European Union stated in the *Kásler* case, the intelligibility and plain language conditions “cannot... be reduced merely to their being formally and grammatically intelligible”, but rather need to be comprehended in a “broad sense” taking into account an “average consumer, who is reasonably well informed and reasonably observant and circumspect” (Case C-26/13). This is in-line with the “reasonably well-informed, observant and circumspect” terminology used in Art. 50 of the AI Act. As a result, a journalist or a media organisation should first define their target audience and establish the average audience member’s level of understanding. However, in practice, it would be difficult to ascertain who is accessing a particular website. For this reason, this article argues to always assume that a vulnerable person (such as a child or vulnerable adult) could interact with a news article and to adapt communication mechanisms by default to such circumstances (Piasecki & Chen, 2022). This would make information clearer for everyone. This is also in-line with Art. 50.5, which mentions “applicable accessibility requirements” and EU regulation more broadly. For example, the GDPR transparency principle mandates organisations to adopt special measures when they provide information to vulnerable people (Piasecki, 2023, p. 13). It is beyond the scope of this research to analyse communication mechanisms in detail, and they would be dependent on the nature of the information provided. This article simply wants to underline the importance of taking this (often overlooked) aspect of information provision into account.

Finally, this article does not discuss the relationship between the AI Act’s transparency provisions and consumer law. Whether the new AI regulation complements or undermines the latter, as well as more generally how to interpret Art. 50 in light of current consumer law provisions, requires further study.

Conclusion

Transparency is no panacea, also not in enabling end users to evaluate the safety and trustworthiness of AI-generated content. Yes, people do want to know whether they are exposed to human-written or synthetic content, and they would feel manipulated if that information was withheld from them. This alone is a strong reason to justify the inclusion of the transparency obligations in Art. 50. And yet, it is also important to realise that in its current form, the transparency obligation for AI-generated text is too narrow to live up to the expectations of the audience in being informed and too limited to convey much more cues than that a content has been artificially generated. In ten years from now, synthetic content may very well constitute a large share of content on the internet – what then will be the added value of the provision? Instead, the transparency obligations in Art. 50 should be the beginning of a conversation on how to make transparency meaningful for the audience.

References

- Adobe. (n.d.). *Content authenticity initiative*. Content Authenticity. <https://contentauthenticity.org>
- Almada, M., & Petit, N. (2023). *The EU AI Act: A medley of product safety and fundamental rights?* (Working Paper No. RSC 2023/59). Robert Schuman Centre for Advanced Studies. https://cadmus.eu.i.eu/bitstream/handle/1814/75982/RSC_WP_2023_59.pdf?sequence=1&isAllowed=y
- Altay, S., & Gilardi, F. (2023). *People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation*. PsyArXiv. <https://doi.org/10.31234/osf.io/83k9r>
- Andrada, G., Clowes, R. W., & Smart, P. R. (2023). Varieties of transparency: Exploring agency within AI systems. *AI & Society*, 38(4), 1321–1331. <https://doi.org/10.1007/s00146-021-01326-6>
- Bastian, M., Helberger, N., & Makhortykh, M. (2021). Safeguarding the journalistic DNA: Attitudes towards the role of professional values in algorithmic news recommender designs. *Digital Journalism*, 9(6), 835–863. <https://doi.org/10.1080/21670811.2021.1912622>
- Busuioc, M., Curtin, D., & Almada, M. (2023). Reclaiming transparency: Contesting the logics of secrecy within the AI Act. *European Law Open*, 2(1), 79–105. <https://doi.org/10.1017/elo.2022.47>
- Case C-26/13. (n.d.). *Judgment of the Court (Fourth Chamber), 30 April 2014: Árpád Kásler and Hajnalka Káslerné Rábai v OTP Jelzálogbank Zrt*. The Court of Justice of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62013CJ0026>
- Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence*, 1, 74–78. <https://doi.org/10.1038/s42256-019-0020-9>
- Circiumaru, A. (2022). *People, risk and the unique requirements of AI: 18 recommendations to strengthen the EU AI Act* [Policy briefing]. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act/>

De Andrade, N., Galindo, L., Zarra, A., Heal, J., & Rom, S. (2023). *Towards informed AI interactions: Assessing the impact of notification styles on user awareness and trust* (Open Loop Artificial Intelligence Act: A Policy Prototyping Experiment Operationalizing the Requirements for AI Systems – Part III, pp. 1–27). Open Loop. https://openloop.org/wp-content/uploads/2023/06/AI_Act_Towards_Informed_AI_Interactions.pdf

