Trust Indicators and Explainable AI: A Study on User Perceptions

Delphine Ribes[1](✉), Nicolas Henchoz[1], Hélène Portier[1], Lara Defayes[1], Thanh-Trung Phan[4,5], Daniel Gatica-Perez[4,5], and Andreas Sonderegger[2,3]

[1] EPFL+ECAL Lab, Ecole Polytechnique fédérale de Lausanne, Lausanne, Switzerland
delphine.ribes@epfl.ch
[2] Bern University of Applied Sciences, Bern, Switzerland
[3] Université de Fribourg, Fribourg, Switzerland
[4] Idiap Research Institute, Martigny, Switzerland
[5] LIDIAP-STI, Ecole Polytechnique fédérale de Lausanne, Lausanne, Switzerland

Abstract. Nowadays, search engines, social media or news aggregators are the preferred services for news access. Aggregation is mostly based on artificial intelligence technologies raising a new challenge: Trust has been ranked as the most important factor for media business. This paper reports findings of a study evaluating the influence of manipulations of interface design and information provided in the context of eXplainable Artificial Intelligence (XAI) on user perception and in the context of news content aggregators. In an experimental online study, various layouts and scenarios have been developed, implemented and tested with 266 participants. Measures of trust, understanding and preference were recorded. Results showed no influence of the factors on trust. However, data indicates that the influence of the layout, for example implicit integration of media source through layout structuration has a significant effect on perceived importance to cite the source of a media. Moreover, the amount of information presented to explain the AI showed a negative influence on user understanding. This highlights the importance and difficulty of making XAI understandable for its users.

Keywords: Trust indicators · Fake news · Transparency · Design · Explainable AI · XAI · Understandable AI

1 Introduction

News aggregators are considered as one of the most phenomenal changes in the media industry over the last decade [1], leading to disruption of the content supply chain and therefore of the business models in the media industry. They can be considered as a curation of news from various media organisations [2]. This curation can be performed by machine based algorithms, human judgement or a mix of both. However, due to the volume of content processed, automation has gained in importance in recent years. In 2009, for instance, Google was already curating over 25000 media sources around the world without human assistance. The phenomenon continues to increase: according to the 2019

Published by Springer Nature Switzerland AG 2021
C. Ardito et al. (Eds.): INTERACT 2021, LNCS 12933, pp. 662–671, 2021.
https://doi.org/10.1007/978-3-030-85616-8_39

Reuters report [3], across all countries, search engines, social media or news aggregators are the preferred services for news access. However, previous research [4, 5] has indicated the importance of trust for the acceptance and perceived usefulness of such tools. But the rise of automated aggregation generates new trust issues. First, the multiplication of sources and the ease of creating and sharing content has spawned the massive distribution of fake news [6]. Additionally, users may get confused by the variety of news sources [7]. Finally, coming along with the increasing complexity and opacity of aggregating algorithms, users may not understand how algorithms make decisions, resulting in a loss of confidence in the aggregation [8]. In this regard, the question arises as how trust and confidence can be influenced when designing interfaces.

The work presented in this paper aims at evaluating the influence of design measures in the particular context of news content aggregators and specifically the effect of layout design and algorithm transparency on user trust.

2 Background and Related Work

2.1 Trust Indicators

To address trust issues, various projects have been initiated. For example, following the US 2016 election, “The Trust Project” [9] was launched, putting forward a set of eight trust indicators with the objective to provide information on the news article (*e.g.* the source, the publication date, the funding). Similarly, other sets of indicators have been proposed [10]. Such indicators are used, for example, for the “Facebook context button”. From the proposed list of indicators, the source appears to be essential to assess the reliability of the content [11–14] and evidence from literature [15] suggests that source visibility should be designed to favor visibility. In a more intrusive way, some social networks have launched warning labels such as the “Facebook disrupted label” or “warnings on misleading Tweets label”. Their purpose is to inform readers about the veracity or the trustworthiness of the information. Whether produced manually by external and independent fact checking organisations or in a more automatic way based on mathematical models [16], they may be harmful if misused or used inadequately [17, 18].

