KARPETA

Teknokultura. Revista de Cultura Digital y Movimientos Sociales
ISSNe: 1549-2230

EDICIONES COMPLUTENSE

https://doi.org/10.5209/tekn.87552

The epistemic ethical concerns involving algorithms in intelligent communication

Jialin Lin[1]; Changfeng Chen[2]

Received: March 12, 2023 / Accepted: May 22, 2023 / **OPR**

Abstract. With the development and application of algorithms as catalysts, the changing modes of information production, dissemination, and consumption have also given rise to a myriad of serious ethical challenges. This study employs a multi-case approach and semi-structured in-depth interviews to examine three prominent international information technology companies, namely, Meta, Sina, and Byte Dance. By investigating the utilization of algorithms in content creation and distribution and adopting an epistemic, ethical framework, this paper analyzes the phenomenon of information cocooning resulting from inconclusive algorithmic evidence, the presence of algorithmic black boxes stemming from inscrutable evidence, and the issue of algorithmic bias caused by misguided evidence. Consequently, this paper proposes three fundamental ethical principles for algorithmic systems: certainty, interpretability, and reliability.
Keywords: algorithm bias; algorithmic black box; algorithmic ethics; information cocoon; interpretability.

[es] Consideraciones ético-epistémicas en la comunicación inteligente: El rol de los algoritmos

Resumen. El desarrollo y la aplicación de los algoritmos ha traído consigo notables transformaciones en los modos de producción, difusión y consumo de información al igual que desafíos éticos de gran calado. Este estudio utiliza el estudio de casos múltiples y entrevistas en profundidad semi-estructuradas para examinar tres empresas internacionales de tecnología de la información: Meta, Sina y Byte Dance. Desde un marco ético-epistémico analizamos el fenómeno del aislamiento informativo resultante de pruebas algorítmicas inconclusas, la presencia de cajas negras algorítmicas derivadas de pruebas incomprensibles, y el problema del sesgo algorítmico causado por pruebas equivocadas. En consecuencia, este artículo propone tres principios éticos fundamentales para los sistemas algorítmicos: certeza, interpretabilidad y confiabilidad.
Palabras clave: aislamiento informativo; caja negra del algoritmo; ética algorítmica; explicabilidad; sesgos algorítmicos.

Table of contents.

How to cite: Lin, Jialin and Chen, Changfeng (2023). The epistemic ethical concerns involving algorithms in intelligent communication. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales, 20*(Special Issue), 23-36. https://doi.org/10.5209/tekn.87552

1. Introduction

Artificial intelligence has profoundly altered the way we produce and disseminate the news. Algorithms and big data represented by deep learning and its derivative technologies have transformed information exchange and the public opinion ecology for individuals, societies, countries, and the world. On this basis, the concept of 'intelligent communication' has emerged.

Intelligent communication is built upon developments and breakthroughs in new energy, artificial intelligence (AI), big data, the Internet of things, robotics, and other technologies. It involves the placement of machines as nodes replacing traditional production factors, such as manpower, technology, and equipment, with the Internet of things and data streams. Guided by users and markets, intelligent technology is used to determine the direction of content production, distribution, marketing, and integrated decision-making in the communication model (Meng, 2018).

Solving problems with the help of algorithms has become a preferable option, as it circumvents human cognitive limitations and executive ability. Concomitantly, the problems of algorithms and their impacts on the public are also projected into current production and the lives of individuals, raising significant ethical questions. Therefore, this paper attempts to identify the possible risks produced by the development and application of algorithms in content production and distribution. As such, several multi-

[1] Sun Yat-Sen University (China)
E-mail: linjlin7@mail.sysu.edu.cn
ORCID: https://orcid.org/0009-0008-9655-488X
[2] Tsinghua University (China)
E-mail: fengchen5266@163.com
ORCID: https://orcid.org/0009-0003-2276-4749

case studies and semi-structured in-depth interviews were conducted to gain insight into the ethical issues associated with the risks of algorithms and address the algorithm trust crisis.

