

## EXPLAINABLE AI: HOW EXPLAINABILITY IMPACTS THE HUMAN INTERACTION WITH AI

Donghee Shin<sup>1</sup>

### Abstract

With the rapid growth in the use and implementation of AI in the journalism industry, concerns on the ethical implications have surfaced recently. The issue of explainable algorithmic journalism has grown rapidly and resulted in a large body of attempts that embody normative qualities such as transparency, privacy, and fairness. This study investigates the effect of explainability on privacy in newsbots to understand how users' information processing leads to data disclosure. We discuss the conceptual mapping of explainability and privacy in algorithms, followed by empirical modeling of how the explanatory heuristics play out in the cognitive processes to which they contribute to user privacy and data disclosure. A mixed-method design incorporating both qualitative and quantitative approaches was used to discover user heuristics and to test the effects of incorporating explanatory cues into newsbots on user privacy. Together, our results reveal that explanatory heuristics exert a key role in users' intentions to disclose more data in algorithm interactions. Explainable newsbots facilitate users' understanding of decision-making in algorithms by evoking transparency and fairness, which serve as the cues for user trust and privacy. To suggest best practices for journalism and identify ongoing challenges for newsbots, we discuss algorithmic information processing and show how the process can be utilized to improve user privacy and trust.

**Keywords:** Algorithmic Privacy, Newsbots, Data Disclosure, Explainable Algorithmic Journalism, Privacy Paradox Explainable AI: How Explainability Impacts the Human Interaction with AI

The rise of artificial intelligence (AI)-driven personalization in newsrooms has had profound effects on the journalism industry and has spurred the emergence of newsbots, pushing journalism into a novel era of personalized news with algorithmic modernity (Ozkul, 2021). Advancements in AI technologies have provided breakthrough improvements to content curation, digital reporting, and user analytics (Wölker & Powell, 2020). Newsrooms

apply intelligent automation and audience analytics to collect, aggregate, filter, edit, and distribute content, processes of production and utilization have become highly automated and algorithmicized (Shin, 2021). Newsbots have been reinventing journalism in an AI era (DeVito, 2017; Thurman et al., 2019) as they create new sets of journalistic values, cultivate new areas of markets, and redefine the relationship of readers and editorial staff (Pietrobruno, 2016).

---

<sup>1</sup>Dr. Donghee Shin has been a Professor and Chair at the College of Communication and Media Sciences at Zayed University, Abu Dhabi Campus since 2019. Over the last 22 years, he taught at various universities in the US and Korea, including Penn State University, US and Sungkyunkwan University in Korea. Dr. Shin received his master's degree and Ph.D. degree from Syracuse University.

Through advances in AI, conversational agents that engage in a dialog with users have become popular in journalism. Newsbots (news robots) interact with their readers through interactive cues and can provide personalized news to readers based on their interests (Montal & Reich, 2017). While newsbots offer personalized and relevant news in many interactive manners, the ethical and privacy issues are complicatedly intertwined with algorithmic personalization (Araujo, 2018). Issues regarding how to safeguard the goals, values, and personalizing processes of newsbots, to what extent users need to share personal information with algorithms, and how to balance privacy and algorithmic performance remain controversial (Crain, 2018). As privacy concerns have peaked recently with the rise of newsbots, the opacity of black-box algorithm processes had led to calls for research on explainability and privacy (Dörr & Hollnbuchner, 2017). Sundar et al. (2020) examine the effects of privacy and data disclosure in the adoption of a personalized recommender system. A concept of algorithmic privacy has been proposed in terms of transparency and trust in algorithm-mediated systems (Dienlin & Metzger, 2016; Shin, 2021). How users make sense of privacy in their algorithm use, how users construe algorithmic personalization, and how they perceive algorithm-based processes will be vital issues to consider as AI develops even more prominent in journalism (Graefe et al., 2018). These issues become even more critical in advanced newsbots because news algorithms face a dilemma between privacy and personalization. Users expect high personalization, customized news recommendations, and at the same time, control over their data (Privacy Paradox). As awareness grows around data privacy, along with new data policies like the General Data Protection Regulation (GDPR), users demand both personalization and privacy, as they seek to control their data (Acquisti, Brandimarte, & Loewenstein, 2020).

Reflecting the rising concern over privacy and transparency, there is mounting

pressure in newsbots for legitimate justifications on why and how results are derived through algorithmic curations (Crain, 2018). As users have little control over how the news they view is generated, explanations have been requested for the type of data utilized for personalization and how data were collected. Recent studies on algorithmic privacy have shown the facilitating role of explainability in user experiences with algorithm-driven news (Iscehen et al., 2020; Rai, 2020). There have been increasing debates over the methods and effects of explanations (Gunning et al., 2019), how to effectively provide explanations, how to balance the right to privacy and the right to an explanation, and to what extent privacy concerns can be reconciled with the transparency of the system. Incorporating explanations is a response to the social and user rights to the algorithms that transparent explanations can enhance the user experience of the content of service by establishing privacy and trust AI even if there is no regulatory or legal requirement (Iscehen et al., 2020). This topic will be even more critical when privacy is context-dependent (Wang et al., 2020) and the data disclosure is dependent upon users' perceived privacy (Benson, Saridakis, & Tennakoon, 2015; Dienlin & Metzger, 2016).

Despite the significance, few studies have researched the heuristics of privacy and explainability in AI. The process of how explainability influences user privacy and which factors influence users' willingness to disclose personal information in newsbots remains unanswered (Rai, 2020). In view of the research gap in the literature, we examine the relation of explainability with privacy and trust by focusing on users' cognitive processing with the following research questions (RQs).

RQ: What can be done to make newsbots more transparent and trustworthy?

1. How do users perceive algorithmic personalization when they interact with newsbots? How do they evaluate algorithmic personalization in terms of privacy?

2. How do explanatory cues newsbots help users to perceive normative value and how does that perception lead to user trust, privacy, and data disclosure?

The RQ we pursue to address in this study is how cognitive heuristics influence privacy decisions and predict data disclosure in contexts where algorithms are used with personal data to predict user needs and expectations. We propose that users disclose personal information based on a cognitive process through which the users interpret their experiences and come to their understandings heuristically and systematically. Findings reveal the internal process of how algorithmic personalization is cognitively processed, what the roles of trust and privacy play in the process, and how users come to the decision of data disclosure (Koochikamali, Peak, & Prybutok, 2017). These processes illustrate users' information processing of newsbots, which is characterized by a users' dual-process in newsbot interaction (Wang et al., 2020). Data disclosure intention develops along with a dual-venue including the heuristic route and the rational route in evaluating privacy conditions and deciding how and whether to disclose their personal information. Users' privacy and trust decisions are not always rational but subject to bounded rationality and imperfect or asymmetric information regarding how their data is collected, for what purposes, and with what consequences (Acquisti et al., 2020). There is a wide range of cognitive processes and dispositions as to whether, how, and to what extent to trust algorithm-based services (Sundar et al., 2020). Such decisions are based on user heuristic assessments of normative values and systematic evaluations of rational weigh the trade-offs associated with the privacy and the sharing of personal data (Keith et al., 2013). Based on the dual-process, we propose cognitive mechanisms of explainability-trust-privacy-data disclosure, which provide significant contributions and advances in the better understanding of the algorithmic heuristics, in particular algorithmic privacy and related behavior. Our work constitutes

a conception of algorithmic information processing development, how users perceive, analyze, control, use, and accept algorithms.