2.2 Aggregation Algorithm Transparency

Transparent AI or eXplainable Artificial intelligence (XAI) systems provide understandable justifications for algorithm outputs [19]. XAI is considered critical for building trust with AI [20]. To successfully implement a XAI system, it is important to determine the XAI system goals, the information to be explained and the way it is explained [21]. In the particular context of news content aggregators, the goal for the XAI system is to make the user understand why a content has been aggregated and how well the content matches the user search criteria. In this regard it has been suggested that information should be explained and represented in a simplified way in case that users are AI-novices [22]. In addition, it should be designed in an engaging way [23] and encourage users to interact with and further inspect the result [24]. In addition, previous research on automation has shown that the accuracy of the automated system (also referred to as automation reliability) considerably influences trust and usage behaviour [25]. Furthermore, research on technology acceptance has shown the important role of system usefulness and ease of use [26].

2.3 The Present Study

In order to address the above-mentioned research question, different versions of an AI-based news content aggregator prototype for an important national event (i.e. winegrowers' festival) were developed. The versions differed with regard to detail of explanation of the AI (level of XAI) and the visual representation of the results (design manipulation). With regard to the design manipulation, it was expected that the visual organisation of the content based on its source (i.e. curated heritage archives, news media archives, or social media) would lead to an increase in trust in the result of the content aggregator compared to an unstructured presentation. In a similar vein it was hypothesized that increasing detail of explanation of the AI would lead to a better understanding and hence to higher trust ratings.

3 Method

3.1 Participants

A total number of 226 participants (mean age of 32.38, SD = 13.74, ranging from 18 to 80 yrs., 128 women, 94 men and 4 identifying themselves differently) were recruited for this study. About half of participants (N = 108) were students while the other half consisted of employees working in various domains.

3.2 Experimental Design

We conducted an online experiment following a 2 × 3 between-subjects design, with the factors 'XAI' (detailed explanation vs. short explanation vs. no explanation) and design (source order vs. rank order). Figure 1 shows the different experimental manipulations. In a second stage of the study, one variable (*Perceived relevance of the interface design)* was recorded and analysed as repeated measures variable.

For the manipulation of the level of XAI in the *short explanation* condition, a color-coded matching rate number (CCMRN) in percentage was presented for each search result, together with the relevant keywords next to the aggregated content. This information helps understanding the prioritisation of the content and the accuracy with which the content matches the user search. In the *detailed explanation* condition, the same information as in the short explanation condition was presented. Furthermore, users could obtain additional information on content features, keywords extracted from indexation as well as success rate in percentage regarding the AI indexation by clicking on a button labeled "+ display more details". In the *no explanation condition*, no information regarding matching rate and search performance was presented.

In the *source order* condition of the design manipulation, search results were organized with regard to the source they were extracted from. The different sources of content were: "heritage" for content originating from curated cultural heritage archives such as museums, "news media" for content being produced by national and regional news and television companies, and "social network" for content being generated by users on platforms such as Instagram and Facebook. This version of the news aggregator interface was designed to clearly indicate the source of the presented content. Therefore, results were sorted into labeled and color-coded columns, each column representing a media type. The search results within each column were organized with regard to their matching rate with the search keywords, with high-matching content being placed on top of the column. Each column could be scrolled individually while the title always stayed visible. In the *rank order* condition, search results were ordered with regard to matching with the search keywords, without being organized by source. However, the source of the content was indicated by a color code in the description.