2. Literature review

Algorithmic ethics is a new interdisciplinary topic related to information, technology, and humans in the era of intelligent communication. However, current research on this topic is limited, unsystematic, and fragmented between disciplines, with no clear way to unify interdisciplinary research within a holistic framework (Mulvenna et al., 2017). Furthermore, questions regarding what can be understood as proper AI and what kind(s) of algorithms can truly produce a flourishing future for humans in a broad sense are difficult to answer (Acemoglu & Restrepo, 2020).

Initially considered a mathematical concept, algorithms were introduced into computer science in the middle of the last century under the influence of production practices. They are currently being further integrated into various areas. The etymology of the word algorithm can be traced back to the name of the Arab mathematician Mohammed ibn Musa al-Khwarizmi (Bagdad, 780?, 850) (Zhang, 2021, p. 3). The earliest use of the word 'algorithm' appeared in Spain in the 12th century. Unlike general algebraic formulas, an algorithm, as an abstract mathematical structure, is characterized by the irreversibility of assignment or instruction (Miyazaki, 2012).

Along with the birth of computers, the term was introduced into computational science. At present, algorithms are widely used in various production fields. Different disciplines have developed their own understanding of the concept, which has multiple meanings (e.g., mathematical, productive, social, and philosophical). Existing algorithms have been designed to follow statistical thinking (Mullainathan & Spiess, 2017). However, in terms of specific conceptual definitions, a consensus on the definition of the term has yet to be achieved. In the academic domain, different disciplines have divergent algorithmic perceptions and properties.

In production and real life, various types of groups have different perceptions and understandings of algorithms. The ambiguity of the concept also hinders further exploration of algorithms and their risks. This paper regards an algorithm as a comprehensive and integrated concept in the sense of mathematics, computer science, and production. We explore the ethical concerns, risks and possible hidden dangers arising for humans and society when algorithms applying statistical models are employed as the basic mathematical logic, as computer thinking models, and the practical outcomes for news production and dissemination.

Previous studies on algorithm risk highlighted six major issues: information cocooning, the black box effect, algorithm bias, algorithm discrimination, the digital divide, and infringement on the right to information. Computer science studies typically discuss the technical causes of problems in terms of model optimization, thereby exploring only the technical solutions (Calders et al., 2009; Hajian et al., 2016; Mittelstadt et al., 2019; Sun et al., 2019). Social disciplines, such as journalism and communication on the other hand, focus on specific risks caused by algorithm use in social practices (i.e., news production and dissemination), analyze the manifestations and potential harms, and conduct inquiries applying ethical and philosophical frameworks (Schermer, 2011; Jalonen, 2012; Macnish, 2012; Barocas, 2014; Baldini et al., 2018).

Current algorithm research within applied ethics mainly focuses on empirical investigations of the demonstration phenomenon and speculative discussions regarding such investigations, including ethical anomie and discussions surrounding AI norms based on these phenomena. Brent Daniel Mittelstadt (2016) proposed a conceptual map classification method, identifying six issues related to algorithm design logic. This conceptual map suggests three ethical dimensions based on the process of algorithm application: epistemic ethics, normative ethics, and algorithm traceability. These three ethical dimensions are further subdivided into six categories. Epistemic ethics, which corresponds to the machine learning stage, is the first stage of algorithm development and application. It is related to the following three types of ethical concerns: 1) inconclusive evidence, which leads to unjustified actions due to the uncertainty of algorithmic logic based on probability; 2) inscrutable evidence, which leads to algorithmic opacity, along with unrecognizable algorithm models; and 3) misguided evidence, which leads to algorithm bias caused by the poor-quality data that machine learning models rely on.