Theoretically, clarifying the privacy heuristics along with data disclosure in algorithmic personalization would make momentous contributions to the ongoing developments of human-AI interactions. As there is no established theory of algorithmic heuristics yet, proposing an internal cognitive logic and process of how algorithmic cues contribute to privacy-related decision-making in newsbots. Clarifications of these processes would contribute to the debate of how humans and algorithms co-construct and how to design algorithms human-centered AI. Understanding the use and formulation of heuristics is critical for academia, industry, and for regulators.

## Literature Review

Understanding user heuristics for the news from AI-driven systems is key not only for theory to academia, but also for practice to the algorithm industry, and regulation for policymakers. A suitable theory of algorithmic privacy would help the industry to use it effectively as well as would help them to design effective newsbot interfaces. The section provides a conceptual overview of the heuristic perspective in users' decision-making about privacy and data disclosure in algorithmic contexts.

### Newsbots: Why Do News Readers Need Explanations?

Newsbots in this study are defined in two ways. First, newsbots are based on a machine-learning process whereby data in the form of facts and figures is input to a machine-learning process or algorithmically-informed content production software, and then news articles are output (Montal & Reich, 2017). Under this definition, news articles are personalized, generated, and even written by algorithmic programs. Through machine learning software, articles are produced algorithmically by AI rather than by human reporters.

Second, some newsbots mainly use automated functions of news algorithms (Caswell & Dörr, 2018). In this case, newsbots develop and provide automated algorithms to news outlets. News providers such as Automated Insights, ProPublica, and Associated Press can be under this category. Newsbots in this line of category deal with topical news like sports recaps, weather, financial reports, and earning reviews. These newsbots commonly use a combination of algorithms and human editing. Newsbots generate news based on the data from users, which are used in algorithmic recommender systems. Human reports receive the contents from algorithms and then edit the contents to make them more user-likely news (Shin, Chotiayaputta, & Zaid, 2022).

Both forms of newsbots have faced the concern of the opacity nature of algorithm systems: black-box systems that take an input, generate an output but give no clue why (Dörr & Hollnbuchner, 2017). Against the increasing concern over black-box, explainability has been proposed as a measure to solve the problem. Explainability is the extent to which the internal mechanics of newsbots can be explained in understandable human language (Arrieta et al., 2020). Explainability is crucial in establishing trust and understanding between an AI agent and its user, particularly in the case of undesirable consequences and unanticipated faultiness. Explainability allows newsbots' editorial staff, journalists, and newsrooms to better understand why certain news algorithms are made—and to correct as needed. This is important in the context of journalistic values in AI, as it will enable journalism to identify potential bias and discrimination against certain values.

### **Explainability: A Reconciliation between Privacy and Personalization**

With increasing AI evolution, users expect high personalization, bespoke content, and tailor-made news recommendations. For personalization, AI needs to collect users' data to make predictions about their

preferences, their needs, and their influences on them. Thus, the use of user data inevitably raises privacy concerns, which is the personalization-and-privacy paradox (Wang et al., 2020). As the tension arises between privacy and algorithmic personalization, explainability has been proposed as a possible solution to mitigate the user's privacy concern as well as effectively enhance personalization (Dienlin & Metzger, 2016). Explainability enables journalism to provide greater context than would be presented in traditional news sources. For example, Vox, the explanatory news website, presents ongoing news in a more interpretable manner by including photos, charts, or videos to explain complicated issues readers don't understand. Vox uses algorithms to detect how important certain content is rated to increase the reader viewership. Also, it uses machine learning to introduce new products, news, or people who may expand the social networks of others. The Conversation, conversational explanatory journalism, provides information that is explained and accessible to readers. It has started to automate the production of its news; not only do algorithms create thousands of news stories for a particular topic, but they also do it effectively and in a trustworthy way with understandable explanations. The use of an algorithm in emerging media helps to provide transparency of editorial processes and build user trust in algorithmic journalism.

Explainable approaches can transparentize the algorithm's inner processes and reveal insight into the factors that influenced its result. Explainability can help users to develop a sense of reasonableness around their heuristics, by giving users a sense of control insofar as users can decide whether or not to rely on algorithmically curated news. For instance, readers are more likely to trust health-related news when it comes to explanations of understandable terminology and human reliable sources like human doctors. The presence of explanations can hold newsbots liable for their personalization by ensuring that they are compliant with ethical standards, by verifying the inner

workings, and by making the best efforts to generate accurate recommendations.

Explanatory journalism is nicely in line with emerging trends of explanatory AI (Mann, 2016). Explanatory journalism attempts to present news stories in a more understandable and user-centered manner. This attempt is consonant with explainable AI, which is trying to the recommendation that can be understood by humans. While explanatory journalism aims to make complex topics easier to understand content-wise, explainable AI is to provide explanations of how certain contents are generated from a procedural and technical perspective. Thus, in coupling the two ideas, explainable AI journalism can provide users with dual explanations on content aspects of how the news story has been developed as well as procedural aspects of how the specific news is algorithmically curated, generated, and presented to users. The dual-explanation can help users to have a sense of readability on the news carried and trust in the process it delivers.

### **Context-Dependent Privacy and User-Dependent Algorithms**

AI raises concerns about data privacy and data disclosure as it uses sensitive personal data for training the algorithm logic, particularly in an era where fake news and misinformation are rampant, and users' right is valuable to identity. A newsbot is an AI system that can modify its behavior based on user information or machine learning. To generate recommendations, algorithms need personal information. Personal data can be gathered by direct requests to disclose this data, which might induce privacy concerns (Ischen et al., 2020). Privacy is a critical topic in newsbots as such services need huge volumes of user data (Sundar et al., 2020). Since data includes user private data, there are concerns about the privacy implications of newsbots such as what is required for user information to be used and what privacy measures are required to protect user data (Dienlin & Metzger, 2016). Privacy in an algorithm context can be defined as the extent

to which users are concerned with the potential risks of the right to prevent the disclosure of individual information to others (Sundar et al., 2020). This definition renders a right to receive an explanation of decisions reached by algorithms, but this explanatory process of demand leads to a scrutiny of the data that trained the algorithms. It will lead to demands to see the data and thus will breach the privacy rights of users from whom the training data was derived. Users have the right to know which of their personal data are gathered, utilized, analyzed, or otherwise processed and to what extent the personal data are or would be processed. Ischen et al. (2020) have shown that users have a desire to have some control over how their information is gathered and analyzed and have a concern for privacy when it comes to interactions with AIs. Since the interaction with chatbots is a new feature and the black-box nature of AIs, users may be more cautious about how their data are collected, used, and represented (Acquisti, Brandimarte, & Loewenstein, 2020). A question of how to balance between privacy and explainability emerges as a key question for trustworthy algorithms (Rai, 2020). Algorithmic explanations offer insights into an AI's decisions on input data, whereas privacy is primarily concerned with safeguarding information about the training data.

### **Algorithmic Information Processing Theory**

The algorithmic information processing (Shin, 2021) is applied as a theoretical frame to trace user sensemaking in newsbots by analyzing the role that algorithmic attributes play in shaping user privacy and trust, and sensemaking of algorithms, as well as how user behaviors affect their data disclosure. With the emergence of algorithmic services, it is important to understand how algorithmic information is processed, consumed and how users interact with newsbots (Shin, 2021). The theory has been used to seek a better understanding of the cognitive process of algorithm rationalization that affects users' private heuristics and related behaviors by

examining how the cognitive processes people use to sense-making influence their decisions to disclose personal information. The main tenet of the theory addresses user motivation, heuristics, and behaviors. User heuristics concerning algorithm processing include questions like: How do users assess the qualities and attributes of algorithms? How do people sense them and with what meanings? The theory is a constructive theoretical lens for these inquiries as our model theorizes that attitude and behavior changes occur through different routes of processing (Wang et al., 2020). This theoretical frame allows us to discover how users process the algorithmic information they have, beyond simply responding to stimuli.