El Ali, A., Venkatraj, K. P., Morosoli, S., Naudts, L., Helberger, N., & Cesar, P. (2024). Transparent AI disclosure obligations: Who, what, when, where, why, how. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–11. <https://doi.org/10.1145/3613905.3650750>

EY. (2023). *Political agreement reached on the EU Artificial Intelligence Act* (pp. 1–15) [Report]. https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/ai/ey-eu-ai-act-political-agreement-overview-10-december-2023.pdf

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951719860542>

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, 26, 3333–3361. <https://doi.org/10.1007/s11948-020-00276-4>

Gyevnar, B., Ferguson, N., & Schafer, B. (2023). *Bridging the transparency gap: What can explainable AI learn from the AI Act?* (Version 5). arXiv. <https://doi.org/10.48550/ARXIV.2302.10766>

Hacker, P. (2023). *AI regulation in Europe: From the AI Act to future regulatory challenges*. arXiv. <https://doi.org/10.48550/arXiv.2310.04072>

Haresamudram, K., Larsson, S., & Heintz, F. (2023). Three levels of AI transparency. *Computer*, 56(2), 93–100. <https://doi.org/10.1109/MC.2022.3213181>

Helberger, N., Lynskey, O., Micklitz, H.-W., Rott, P., Sax, M., & Strycharz, J. (2021). *EU consumer protection 2.0: Structural asymmetries in digital consumer markets* [Report]. BEUC The European Consumer Organisation. https://www.beuc.eu/sites/default/files/publications/beuc-x-2021-018_eu_consumer_protection_2.0.pdf

Helberger, N., Sax, M., Strycharz, J., & Micklitz, H.-W. (2022). Choice architectures in the digital economy: Towards a new understanding of digital vulnerability. *Journal of Consumer Policy*, 45(2), 175–200. <https://doi.org/10.1007/s10603-021-09500-5>

Italie, L. (2023, November 27). What's Merriam-Webster's word of the year for 2023? Hint: Be true to yourself. *AP News*. <https://apnews.com/article/merriam-webster-word-of-year-2023-a9fea610cb32ed913bc15533acab71cc>

Jabłonowska, A., & Pałka, P. (2019). EU consumer law and artificial intelligence. In L. De Almeida, M. C. Gamito, M. Durovic, & K. P. Purnhagen (Eds.), *The transformation of economic law: Essays in honour of Hans-W. Micklitz* (pp. 91–112). Hart Publishing. <https://doi.org/10.5040/9781509932610>

Jasnoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasnoff & S.-H. Kim (Eds.), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power* (pp. 1–33). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226276663.03.0001>

Kiseleva, A. (2021, July 29). Making AI's transparency transparent: Notes on the EU proposal for the AI Act. *European Law Blog*. <https://web.archive.org/web/20211006121316/https://europeanlawblog.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/>

- Lewis, S. C., Holton, A. E., & Coddington, M. (2014). Reciprocal journalism: A concept of mutual exchange between journalists and audiences. *Journalism Practice*, 8(2), 229–241. <https://doi.org/10.1080/17512786.2013.859840>
- Lippi, M., Contissa, G., Jabłonowska, A., Lagioia, F., Micklitz, H.-W., Palka, P., Sartor, G., & Torroni, P. (2020). The force awakens: Artificial intelligence for consumer law. *Journal of Artificial Intelligence Research*, 67, 169–190. <https://doi.org/10.1613/jair.1.11519>
- Madiega, T. (2023). *Generative AI and watermarking* [Briefing]. European Parliamentary Research Service. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)757583](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)757583)
- Micklitz, H.-W., Palka, P., & Panagis, Y. (2017). The empire strikes back: Digital control of unfair terms of online services. *Journal of Consumer Policy*, 40(3), 367–388. <https://doi.org/10.1007/s10603-017-9353-0>
- Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User perspectives on the news personalisation process: Agency, trust and utility as building blocks. *Digital Journalism*, 8(9), 1142–1162. <https://doi.org/10.1080/21670811.2020.1773291>
- Naudts, L., Dewitte, P., & Ausloos, J. (2022). Meaningful transparency through data rights: A multidimensional analysis. In E. Kosta, R. Leenes, & I. Kamara (Eds.), *Research handbook on EU data protection law* (pp. 530–571). Edward Elgar Publishing. <https://doi.org/10.4337/9781800371682.00030>
- Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., & Nielsen, R. K. (2023). *Digital news report 2023* [Report]. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf
- News values and newsworthiness. (2018). In *Oxford research encyclopedia of communication*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190228613.013.850>
- OpenAI. (2022). *Sharing & publication policy* [Policy overview]. <https://openai.com/policies/sharing-publication-policy>
- Piasecki, S. (2023). Expert perspectives on GDPR compliance in the context of smart homes and vulnerable persons. *Information & Communications Technology Law*, 32(3), 385–417. <https://doi.org/10.1080/13600834.2023.2231326>
- Piasecki, S., & Chen, J. (2022). Complying with the GDPR when vulnerable people use smart devices. *International Data Privacy Law*, 12(2), 113–131. <https://doi.org/10.1093/idpl/ipac001>
- Prifti, K., Krijger, J., Thuis, T., & Stamhuis, E. (2023). From bilateral to ecosystemic transparency: Aligning GDPR's transparency obligations with the European digital ecosystem of trust. In S. Kuhlmann, F. De Gregorio, M. Fertmann, H. Ofterdinger, & A. Sefkow (Eds.), *Transparency or opacity: A legal analysis of the organization of information in the digital world* (Vol. 1, pp. 115–140). Nomos. <https://doi.org/10.5771/9783748936060>
- Quiring, O., Ziegele, M., Schemer, C., Jakob, N., Jakobs, I., & Schultz, T. (2021). Constructive skepticism, dysfunctional cynicism? Skepticism and cynicism differently determine generalized media trust. *International Journal of Communication*, 15, 3497–3518. <https://ijoc.org/index.php/ijoc/article/view/16127>
- Regulation 2024/1689. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial*