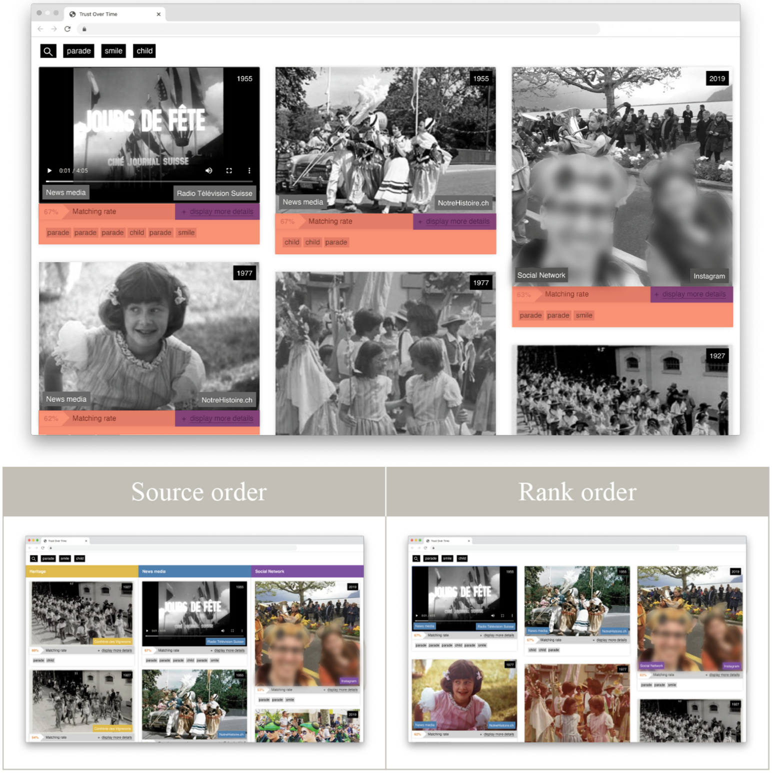

Fig. 1. Representation of the different interfaces as a function of XAI (detailed explanation vs. short explanation (without the purple) vs. no explanation (without the purple and salmon - top image) and design (source order vs. rank order - bottom image). (Color figure online)

3.3 Measures

Trust in the search results was assessed with the items 'Do you trust the results obtained by this search engine?', 'Did you find the results obtained by this search engine relevant?', ‘Did you find that the images that came up in the search matched the keywords?’ with a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’.

Perceived adequacy of the sorting order was assessed with the item ‘Would you have liked to change the order in which the images appeared?’ using a Likert scale ranging from 1 ‘not at all’ to 7 ‘absolutely’.

Perceived usefulness of citing the source was assessed with the item ‘Do you find it useful to have the source of the content cited?’ on a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’.

Subjective understanding of the CCMRN was assessed with the item ‘Did you understand what the matching rate number represents?’ on a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’. Two screenshots displaying two different CCMRN (one with a low value and one with a high value) with their associated content were shown above the item.

Objective understanding of the CCMRN was assessed with the item ‘Try to explain, as simply as possible, how this matching rate number is calculated’. Participants’ answers were coded into three categories, 1: not understood, 2: moderately understood, 3: understood.

Perceived usefulness, trustworthiness and interest in the CCMRN was assessed with the items ‘Do you find the matching rate number to be a useful indicator?’, ‘Do you trust this indicator?’ and ‘Would you like to have an indicator like this when doing online search?’ on a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’.

Perceived accuracy of the CCMRN was assessed with the item ‘Do you find that the percentages shown match the similarity between the search keywords and the found content?’ on a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’.

Perceived usefulness, trustworthiness and interest in the detailed explanation available was assessed with the items ‘Does this information help you to understand how the matching rate number is calculated?’, ‘Would you like to have access to this kind of information the next time you do an online search?’, ‘Do you trust this information’ on a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’. A screenshot displaying the content together with its associated CCMRN and its associated details was displayed above the item.

As a repeated measures variable, *Perceived relevance of the interface design* was assessed by asking the participants ‘Do you find it relevant to sort the results according to the interface below? on a scale ranging from 1 ‘not at all’ to 7 ‘absolutely’. The question was asked twice. First, a screenshot of the source order interface was displayed and then, a screenshot of the rank order interface was displayed.

Additional variables such as perceived usefulness, relevance and accuracy of human indexation versus AI indexation were recorded but are not explicitly reported in this publication and will be part of a future publication.

3.4 Procedure

Participants recruited via institutional mailing lists and social media were equally distributed over the six experimental conditions. They answered an online questionnaire before and after seeing one of the online versions of the interface. They were asked to use a desktop computer or a laptop to complete this online-study. The search query feature was removed in order to control for results quality (i.e. all participants obtain the same search result interface for a maximal experimental control). Before launching the user interface, a text was displayed explaining the context of the search and indicating the keywords used in the search, which were 'parade', 'smile', and 'child'). The time to complete the survey was about 30 min.

3.5 Statistical Analysis

Data was analysed with two-factorial ANOVAs with post-hoc analyses using Sidak corrections. Data regarding the *perceived relevance of the interface design* was analysed with a three-factorial mixed ANOVA with the two questions regarding the relevance of the design versions (source order vs. rank order) as repeated measures variables. The link between two measures was calculated using the Pearson correlation.