### **Algorithmic Information Processing in Newsbots**

The proposed model is to examine the effects of explainability, anthropomorphism, and privacy on data disclosure in newsbots (Figure 1). This model is particularly useful in examining the effects of explanations as the explanatory processes involve users' sense-making of privacy and data disclosure in newsbots.

### **How News Readers Make Sense of Algorithmic Explainability**

As algorithms are increasingly adopted in journalism practice, where values and morals are important, FAT is becoming an underlying operational requirement and normative value (Shin, 2021). Reader perceptions of newsbots' content often depend on their understanding of how the algorithm works (Graefe et al., 2018). How and why certain news articles are generated, together with how user input contributes to outcomes, have been confirmed as being important to users (Arrieta et al., 2020). The increasing notion is that readers deserve to be given a transparent explanation of how algorithm tools were used to conduct an analysis, identify a pattern, or report news. Providing explanations in recommendation systems can enhance positive attitudes and subsequent behaviors among users (Shin, 2021).

Previous studies have suggested a causal link between explainability and reader assurance in the context of newsbots (Rai, 2020).

Ongoing research consistently shows that newsbots should assist users in understanding the algorithmic process, and thereby increase user trust. Explainability can help address issues of FAT and trust (Arrieta et al., 2020). People are more likely to trust and accept transparent systems when they are aware of how data are collected, handled, analyzed, and presented. Newsbot readers should also be able to understand the process of a personalization system (Rai, 2020). Also, accountability and fairness in algorithms are important since black-box algorithms must be verified to ensure they are accurate and unbiased (Diakopoulos, 2019). Accountability in the newsbots context indicates that newsrooms should be held responsible for the results of their programmed algorithms. As the product of human editors, newsbots can have issues resulting from human bias or simple oversight. In this light, explainability can be hypothesized with normative values.

H1: Explainability positively influences user perceptions of newsbot transparency.

H2: Explainability positively influences user perceptions of newsbot fairness.

H3: Explainability positively influences user perceptions of newsbot accountability.

Procedural Evaluation: How Do Algorithms Work?

There is a very limited understanding as to exactly how trust is constructed when users interact with algorithms. Over the last years, there has been extensive research on how users perceive algorithmic news (Graefe et al., 2018). Beyond general credibility and quality assessments, one of the key issues has been the impact of perceived normative values on algorithmic trust. In newsbots, trust is considered one of the central questions concerning transparency in journalism representing how much readers believe the news and news sources and reflects the belief in the accuracy of news filtering and user willingness to accept a recommender system's

capacity (Kim & Gambino, 2016). Research has shown that when it comes to news personalization, people place more trust in algorithmic news curation than human selection (Shin, Chotiyaputta, & Zaid, 2022). Such greater trust appears to be due to neutrality and objectivity in automated news curation, as perceived by users. Thus, trust was introduced as an intermediary between the normative value of newsbots and the performance value of news. This intermediary was usually an opinion leader of influence with greater access to information in the traditional two-step flow of communication. In newsbots, trust is seen as an intermediary in the interaction between readers and AI. When there is a fair process, users increase their input data to improve recommendation outputs (Graefe et al., 2018). The quality of the output is highly dependent on the quality of the data that is fed into it. The quality of the output is contingent on the quality of the data that is fed into it. If users have notions of accountability—including that newsbots are supposed to operate in a manner that contributes to the public good—they feel reassured, and trust is established (Montal & Reich, 2017). Together, the three ethical factors are considered paramount in the development of newsbots insofar as they help establish user trust in AI.

H4: Perceived transparency positively influences user trust in newsbots.

H5: Perceived fairness positively influences user trust in newsbots.

H6: Perceived accountability positively influences user trust in newsbots.

### **Effects of Anthropomorphism**

The feature of humanness has gained a spotlight in recent AI as the goal of AI is to make more human-like machines (Go & Sundar, 2019). Relevant research has shown that increased perceived humanness leads to a positive user experience (Rai, 2020). Perceived humanness in AI contexts is defined as the degree to which AI or algorithms is being experienced as a human being. Users' acceptance of algorithms depends on

perceived humanness, which is defined as the degree to which users perceive that algorithms might be human (Westerman, Aaron, & Lindmark, 2019). Perceived humanness becomes relevant in explainable AI as the explanations are hard to understand to people. Research has suggested that human-like attributes positively influence the acceptance of speech-based chatbots (Araujo, 2018). Users have positive feelings if they perceive high humanness (de Visser et al., 2016). A human-like chatbot with an explanation and humanized communication skills can enhance user trust, which in turn leads to an improved user experience (Go & Sundar, 2019). Perceived humanness can increase explainability and user trust and assured feeling and improve user's experience using algorithms (Westerman et al., 2019).

H7: Perceived humanness positively affects the explainability of newsbots.

H8: Perceived humanness positively affects user privacy in newsbots.

### **Trust, Privacy, and Data disclosure**

News personalization built around algorithms offer a series of benefits to users, but they require huge amounts of user data in return. Privacy becomes vulnerable because algorithms can use the data for good purposes but also, they can misuse or abuse data. Thus, in algorithm contexts, trust, privacy, and data disclosure are closely related concepts (Fast & Jago, 2020). In news recommendation contexts, trust refers to a reliable belief in the accuracy of news recommendations and for user readiness to use the recommender system's capabilities. Privacy is considered a user right and as such, inhibits the disclosure of personal information. Perceptions of privacy in algorithms are significantly conditioned by factors in the algorithmic environment, particularly factors such as personalization and trust. Privacy is critical in newsbots because algorithms need access to huge amounts of user data to generate relevant news. Users are increasingly paying more attention to their information and becoming uncomfortable

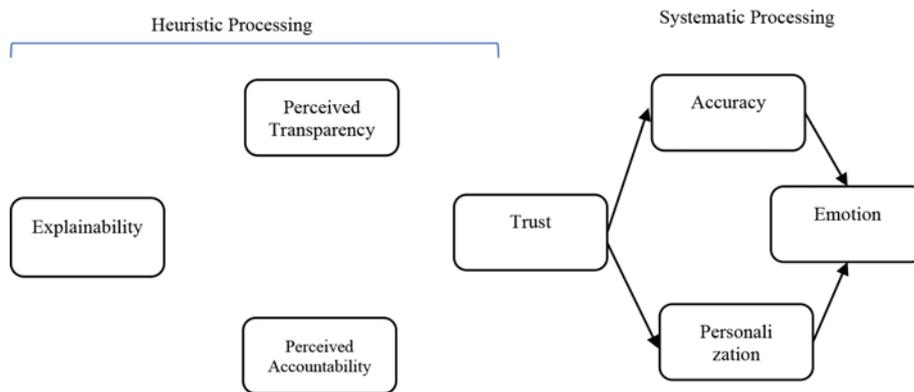
with how data about their preferences, search histories, locations, and interests are used by newsbots they interact with (Sundar et al., 2020). Together, trust and privacy play critical roles in algorithm services by contributing to data disclosure in algorithms (Koochikamali, Peak, & Prybutok, 2017). High levels of trust in algorithms can afford users a sense of privacy and data disclosure. Trustable recommendations afford users a sense of privacy, which, in turn, triggers willingness to disclose personal information (Keith et al., 2013). Relevant research has often found cognitive processes such as users' assessment of normative values and trust as determining factors to users' decision to disclose personal information. Particularly, Shin

(2021) proposed the algorithmic information processing model that found normative values, trust, privacy concerns, and anthropomorphism as salient in the users' decisions to disclose their personal information. By applying the findings from algorithmic information processing, the nature of these cognitive processes could be better understood with regards to their relationship to personal data disclosure.