Intelligence Act). European Parliament and Council. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

The Amsterdam paper: Recommendations for the technical finalisation of the regulation of GPAI in the AI Act. (2024). [Report]. AI, Media & Democracy Lab. <https://www.aim4dem.nl/the-amsterdam-paper-recommendations-for-the-technical-finalisation-of-the-regulation-of-gpai-in-the-ai-act/>

Van Ooijen, I., & Vrabec, H. U. (2019). Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. *Journal of Consumer Policy*, 42(1), 91–107. <http://doi.org/10.1007/s10603-018-9399-7>

Varošanec, I. (2022). On the path to the future: Mapping the notion of transparency in the EU regulatory framework for AI. *International Review of Law, Computers & Technology*, 36(2), 95–117. <http://doi.org/10.1080/13600869.2022.2060471>

Appendices

Appendix A



Jeron van Hoorn

Woensdag 13 september, 08:04

**Rijksmuseum krijgt grootste particuliere
schenking ooit**



Geschreven door kunstmatige intelligentie

Woensdag 13 september, 08:04

Rijksmuseum krijgt grootste particuliere schenking ooit



Jeron van Hoorn

Woensdag 13 september, 08:04

Meer immigranten naar Nederland, vooral door oorlog in Oekraïne



Geschreven door kunstmatige intelligentie

Woensdag 13 september, 08:04

Meer immigranten naar Nederland, vooral door oorlog in Oekraïne

Appendix B

Items empowerment scale:

1. I would continue consuming news from that source. ($M = 3.38$, $SD = 1.76$).
2. I want to be able to filter news content that has been written by this source. ($M = 3.85$, $SD = 1.85$).
3. I want to be able to inform the news organization about biases that I see in their news produced by this source. ($M = 3.50$, $SD = 1.85$).
4. I want additional information about the news production and distribution. ($M = 3.48$, $SD = 1.93$).
5. I want to be able to report the article. ($M = 3.85$, $SD = 1.98$).
6. I want to be able to complain to the news organization. ($M = 3.74$, $SD = 2.04$).
7. I want to be able to have the option to talk to the editor. ($M = 3.11$, $SD = 1.94$).

Appendix C

TABLE A1: One-way ANOVA for trust

	DF	SUM SQ	MEAN SQ	F-VALUE	P-VALUE
TRUST IN SOURCE	3	21.90	7.30	2.85	0.04
RESIDUALS	203	520.6	2.56	-	-
TRUST IN INFORMATION	3	21.40	7.13	2.81	0.04
RESIDUALS	203	515.0	2.54	-	-

The Tukey post hoc test revealed that for both – trust in the source and trust in the information – significant differences between the politicised AI group and the politicised human journalist group exist. Individuals who received the politicised headline with the AI transparency cue were significantly more distrusting of the source than individuals who received the politicised headline with the human journalist label. This is identical for the perceived trust in the information.

Published by



ALEXANDER VON HUMBOLDT
INSTITUTE FOR INTERNET
AND SOCIETY

in cooperation with



CREATE



centre
— internet
et societe



R&I

IN3
Internet
interdisciplinary
Institute

Universitat Oberta de Catalunya



UNIVERSITY OF TARTU
Johan Skytte Institute of
Political Studies