4 Results

Trust in the Search Results. Participants in the source order condition tended to trust the results more ($M = 4.91$, $SE = 0.12$) than participants in rank order condition ($M = 4.69$, $SE = 0.12$) but the difference is not significant, $F(1, 220) = 1.70, p = .19$, $\eta^2 = .01$. Interestingly, no significant differences regarding participants' trust in search results were found for the XAI manipulation ($F < 1$), and also the interaction of the two factors did not reach significance level (F < 1). Since no interaction effect reached significance level for all the dependent variables, they are not reported for the following dependent variables.

Perceived Adequacy. Participants in the source order condition were significantly less inclined to change the order in which the search results were presented ($M = 3.59$, $SE = .18$) than those in the rank order condition ($M = 4.20$, $SE = .18$), $F(1, 220) = 5.53, p = .02$, $\eta^2 = .03$. No significant difference was found for the XAI manipulation $F(2, 220) = 1.45, p = .23$, $\eta^2 < .01$.

Perceived Relevance of the Interface Design. The repeated measures comparison of the two design versions (source order vs. rank order) indicated significantly higher ratings for the source order ($M = 5.28$, $SE = .10$) compared to the rank order ($M = 3.74$, $SE = .12$), $F(1, 220) = 113.41, p < .001$, $\eta^2 = .34$. Both between-factors did not influence this measure, $F_{\text{design}} < 1$; $F_{\text{XAI}}(2, 220) = 1.52, p = .22$, $\eta^2 = .014$.

Perceived Usefulness of Citing the Source. Participants in source order condition (S-) found it significantly more useful ($M = 6.58$, $SE = .10$) to have the source cited than participants in the rank order condition ($M = .30$, $SE = .10$), $F(1, 220) = 3.94, p = .048$, $\eta^2 = .02$). No significant differences were found for the XAI manipulations ($F < 1$).

Subjective Understanding of the CCMRN. Participants in the source order condition rate their understanding significantly higher ($M = 5.63$, $SE = .15$) than participants in the rank order condition ($M = 5.18$, $SE = .15$), $F(1, 220) = 4.45, p = .036$, $\eta^2 = .02$. No significant difference was found for the XAI manipulation ($F < 1$).

Objective Understanding of the CCMRN. No significant differences were found for the design manipulation ($F < 1$), while the effect of XAI reached significance level $F(1, 220) = 3.16, p = .044, \eta^2 = .03$. Sidak-corrected post-hoc comparisons indicated that participants in the detailed condition explained significantly less well how the *CCMRN* is calculated ($M = 2.38, SE = .08$) compared to participants in the short explanation ($M = 2.66, SE = .08; p = .04$), while the comparisons with no explanation condition ($M = 2.58, SE = .08$) did not reach significance level. There is a positive relationship between subjective and objective understanding of the *CCMRN*, $r = .27, p < .001$ indicating that participants who subjectively think they understood the *CCMRN* were better able to explain it.

Usefulness, Trustworthiness and Interest in the CCMRN. No significant differences were found for the design manipulations $F(1, 220) = 1.2, p = .26, \eta^2 = .006$ nor for the XAI manipulations $F(2, 220) = 2.2, p = .11, \eta^2 = .02$.

Perceived Accuracy of the CCMRN. No significant differences were found for the design manipulations ($F < 1$) nor for the XAI manipulations ($F < 1$).

Perceived Helpfulness, Trustworthiness and Interest in the Detailed Explanation. No significant differences were found for the design manipulations F (1, 220) = 1.2, p = .26, $\eta^2 = .006$ nor for the XAI manipulations F (2, 220) = 2.2, p = .11, $\eta^2 = .02$.

5 Discussion and Conclusion

This study addressed the effect of design manipulations and XAI on user trust and understanding. Surprisingly, results indicate that user trust is not influenced by the proposed design manipulations. We believe there are several reasons for this. The first one may be the content used. The winegrowers' festival is a popular event whose media content may not be prone to fake news, whatever the source. Therefore, the implemented design manipulations might have shown little influence on trust. To pursue efforts in understanding the influence of interface design on user trust, it might be important to conduct additional research with the suggested design manipulations on different content more connotated to include fake news (e.g. political, ecological or economical information). The second explanation for the unexpected outcomes may be due, as discussed in [15], to the similitude in the presentation of the content for both design conditions, resulting in a comparable effect on trustworthiness of the design manipulations. We manipulated the design of results sorting, either by source or by ranking order but the presentation format of the content remained identical for both conditions (*i.e.* position of the image or video within the content, size of the content, size of the elements surrounding the content), leading to a similar effect on user trust. Finally, the lack of effect on user trust might be due to the visibility of the sources in both conditions. Since the source appears to be essential to assess the reliability of the content [11–14], the effect on trust remains the same as long as the source is clearly highlighted.