H9: Trust positively influences user privacy in newsbots.

H10: Trust positively influences data disclosure in newsbots.

H11: Privacy positively influences data disclosure of newsbots.



**Figure 1** Heuristic and systematic processing in AI-based CA

## Methods

Ethical issues were considered prior to data collection in accordance with the guidelines provided by the Korean National Research Foundation. Upon opening the experiment, information was given to participants that their participation was on a confidential basis.

## Mixed Methods

We used a triangulated mixed method to examine the users' heuristics and the effects in newsbots. Mixed methods are particularly useful in understanding algorithm users' points

of view as they produce a voice to examine users and ensure that findings are grounded in users' experiences. Also, the methods are constructive in exploring the algorithmic information processing of newsbots as they confirm factors in users' perceptions and responses, then identify users' underlying heuristics regarding algorithms. Through qualitative methods, we develop diverse user heuristics in newsbots, and through quantitative methods, we test the derived heuristics to evaluate the newsbot system.

**Qualitative Methods.** Using the interpretive nature of qualitative methods,

we focused on respondent interpretations and experiences of algorithmic features to uncover user heuristics. To determine the notion of FAT that is most compatible with a user's perception of normative values, we design a triangulated qualitative method using both in-depth interviews and field experiments wherein respondents provided rationales for their behaviors. We performed qualitative methods with 29 respondents from different populations. First, in-depth interviews were performed with algorithm users. Thirty respondents were randomly selected from public universities located in Seoul, South Korea. We performed qualitative in-depth interviews about their practices in using online recommender systems. This qualitative interview was designed to develop the dimensionality of user experience with algorithms, identify the subjective relevance of FAT in relation to algorithms, and analyze awareness, perception, and evaluation of algorithms in social platforms, news comment prioritization, and online recommendation.

Second, based on the interviews, we designed field experiments in which we tested the effect of explainability on FAT and privacy perceptions, search behavior, and content usage. Thinking aloud was designed to get insights into the thought processes of users while they experience newsbots. The results from thinking aloud can clarify the question of how users get certain heuristics and certain sequences of activities. A total of 100 students were incentivized to participate in the experiments which varied recommender prototypes over treatment groups in order to examine algorithmic media consumption, behavior in user forums, and content choices, each over three months. For online recommendations, we used the Twitter application programming interface, which lets people read and write Twitter data, to rebuild the newsfeed of users. For news article comment recommendation, we utilized data from Joongang Daily, the online newspaper in South Korea with the most active and largest user community, to control suggestions of

relevant posts and opinions. To enable the field experiments, the developed smartphone apps resembled the functions under investigation in popular usage settings. Supplied algorithms were designed according to different notions of fairness, for example through different objectives of parity and diversity, partly reflecting the objectives of various stakeholders. The apps served to request ratings of qualities and feedback. We have initiated an ethics clearance procedure with the ethics committee of the university where the authors are affiliated.

**Quantitative Methods.** Using the conceptual clusters identified from the qualitative approaches, a survey questionnaire was designed through multiple rounds conducted by researchers, practitioners, and experts (Table 1). Respondents were recruited for the experiment and survey. The experiment was designed to test user experiences of the data-in/content-out mechanism by algorithms. This design was made because there was a difference between participants having the impression that content is produced by algorithmic curation and the actual experience that content is generated through user data by algorithms. Upon entering a research lab equipped with algorithmic media platforms, participants were asked to log in to their preferred platforms and instructed (1) to search/read recent trend news and (2) to surf/shop video contents (film/drama/TV series) that were algorithm-generated for about 1-2 hours. The respondents were given a list of news items and video clips that they could choose. They were briefed that the news items/contents were generated by algorithmic platforms. These informational sessions were critical because we wanted to ensure that there was no difference between what we tried to measure newsbot experience and what was understood by participants. After viewing and consuming, they were given a survey to fill out. The respondents were given class credits as compensation for their time spent in the experiment and recruited through undergraduate/graduate courses related to new media, digital

technologies, and consumer research. We were sure to avoid any kind of Hawthorne Effect that respondents possibly modify their

algorithmic behavior in response to their awareness of being experimented upon.

**Table 1** Descriptive Statistics

| Status                | Group 1     | Group 2     |
|-----------------------|-------------|-------------|
| Age (Mean/S.D.)       | 39.12/14.24 | 40.11/16.23 |
| Gender (Female rate)  | 52.30       | 51.11       |
| College-educated      | 27.11       | 29.34       |
| CA service experience | 1.3 years   | 1.1 years   |

### Reliability and Validity

| Factor          | Group 1          |       | Group 2          |       |
|-----------------|------------------|-------|------------------|-------|
|                 | Cronbach's alpha | AVE   | Cronbach's alpha | AVE   |
| Explainability  | .725             | .627  | .701             | 0.631 |
| Transparency    | .714             | .671  | .854             | 0.665 |
| Fairness        | .785             | .626  | .797             | 0.558 |
| Accountability  | .727             | .7613 | .842             | 0.823 |
| Personalization | .796             | .7170 | .797             | 0.477 |
| Accuracy        | .802             | .8380 | .805             | 0.632 |
| Trust           | .905             | .8520 | .905             | 0.665 |
| Emotion         | .772             | .8196 | .768             | 0.812 |

### Scales and Measurements

Per the methods of mixed triangulation, we developed measurement items from the qualitative data as well as existing literature. Based on the qualitative analyses, possible measurement items were drawn and prototyped (27 items). These measurements were then compared and contrasted with the ones from existing literature that have been developed and validated in other studies. This two-step process was necessary to ensure reflecting user perspectives and to warrant consonant with existing measurements used in other AI services. The measurements of perceived humanness and FAT were based on the work of Shin (2021) and the explainability measurements were modified from research by Rai (2020). Trust measurements were derived from the works of Rai (2020). Fifteen users with prior experience using newsbots finished a pretest about a specific news topic. The purpose of the pretest was to identify potential problems for participants in either understanding or

interpreting a question. The pretest exactly replicated the actual experiment online. The pretest was necessary to evaluate the practicability of the FAT concepts in the newsbot context. The results of the pretest revealed problems relating to reading news, answering questions, and completing a survey, thus helping to point out weaknesses in the administration, organization, and execution of the instrument. Some wording on the concept of FAT was altered for normal laypersons to more easily comprehend. The sequence of components in the pretest was modified in the actual experiment to help respondents better understand.

Explanatory stimuli were given as a form of explanatory and anthropomorphic cues in newsbots. Explanatory cue phrases popped up when respondents clicked on recommended news. Explanatory cues provided users with relevant explanations on why specific news was recommended, such as “because you watched it before” or “most similar [to] watched news”.

The explanatory cues made it possible for respondents to understand, interpret, and experiment with newsbots. The explainability measures were derived from previous studies (Arrieta et al., 2020) and modified to suit the newsbot context.

## Results

### Qualitative Findings: Discovering User Heuristics

We extracted factors and grouped the related factors using semantic analysis into common factors. We identified a range of algorithmic factors based on the interpretive and dialectic methods, focusing on the user heuristics with algorithmic media (Sundar et al., 2020). Using qualitative methods that elicited participants' disclosure intentions in algorithmic news scenarios, we identify six distinct heuristics relevant to privacy and show that they are systematically connected with data disclosure.