This conceptual map effectively encompasses the issues related to algorithms within AI technology applications and provides a theoretical tool for organizing and summarizing the broad and fragmented academic debates surrounding algorithm technologies, issues, and ethics. However, room for further refinement of this research topic remains. This theoretical framework and its verification are based on the discursive framework and technological practices of the Western world. A study on the WeChat platform showed that China's technological practices are different from the Western model of internet platforms adopted by media platforms such as Google and Meta (formerly known as Facebook) (Plantin & De Seta, 2019). The present study seeks to develop an ethical reflection grounded in the academic literature rather than first-hand data from field work. Furthermore, it has been five years since the publication of this study (Platin & De Seta, 2019), during which time algorithm technology has undergone continuous development, such that further validating the reliability and precision of this concept map in both Chinese and Western contexts becomes imperative.

This paper aims to analyze the practical risks presented by algorithm application in news content production and communication, and to reshape perceptions regarding algorithm ethics associated with algorithm use in intelligent communication by studying the causes of these risks.

3. Method

In this paper, we use a multiple-case study approach to analyze the algorithm applications and issues of three large tech companies involved in intelligent communication. The three selected cases were chosen following theoretical sampling. Compared with random sampling, theoretical sampling selects cases that are more suitable for elucidating and extending the relationships and logic between structures (Eisenhardt & Graebner, 2007).

To reduce omissions in the identification of ethical issues and to test the scientific validity of the conceptual map proposed by Mittelstadt et al. (2016), a semi-structured in-depth interview involving fourteen practitioners from ten companies that apply algorithmic technologies for production was conducted.

4. Epistemic ethical concerns of algorithms

4.1 Inconclusive evidence of algorithms

The algorithms currently used in content production and dissemination mainly rely on supervised machine learning, in which «software programs take as input training data sets and estimate or 'learn' parameters that can be used to make predictions on new data» (Athey, 2017, p. 483). This is the principle of algorithmic decision-making. Following this principle, algorithmic programs sift through a large number of variables in the dataset to find the combination of variables with the highest reliability for predicting outcomes. To predict reliable outcomes, the dataset used for training is always large, with extremely diverse, complex, and even jumbled content. Notably, to obtain results from big data, we need to use reliable algorithms as information bridges. However, algorithms can 'overfit' the prediction to spurious correlations within the data, and multicollinearity correlation predictors may produce unstable estimates. Any of these factors can yield overly optimistic readings of the model's accuracy and exaggerate real-world performance (Obermeyer & Emanuel, 2016). As such, pure predictions, which rely on big data and probabilistic calculations, present inherent limitations.

Mittelstadt et al. (2016) used the term 'inconclusive evidence of algorithm' to describe this kind of limitation. When the algorithm draws conclusions from the data obtained by inferential statistics or machine learning, it will inevitably produce inconclusive knowledge. Although the process of algorithmic decision-making is based on accurate calculations, algorithmic decision-making and data mining rely on correlations between data. The final result is formulated according to the probability of correlation. Therefore, the result of this calculation relies heavily on probabilities but is not necessarily precise.

4.1.1 Algorithm risks caused by inconclusive evidence

When tech companies use algorithms for content distribution, they usually determine the social identities of various subjects. They precisely target information at specific user profiles, wherein the non-deterministic nature of algorithms leads to a high degree of information homogeneity for specific users, ultimately 'encasing' them in an invisible cocoon generated by their personal data. This phenomenon called the 'information cocoon', is often referred to as the 'filter bubble'. The recurrence of extreme information can affect user perceptions of the world. Among the cases selected for this paper, two companies have generated information cocoons in the process of content production and distribution.

In Case 1 (Byte Dance - 2017), the client of Today's Headlines replaced the gatekeeper in the traditional content distribution process with personalized recommendations. This change increased the concentration of the same or similar content received by users, causing them to reinforce their inherent biases and preferences through constant repetition and self-verification, leading to a self-produced cocoon, or 'information island'. In Case 2 (Meta - 2017–2021), even after knowing that without human intervention the platform's recommendation system would lead to increasingly homogeneous information received by the same user, Facebook still allowed the system to recommend content to users solely through its algorithm. As a result, the concentration of homogeneous information reached a point where users felt their worldview was distorted.