While the proposed manipulations of XAI and design did not influence trust ratings, they did show an effect on subjective and objective understanding of the CCMRN. For participants in the detailed explanation condition, a discrepancy was observed between their subjective and objective understanding of the CCMRN: in the detailed explanation condition, participants performed worse in explaining the CCMRN. Nevertheless, they rated their subjective understanding similarly when compared to the other two XAI conditions. This indicates that providing an additional amount of information does not automatically lead to a better understanding of the AI system. Similar findings were presented in previous research where it was shown that XAI systems must be designed in a tailored way as otherwise this too much information can erode trust [25]. We observed however that the understanding of the CCMRN is not converted at this point into an effect on trust. This could be due to the fact that participants found it difficult to imagine the meaning of a percentage score. It would therefore be interesting to test the effect using only the red-orange-green colour code, the highlighted keywords, or a sentence as proposed in [25, 27, 28]. We believe that the effect on comprehension would be comparable, but users would be more inclined to use it. Surprisingly, we found that participants in the *source order* condition rated their subjective understanding of the CCMRN higher compared to participants in the *rank order* condition, but their objective understanding was similar. In the *source order* condition, results are sorted into categories and displayed as columns which can be scrolled individually. On the contrary, in the *rank order* condition, results are scrolled all together. The clear organisation in three separate columns might facilitate the reading and comprehension of the relevant information, which helped participants to focus on the meaning of the presented information.

Interestingly, the design manipulations showed effects on specific user perceptions. Participants in the *source order* condition found it more important to present information about the source in addition to the content compared to participants in the *rank order* condition. This corroborates results found in previous research [15] indicating the importance of clearly indicating the source of aggregated content in order to favor its visibility. However, as the results of this study indicate, the way the source is highlighted is also an important factor to influence people's perception of its importance. Moreover, results indicated that participants in this study preferred the *source order* layout. As for the subjective and objective understanding of the CCMRN, the proposed source order layout made the content easier to read, which might have influenced their preference ratings.

Limitations of this study are that results are based on data of a popular event dataset in Switzerland and hence need to be interpreted within this cultural context. Moreover, results are based on self-report data, which must be taken into account when interpreting the results [29].

To conclude, the presented study established a connection between layout design and user understanding. More specifically, we related the sorting of news into column categories to the importance of citing the source. This knowledge is useful in the fight against misinformation when designing news content aggregators. The results of this study also demonstrated a link between the amount of information provided to explain AI and the understanding of the AI, indicating that more information does not forcefully lead to a better understanding. In this regard, the real challenge for the future development of XAI environments might not be to make the system explainable but to make it understandable for the person that is supposed to use and trust it.

This project has received funding from the Initiative for Media Innovation based at Media Center, EPFL, Lausanne, Switzerland.