Respondents were generally familiar with AI-related issues and key concepts: they understood that algorithms are widely used to curate news feeds and that algorithms bear the ethical issues of FAT. In algorithmic news, the ethical issue revolves around how to engage in practices of producing news articles through machine learning in ethical ways. One example is how to eradicate, or at least reduce, bias while upholding accuracy. Not everyone agrees on what constitutes FAT. There is no common standard for what level of transparency is enough transparency (Perdomo & Rodrigues-Rouleau, 2021). To be truly transparent, should journalism agencies publicly release the algorithm codes behind their news algorithms? To what extent do users need to understand explanations given for how newsroom algorithms work? These questions are recurring topics in our examination of the qualitative data.

**Fairness Heuristics.** The concept of fairness, along with accuracy, was touted during our interviews. As more decisions and processes of greater importance are being

made by algorithms in various contexts, people are becoming increasingly questioning the objectivity of algorithmic results. Some participants cited the example of AI judicial rulings, where algorithms are used for sentencing and could discriminate certain groups. Another respondent asked an important question, "How does algorithmically-informed journalism shape the stability of our democracy?" People voiced that fairness is an objective and impartial practice, without discrimination or favoritism. Respondents explained the concepts of non-discriminatory treatment, indiscriminination, and impartiality. One respondent mentioned neutrality, and others indicated instances of accurate results from the algorithmic filtering and recommendations to elucidate the notion of fairness. Unanimously, people agreed that fairness is a key attribute in algorithmic platforms by asking, "What emerging tools or approaches could mitigate opacity or problems of unfairness/bias?" One respondent put it, "I like [that] the algorithm system does not discriminate and has no favoritism". Another respondent mentioned, "I hope [that] the results of the algorithm system [are] accurate and bias-free". As for the due process, it was noted that "[an] algorithm system should follow [a] due process of impartiality with no prejudice". Based on participant responses, it is proposed that human perception of fairness is beyond mathematical fairness referring to contextual equality such as impartiality, indiscriminination, and accuracy.

### Accountability Heuristics

Algorithmic accountability in algorithms is the notion that platform providers should be held liable for considering platform design, development, and recommendation processes and outcomes. Typically, algorithmic accountability is viewed along with transparency, which mandates that providers be clear about the goals, structure, and underlying actions of algorithms used to collect, process, and deliver results. In this study, respondents expressed concerns that algorithmic decisions are vulnerable to making errors that may lead to unwanted consequences.

Respondents stated that algorithms are likely to create problems due to an inability to deal with bias or due to simple oversight. One respondent stated, “. . . bias and mistakes in algorithms can cause undesired and even hazardous problems”. Another respondent used an egregious example of AI’s facial recognition algorithm, which wrongly labeled some ethnic race as gorilla. A further example was raised involving self-driving vehicles, which were reported to have run a stop sign due to a faulty algorithm. It was concluded that firms using algorithms should somehow be held liable for the results of their programmed machine learnings. In cases where algorithms deliver discriminatory outputs due to bias embedded in data, the systems should be responsible for the harm done as a consequence of this discrimination (thus articulating participants’ definition of accountability in the context of algorithms). One respondent noted, “I wish I could examine and review the behavior of [the] algorithm system” (auditability). Another respondent claimed that algorithms should be equipped to modify a system using only certain manipulations (controllability). Based on this qualitative process, factors of responsibility are inferred—auditability, equity, legibility, and accountability.

**Transparency Heuristics.** Along with fairness, transparency has been considered a key dimension in platforms. As a growing number of everyday matters are powered by algorithms, there is more demand for higher transparency to mitigate potentially opaque processes that carry risks of biased profiling and discrimination. Since transparency itself is a hypothetical issue, people’s views are divergent. In general, people’s views can be categorized into three main areas—understandability, explainability, and observability (or visibility). Respondents in this study articulated that considering the magnitude of algorithms, the decisions automatized by AIs should be interpretable, explainable, and observable to the users who adopt, consume, and control the algorithms and algorithmic services. As to understand-

ability, one respondent stated that “Automated systems are increasingly complex, and they are often hard to understand for laypeople and for the people about [which] decisions are being made”. As to visibility, another respondent cited, “I think the design of algorithms is not visible and observable. A lot of these systems are designed by private companies and their details are proprietary”. Another put it, “It’s hard to know what they are doing and who is responsible for the decisions they make”. Others confirmed the need for explanation, indicating, “Any outputs produced by an algorithmic system should be explainable to the people affected by those outputs”. Based on this interpretative process, the identified factors of transparency are observability, expressivity, verifiability, and visibility.

**Heuristics of Explainability and Humanness.** Respondents see interpretability as the most important, but least touted feature of algorithms. Respondents described explainability as how the mechanisms and processes in the use of machine learning should be perceived by humans. Respondents expressed that algorithm can be rationalized in terms of how certain recommendations are produced in such a manner that humans can understand. It has been raised that some AI systems use algorithms based on black-box practice that laymen user can’t comprehend. Another respondent said, “in most current AI systems, there is no clear explanation of how the AI produced certain results or values”. Related to this, it is noted that “providing interpretable explanation can be critical in certain fields such as medicine, the judicial system, finance, public service, and healthcare”. Respondents raised a need for interpretability beyond explainability to make algorithms understandable to people. There was a discussion regarding transparency to whom and for what purpose and what comprises a transparent explanation as well as to what extent of transparency is needed.

Respondents expressed the idea of humanized explanations that they wished explanations could be provided by human-

like agents like chatbot news. Other respondents echoed these observations that they prefer interactive forms of human-agent explanations to text-based explanations. Many respondents expressed interpretable explanations that can be understood by people easily and they concurred that human-enabled explanations are the best way to comply with the right to explanation outlined in the GDPR. It was noted that while newsbots need to comply with the GDPR, there are currently no specific criteria for how algorithmic processes should be explained and to what extent algorithmic allow to collect user data.

### Heuristics of Trust, Privacy, and Data disclosure

Respondents discussed an idea of a two-step flow of communication with newsbots that most people form their opinions under the process of trust mechanism, which in turn are influenced by the algorithms. They emphasized that they are not directly determined by algorithmic messages, but the message flows from algorithms to them through trust, and from them to an actual usage intention. Trust mechanism passes on the trust evaluation of information in addition to the actual algorithmic content. Respondents understood a gatekeeping role by algorithms

and worried about filter bubbles, echo chambers, and polarizes public discourse.

Respondents were concerned about privacy issues as AI firms feed more and more consumer and vendor data into advanced, AI-fueled algorithms to create information. Most respondents concurred that data is important in AI operation and data is needed to train machine learning algorithms. Yet, many respondents feel uneasy when they do not understand how their personal data are garnered and used by algorithms. One respondent says, “I understand we must provide our data to AI in order to receive personalized content, but it makes me uncomfortable when I do not know how my data are handled”. Related to this concern, numerous respondents related privacy to trust implying the connection between privacy and trust, with regard to data protection. It is noted that data disclosure is directly influenced by trust, it is also indirectly influenced by privacy. As to perceived humanness, respondents expressed that anthropomorphized newsbots increase users’ sensitivity to the issue of privacy. It is noted that the more anthropomorphized features in newsbots, the more likely users’ privacy concern is mitigated, which leads to increased user data disclosure.

**Table 2** Model Fit Indices

| Fit statistics              | Group 1      | Group 2      | Criteria          |
|-----------------------------|--------------|--------------|-------------------|
| $X^2$                       | 1222.19      | 1222.19      |                   |
| $X^2/df$                    | 1222/241=5.0 | 1222/241=5.0 | < 5               |
| <i>p</i> -value             | .000         | .000         | < 0.05            |
| RMSEA                       | .05          | .094         | 0.05 to 0.10      |
| CFI                         | .830         | .826         | > 0.90            |
| RFI                         | .788         | .744         | $0 < x < 1$       |
| <i>Hoelter</i>              | 146 (0.01)   | 99 (0.01)    | $75 \leq x < 200$ |
| IFI                         | .828         | .828         | > 0.80            |
| GFI (Goodness of Fit index) | .967         | .968         | > 0.95            |
| AGFI (Adjusted GFI)         | .917         | .928         | > 0.90            |

### Quantitative Findings: Validating Heuristic and Rationalizing Processes

Consistent with the existing literature, the results of our structural equation analysis confirmed the paths drawn in our hypotheses (Table 3). All the path coefficients were significant at  $p$ -values  $< 0.05$  and  $< 0.001$ , which confirm the predictive validity of the model, reflecting a strong endorsement of the cognitive processing—explainability-normative values-trust-privacy-data disclosure. Trust is significantly affected by FAT and at the same time influences privacy. These factors together account for 57% of the total variance

in the trust variance of the model. Perceived humanness significantly influences explainability and trust. It is confirmed that mediating effects of privacy on the path of trust to data disclosure. Overall, the model explained a significant portion of the variance in the data. The R-square indicates that the model accounts for 55.5% ( $p < .001$ ) of the variance in data disclosure, 60.8% ( $p < .001$ ) of the variance in trust, and 32% ( $p < .001$ ) of the variance in privacy. The statistically validated relationships reveal a conceptual mapping of explain ability, humanness, privacy, and trust.

**Table 3** Summary of Hypothesis Testing

| H   | Group 1 (Explanatory Cue) |        |        |      | Group 2 (Anthropomorphic Stimuli) |      |        |      |
|-----|---------------------------|--------|--------|------|-----------------------------------|------|--------|------|
|     | $\beta$                   | S.E.   | C.R.   | $p$  | $\beta$                           | S.E. | C.R.   | $p$  |
| H1  | .066                      | .041   | 1.592  | .111 | .684***                           | .050 | 13.727 | .000 |
| H2  | .037                      | .050   | .750   | .454 | .106***                           | .032 | 3.301  | .000 |
| H3  | .261***                   | .052   | 5.012  | .000 | .542***                           | .044 | 12.288 | .000 |
| H4  | .390***                   | .071   | 5.486  | .000 | .551***                           | .056 | 9.900  | .000 |
| H5  | .952                      | 11.918 | .751   | .453 | .705*                             | .234 | 3.015  | .003 |
| H6  | .208***                   | .053   | 3.937  | .000 | .414***                           | .056 | 7.385  | .000 |
| H7  | .572***                   | .052   | 10.917 | .000 | .670***                           | .054 | 12.387 | .000 |
| H8  | .607***                   | .060   | 10.153 | .000 | .669***                           | .063 | 10.656 | .000 |
| H9  | .446***                   | .059   | 7.496  | .000 | .350***                           | .061 | 5.780  | .000 |
| H10 | .203*                     | .075   | 2.716  | .007 | .111                              | .077 | 1.431  | .152 |

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

$\beta$ : Unstandardized Coefficient

### Squared Multiple Correlation Comparison

| Factor          | Group 1 | Group 2 |
|-----------------|---------|---------|
| Fairness        | .296    | .498    |
| Accountability  | .148    | .589    |
| Transparency    | .012    | .451    |
| Trust           | .596    | .771    |
| Accuracy        | .391    | .483    |
| Personalization | .525    | .588    |
| Emotion         | .394    | .511    |

**Note:** The result of the multicollinearity test shows no signs of a multicollinearity problem.

### **Can Explainability Solve the Credibility Problem in Newsbots?**

The findings show that explanatory and anthropomorphic cues are effective in triggering privacy heuristics and eliciting users to disclose personal data. Each cue triggers a specific cognitive heuristic that offers a rationale for the decision of data disclosure. The results of this study provide a point of reference for newsbots by establishing a foundation for understanding the relationships among explainability, privacy, normative values, and trust. The relationships constitute users' algorithmic information processing when interacting with newsbots. While our results imply a need for transparent personalization that complies with data privacy requirements, the significance of the findings can be discussed from three perspectives.

First, the significant relationship between privacy and data disclosure with reference to trust reveals new insight on newsbots, particularly the conceptualization of privacy heuristics in algorithms. Despite the convenience and benefits of newsbots, users are reluctant to disclose their information due to privacy concerns, which hinders the development of user-centered algorithms within newsbots (Fast & Jago, 2020). The findings confirm that privacy is a strong determinant of data disclosure and the effect is moderated by trust. The results also show that anthropomorphized features influence privacy, suggesting that incorporating anthropomorphic attributes newsbots can mitigate users' perceptions of privacy risk and increase their intention toward newsbots (Go & Sundar, 2019). While prior research has confirmed the relations of trust and privacy extensively (Ischen et al., 2020), and the effect of privacy on data disclosure (Benson, Saridakis, & Tennakoon, 2015), mediating role of privacy remains unknown in newsbot contexts. The finding fills the gap by illustrating the mediating roles played by privacy heuristics between trust and data disclosure and between anthropomorphism and data disclosure. The privacy heuristics can

be useful for the future design and assessment of newsbot interfaces in terms of explanations and data disclosure (Wang et al., 2020).

Second, we examined how explainability influences value, trust, and performance through a lens of two-step flow. The two-step flow finding suggests that human-like mediators pass on their interpretation of messages to readers (Soffer, 2019). Unlike the common view of algorithmic determinism, the two-step flow finding proposes the active role played by users in the interaction process with newsbots, trying to understand the algorithmic process consciously, attempting to influence the algorithmic decision actively, and striving to make the interaction human-like and trustable. The algorithmic interaction is not a single-step flow in which algorithms communicate directly to the users, but is influenced by cognitive processes and enhanced by human-like agents. Attitudes towards algorithms are context-dependent and users first rely on an intuitive appraisal of FAT, which affects trust and then deliberate evaluation on privacy. Previous research has limited explanatory power in characterizing trust relations because they simply consider technical/visual rather than behavioral/cognitive algorithmic stimuli. Our results echo previous work (Araujo, 2018; Shin, 2021; Soffer, 2019) and further the proposition that users engage with newsbots through trust-based humanized explanations. While our results are consonant with ongoing literature (Westerman, Aaron, & Lindmark, 2019), they further that perceived humanness correlates with positive privacy. The existence of anthropomorphized cues in newsbots activated positive heuristics of trust and legitimacy in journalism. One factor in user trust in newsbots is the degree to which a system is perceived as anthropomorphic, which also triggered positive valences of privacy. This indicates that anthropomorphic cues not only activate explainability but also increase user privacy, which leads to data disclosure. This finding provides a baseline understanding of the relations of explainability, humanness, and privacy by clarifying how explainability can

be effective, what role of anthropomorphism play, and how users make a decision on data disclosure (Padyab et al., 2019).