References

1. Lee, A.M., Chyi, H.I.: The rise of online news aggregators: consumption and competition. Int. J. Media Manage. **17**(1), 3–24 (2015). https://doi.org/10.1080/14241277.2014.997383
2. Isbell, K.: The rise of the news aggregator: legal implications and best practices. SSRN Electron. J. (2012). https://doi.org/10.2139/ssrn.1670339
3. Newman, N.: Reuters Institute Digital News Report 2019, p. 156 (2019)
4. Oechslein, O., Haim, M., Graefe, A., Hess, T., Brosius, H.-B., Koslow, A.: The digitization of news aggregation: experimental evidence on intention to use and willingness to pay for personalized news aggregators. In: 2015 48th Hawaii International Conference on System Sciences, HI, pp. 4181–4190, January 2015. https://doi.org/10.1109/HICSS.2015.501
5. Innovation in News Media World Report 2018. WAN-IFRA. https://wan-ifra.org/insight/innovation-in-news-media-world-report-2018/. Accessed 16 Apr 2021
6. Rubin, V.L., Chen, Y., Conroy, N.K.: Deception detection for news: three types of fakes. Proc. Assoc. Inf. Sci. Technol. **52**(1), 1–4 (2015). https://doi.org/10.1002/pra2.2015.145052010083
7. Reuters Institute Digital News Report 2017, p. 136 (2017)
8. European Commission: Final report of the high level expert group on fake news and online disinformation. Shaping Europe's digital future - European Commission, 12 March 2018. https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation. Accessed 14 Dec 2020
9. The Trust Project Homepage. https://thetrustproject.org/
10. Zhang, A.X., et al.: A structured response to misinformation: defining and annotating credibility indicators in news articles, p. 10 (2019). https://doi.org/10.1145/3184558.3188731.
11. Kiousis, S.: Public trust or mistrust? Perceptions of media credibility in the information age. Mass Commun. Soc. **4**(4), 381–403 (2001). https://doi.org/10.1207/S15327825MCS0404_4
12. Hovland, C.I., Weiss, W.: The influence of source credibility on communication effectiveness. Public Opin. Q. **15**(4), 635–650 (1951). https://doi.org/10.1086/266350
13. ACUNA, T.: The digital transformation of news media and the rise of disinformation and fake news. EU Science Hub - European Commission, 25 April 2018. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/digital-transformation-news-media-and-rise-disinformation-and-fake-news. Accessed 14 Dec 2020
14. Pornpitakpan, C.: The persuasiveness of source credibility: a critical review of five decades' evidence. J. Appl. Soc. Psychol. **34**(2), 243–281 (2004). https://doi.org/10.1111/j.1559-1816.2004.tb02547.x
15. Kim, A., Dennis, A.R.: Says who?: how news presentation format influences perceived believability and the engagement level of social media users. In: 51st, Hawaii, vol. 43, Issue 3, pp. 1025–1039 (2018)
16. Zhou, X., Zafarani, R.: Fake news: a survey of research, detection methods, and opportunities, 1 (2018). http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/
17. Pennycook, G., Rand, D.G.: The implied truth effect: attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings. SSRN (2017). https://doi.org/10.2139/ssrn.3035384
18. Clayton, K., et al.: Real solutions for fake news? measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Polit. Behav. **42**(4), 1073–1095 (2019). https://doi.org/10.1007/s11109-019-09533-0

19. Gunning, D.: Explainable artificial intelligence (XAI). Mach. Learn. 18
20. Shin, D.: The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int. J. Hum. Comput. Stud. **146**, 102551 (2021). https://doi.org/10.1016/j.ijhcs.2020.102551
21. Mohseni, S., Zarei, N., Ragan, E.D.: A multidisciplinary survey and framework for design and evaluation of explainable AI systems, arXiv181111839 Cs, August 2020. http://arxiv.org/abs/1811.11839. Accessed 08 Apr 2021
22. Lage, I., et al.: Human evaluation of models built for interpretability. In: Proceedings AAAI Conference Human Computation Crowdsourcing, vol. 7, no. 1, October 2019. Art. no. 1
23. Muir, B.M.: Trust between humans and machines, and the design of decision aids. Int. J. Man Mach. Stud. **27**(5), 527–539 (1987). https://doi.org/10.1016/S0020-7373(87)80013-5
24. Kulesza, T., et al.: Explanatory debugging: supporting end-user debugging of machine-learned programs. In: 2010 IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 41–48, September 2010. https://doi.org/10.1109/VLHCC.2010.15.
25. Kizilcec, R.F.: How much information? Effects of transparency on trust in an algorithmic interface. In: Conference Human Factors Computing Systems, pp. 2390–2395 (2016). https://doi.org/10.1145/2858036.2858402
26. Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. Decis. Sci. **39**(2), 273–315 (2008). https://doi.org/10.1111/j.1540-5915.2008.00192.x
27. Eslami, M., et al.: I always assumed that I wasn't really that close to [her]: reasoning about Invisible Algorithms in News Feeds. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, New York, pp. 153–162, April 2015. https://doi.org/10.1145/2702123.2702556.
28. Wang, W., Benbasat, I.: Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs. J. Manage. Inf. Syst. **23**(4), 217–246 (2007). https://doi.org/10.2753/MIS0742-1222230410
29. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. **88**(5), 879–903 (2003). https://doi.org/10.1037/0021-9010.88.5.879