Third, the findings provide empirical evidence of users' algorithmic information processes by proposing trust and privacy as an intermediary in the algorithmic interaction. Readers' cognitive process of newsbots is heuristic, contextual, and interactive, as opposed to being prearranged with ready-made programming processes. The processes are not homogenous and may change across different modalities of algorithmic interactions and different levels of trust embedded in algorithms. Algorithmic information processes are consonant with the arguments of the two-flow of interaction that users are not passive recipients of messages recommended by algorithms, but actively construct how they will see the messages and decide how they will use the algorithms by controlling their privacy conditions and by determining the personalization level they would like to see. Unlike the previous information processing perspective about how the user cognition functions, which merely examines how users respond to stimuli, algorithmic processing theory furthers that a useful heuristic is a part of the algorithm or AI. The message is collected through the input devices and processed via a trust mechanism, resulting in evaluating privacy risks. How users appreciate, understand, analyze, and respond to algorithmic information follows two distinctive steps, where an initial assessment of normative values occurs, and subsequently trust and privacy evaluation is done. As algorithms function as gatekeeping agents in lieu of opinion leaders, trust exerts a facilitating role in the two-step flow process. Trust between algorithms and readers plays gatekeeper and liaison roles over the heuristics and reasoning of newsbots.

### **Privacy Heuristics That Determine Disclosure**

We examined users' levels of algorithmic personalization and the effects of such attributes on trust and privacy. Theoretically, this study clarifies the information processes of algorithms

together with the liaison of trust from the perspective of user heuristics (Ferrario, Loi, & Vigano, 2020). We provide a more nuanced conceptualization of the processes along with empirical insights from a mixed-method study. Practically, the findings of the study have editorial guidelines as to what newsbot newsrooms should do in order to forge high-quality accountable news reporting—specifically in terms of how to represent explanations and how to embed FAT in AI interfaces to improve the trustworthiness of algorithmic media.

### **Theoretical Implications: Algorithmic Information Processing Theory**

This study makes conceptual and methodological contributions to the extant literature about trust, privacy, and explainability in the context of newsbots. Our arguments confirm the heuristic roles of explainability to readers who seek to know the reasoning behind the recommendations and predictions of algorithmic news. We explain the antecedents of and associations among algorithmic attributes, highlighting the information processing of these antecedents and trust, clarifying how explainability is applied and how it exploits different social cues and showing the roles of privacy in leading to data disclosure. Explainability constitutes a precondition for readers to trust the information newsbots present in news personalization contexts (Rai, 2020). These findings are important because algorithms have even more become a vital factor in the journalism news-making process; thus, how users value algorithmic news and thus react to recommendations becomes a priori question to address (Shin, Chotiyaputta, & Zaid, 2022).

While existing research has shown the relations of normative values and trust (Shin, 2021), and the mechanisms of anthropomorphic effects (Go & Sundar, 2019), the relation of privacy and data disclosure has not been extensively examined in a specific context. The mediating role of privacy in the path of trust to data disclosure is not only meaningful

but also constitute theoretical development in algorithmic information processing theory through conceptual refinement of how algorithmic trust is formed (Shin, 2021), how it is enhanced by humanness (Westerman, Aaron, & Lindmark, 2020), how it is mediated by privacy (Ischen et al., 2020), and what anthropophilic effects are present in newsbot readers (Araujo, 2018). While ongoing research efforts have advanced to explain AI models, approaches to assess the explainability of actual news services from a user cognitive perspective are currently lacking in the literature. We extend the information processing literature by examining the interactions between the heuristic process and the systematic process. Focusing on different aspects of the user decision-making process, we illustrate the algorithmic conditions under which different types of processes are evoked, how they interact and how they apply to users' processing of trust, the evaluation of privacy, and data disclosure behavior. Previous research using dual-process theory has shown the contexts under which one process plays a more prominent role than the other process (Wang et al., 2020). Understanding how the heuristic process interacts with the systematic process provides new insights into algorithmic information processing.

The interaction of the two processes reveals the two-step algorithmic flow in the model, which expands the literature on user

heuristics with newsbots—specifically the users' sense-making process and how automated algorithms help them with this process—by characterizing the role of privacy-trust and by structuring the causal relationships among its closely associated measures (Thurman et al., 2019). Understanding the extent to which users proactively manage sequences and processes in algorithmic choices has key implications for theories and design in newsbots (Dörr & Hollnbuchner, 2017). Findings on the role/process of explanations and the relationships among its associated measures not only confirm the theory's key proposition—that cognitive decision is guided through a two-track process (Wang et al., 2020)—but also advance the theory by linking this process to the two-step flow of communication (Soffer, 2019)—the dual process by which users create an understanding so that they can act in an informed and principled way. Users actively process and proactively control stimuli and algorithmic curations, evaluating privacy based on normative values and trust. Against the increasing black-box nature of algorithms, users want to understand how algorithms function, and how to protect against fake news/disinformation. The two-step flow mechanism furthers the understanding of how algorithm influences decision-making by partitioning the user cognitive process into dual-step (1) evaluating algorithmic values and (2) algorithmic credibility (Figure 2).



Anthropomorphized Chatbot Services

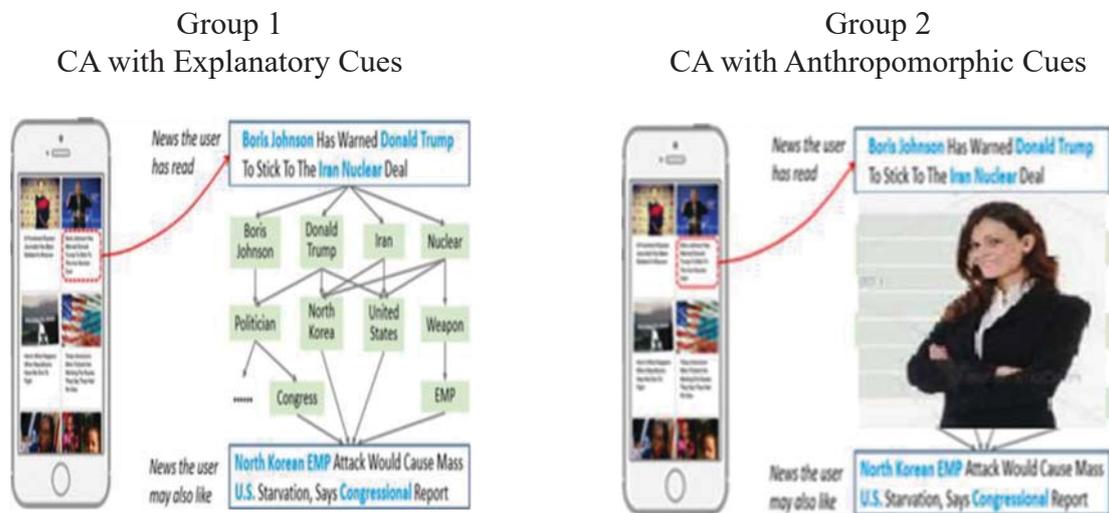


Figure 2 Examples of Chatbots used in experiments

### Practical Implications

Journalism today faces multiple and imminent challenges brought by AI technologies. Journalism is struggling to balance between safeguarding user privacy and making the best use of the impact of personalization. When algorithms misuse user data or process them in any way that would make negative effects on users, the algorithmic news loses credibility and trust. As user expectations continue to outrun journalisms' efforts to be personal, journalisms face due challenges for what users think is an acceptable use of personal data and what measures should be done to gain user

trust and privacy. The main barrier preventing the implementation of user privacy is a lack of understanding of user cognition pertaining to privacy-related decision-making in algorithm contexts. Our results contribute to the issue of explainability in newsbots by presenting how the issue is interwoven with privacy, clarifying how it can be applied, and illustrating how the effects of explainability are measured.

Against the backdrops of the journalism industry facing decreasing credibility, the functions of explainability will certainly offer insights into user trust in algorithms and further improve the credibility of journalism. So far,

the industry would not reveal or share inner algorithms with the public as there are no clear incentives for doing it and there will be possible harms and backfires such as revealing editorial prejudice. Yet soon they will realize opening some of their algorithms at least because the benefits of transparency will outweigh the disadvantages. Just like open-source software has become a norm for software companies these days, opening up algorithm codes not only helps the readers of the news to gain an understanding of the innovative algorithms used by the journalism but also helps the journalism to find a way to improve the quality of news on its platform and to increase the credibility of news sources that are utilized by readers. From a practical standpoint, explainable newsbots may be a necessary innovation insofar as it is sure to more effectively leverage the empowering effects of algorithms, and may also allow users to incorporate legitimate content and avoid fake news or misinformation. Thus, the explainability-privacy mechanism can be proposed as a standard of trustworthy newsbots. The two-step flow of interaction triggered by heuristic explainability leads important paths in how to apply explainable principles in journalism contexts to promote user trust and privacy. Conceptualizing the mechanism of explainability-privacy in algorithms has the potential to make practical contributions to ongoing efforts in realizing explainable journalism (Mann, 2016). Making robot-driven news more explainable is key to correcting the errors which inadvertently lead to bias. An understanding of cognitive heuristics can be valuable for designing interfaces that promote secure and explainable journalism.

## Conclusion

For all the hype about algorithm automation, it is still unsure of how newsbots evolve in what forms (Thornham, 2019). The

prospect of newsbots lies in the reconstruction of the underlying norms of journalism, which divides journalistic practices into the actual news delivery and processes to analyze what and how can be algorithmicized and what are user roles fundamentally. Our results show that explanatory heuristics play a critical role in setting up an essential trust and privacy heuristics exert a mediating role in triggering data disclosure. Together these results highlight the active user role in constructing algorithms that algorithm users are not directly influenced by AI nor passive users of the algorithms and instead form their opinions based on trust heuristic, which leads to evaluate algorithmic qualities and put them into context. Our results further the significance of explainability in newsbots, not only in providing an interpretable mechanism for AI but also for users assessing normative values in providing clues and facilitating the privacy heuristics of the results of newsbots by unlocking the black box of news algorithms.

## Future Studies

The critical challenge for newsbots is how to make use of algorithms in a way that increases users' trust and improves journalistic practice without compromising on quality or credibility. The relationships validated through our results serve as a step toward developing a theory for more user-centered algorithms. Our results are in line with general assumptions that news readers' experiences with newsbots stand to benefit from humanized explanation. Based on the heuristic nature of the findings, we claim that there is potential in this area of research and propose the following propositions for future research. Future studies in algorithmic news recommendations should address ethical issues such as the lack of gatekeepers, the inability to evaluate the social value and authenticity of content, the information cocoon room, and echo room effects.

## References

- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology, 30*(4), 736-758.
- Arrieta, A. et al. (2020). Explainable artificial intelligence. *Information Fusion, 58*, 82-115.
- Araujo, T. (2018). Living up to the chatbot hype. *Computers in Human Behavior, 85*, 183-189.
- Benson, V., Saridakis, G., & Tennakoon, H. (2015). Information disclosure of social media users. *Information Technology & People, 28*, 426-441.
- Caswell, D., & Dörr, K. (2018). Automated journalism 2.0. *Journalism Practice, 12*(4), 477-496.
- Crain, M. (2018). The limits of transparency. *New Media & Society, 20*(1), 88-104.
- deVito, M. (2017). From editors to algorithms. *Digital Journalism, 5*(6), 753-773.
- Diakopoulos, N. (2019). *Automating the news. How algorithms are rewriting the media*. Cambridge, Massachusetts: Harvard University Press.
- Dienlin, T., & Metzger, M. (2016). An extended privacy calculus model for SNS. *Journal of Computer-Mediated Communication, 21*, 368-383.
- Dörr, N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism, 5*(4), 404-419.
- Fast, N., & Jago, A. (2020). Privacy matters or does it? *Current Opinion in Psychology, 31*, 44-48.
- Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally. *Philosophy and Technology, 33*, 529-539.
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H. (2018). Readers' perception of computer-generated news. *Journalism, 19*(5), 595-610.
- Go, E., & Sundar, S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior, 97*, 304-316.
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. (2019). Explainable artificial intelligence. *Science Robotics, 4*(37), 7120.
- Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. In A. Følstad et al. (Eds). *Chatbot Research and Design. Lecture Notes in Computer Science* (pp. 34-48). New York: Springer.
- Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism. *Computers in Human Behavior, 71*, 291-298.
- Keith, M., Thompson, S., Hale, J., Lowry, P., & Greer, C. (2013). Information disclosure on mobile devices. *International Journal Human-Computer Studies, 71*, 1163-1173.
- Kim, J., & Gambino, A. (2016). Do we trust the crowd or information system? *Computers in Human Behavior, 65*, 369-379
- Koohikamali, M., Peak, D., & Prybutok, V. (2017). Beyond self-disclosure: Disclosure of information about others in social network sites. *Computers in Human Behavior, 69*, 29-42.
- Mann, T. E. (2016). *Explanatory journalism: A tool in the war against polarization and dysfunction*. Washington, D.C.: Brookings Institute.
- Montal, T., & Reich, Z. (2017). I, robot. you, journalist. who is the author? *Digital Journalism, 5*(7), 829-849.
- Özkul D. (2021) The algorithmic fix: Location intelligence, placemaking, and predictable futures. *Convergence, 27*(3), 594-608.

- Padyab, A., Päivärinta, T., Ståhlbröst, A., & Bergvall-Kåreborn, B. (2019). Awareness of indirect information disclosure on social network sites. *Social Media + Society*, 4, 1-14.
- Perdomo, G., & Rodrigues-Rouleau, P. (2021). Transparency as metajournalistic performance: The New York Times' Caliphate Podcast and new ways to claim journalistic authority. *Journalism*, 2, 1-17.
- Pietrobruno, S. (2016). YouTube flow and the transmission of heritage. *Convergence*, 24(6), 523-537.
- Rai, A. (2020). Explainable AI. *Journal of the Academy of Marketing Science*, 48, 137-141.
- Shin, D., Chotiyaputta, V., & Zaid, B. (2022). The effects of cultural dimensions on algorithmic news. *Computers in Human Behavior*, 126, 107007.
- Shin, D. (2021). *The perception of humanness in conversational journalism: An algorithmic information-processing perspective*. Retrieved August 16, 2021, from <https://journals.sagepub.com/doi/abs/10.1177/1461444821993801>
- Soffer, O. (2019). Algorithmic personalization and the two-step flow of communication. *Communication Theory*, 14, 6304-6323
- Sundar, S., Kim, J., Beth-Oliver, M., & Molina, M. (2020). Online privacy heuristics that predict information disclosure. *CHI' 20*, 4, 1-12.
- Thornham, H. (2019). Algorithmic vulnerabilities and the datalogical. *Convergence*, 25(2), 171-185.
- Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I. *Digital Journalism*, 7(4), 447-469.
- de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology*, 22, 331-349.
- Wang, L., Hu, H., Yan, J., & Mei, Q. (2020). Privacy calculus or heuristic cues? *Journal of Enterprise Information Management*, 33(2), 353-380.
- Westerman, D., Aaron, C., & Lindmark, P. (2019). I believe in a thing called bot. *Communication Studies*, 70(3), 295-312.
- Wölker, A., & Powell, T. (2020). Algorithms in the newsroom? *Journalism*, 22(1), 86-103.