The information consumption model of information cocoons has two characteristics. The first is the highly homogenous information that results, leading to a homogenization of cognitive content (Case 1: Byte Dance - 2017) and a homeostatic repetition of cognitive approaches (Case 2: Meta - 2017–2021). Homogenization is a major driver of content proliferation in this space. The second characteristic of the information consumption model of the information cocoon is the intense polarization of perspectives (Del Vicario et al., 2016), which manifests as a radical, unidirectional emotional vector (Case 2: Meta - 2017–2021).

In terms of cases, there are two reasons for the emergence of information cocoons. The first is technical, that is, the role of the algorithm itself. Case 2 (Meta - 2017–2021) shows that once human intervention is lost, homogenized information will follow a certain law to keep spreading. The creation of information cocoons also reflects institutional factors of the platform. Both Case 1 (Byte Dance - 2017) and Case 2 (Meta - 2017–2021) demonstrate the influence of the platform's inaction or intentional behavior towards the formation of information cocoons.

When information cocooning affects particular, sensitive issues, it can lead to further negative outcomes. Selective exposure to content is the main driver of content diffusion. The presence of the echo chamber effect reinforces users' exposure to homogeneous content, which, in turn, reinforces the effect of rumor propagation (Del Vicario et al., 2016). Echo chambers are significant determinants of information exchange relationships. The reputation of the information publisher has a clear impact on information dissemination in echo chambers (Malkamäki et al., 2021). Science and policy can better control outbreaks by addressing sources of uncertainty and missing data (Taylor, 2020). Government agencies, as representatives of public power and authority, should play a key role in rumor confirmation and disinformation efforts to avoid the further spread of undesirable information.

4.1.2 Two dimensions of inconclusive evidence limitations

The limitations of inconclusive evidence have two dimensions. The first is related to algorithm dependence on statistical thinking. Statistical thinking follows a basic law – more predictors than observations (Mullainathan & Spiess, 2017). Regardless of how complex the dataset is, it is only a group of samples drawn from the real world. According to the calculation logic of statistics, the samples can only be infinitely close to the population but never completely equal to it. In semi-structured in-depth interviews, one of our interviewees working in a tech company focusing on content production and distribution described the impact of dataset quality problems on modeling quality:

> Taking my field of work as an example, its technical defect may be that it is relatively dependent on data. However, it is difficult for us to exhaust the data when establishing the database. For example, we need to do this vectorization and 3D semantic modeling. If this modeling is data-driven, it needs a very large model library for support. However, there are too many things in the real world, and the model library cannot exhaust everything, which will greatly impact the final effect (Interviewee 07, F. Gill).

The second limitation reflects the dependence of the algorithm on data processing. Given that the output can never exceed the initial input, algorithm information processing is always limited by the original data it is learning from. As such, its reliability can only be equal to or less than the original data's reliability. The conclusion of an algorithmic decision may be predictive of the outcome, but it can never be completely predictive. Probability-based calculations yield correlations, not causality. Correlation can indicate whether there is a statistically significant connection between variables, i.e., how likely things are to be related, but it does not represent a direct causal relationship. In other words, the algorithm does not conclusively connect things with a solid line. The determination of causality requires mathematical calculations more complex than those statistics can realize. At the same time, when predictive analysis of things is required, the reliability of the conclusion's output by the algorithmic logic of correlation is always doubly uncertain (Ananny, 2016). When an algorithm uses big data as its learning base and correlation as the analysis logic to build the decision model, its strong correlation applies to the group rather than the individual. In other words, the output is only valid for a group of people and cannot be read as an accurate description of individuals.

4.2 Inscrutable evidence of algorithms

The inscrutable evidence of algorithms refers to the property that makes them difficult to understand. The intelligibility of such algorithms includes both comprehensibility and accessibility, which are also the most fundamental constituent elements of transparency (Mittelstadt et al., 2016).

The algorithmic black box refers to how algorithms are unknown to the general public due to the opacity of the technology itself and the opaque commercial policies of media organizations and tech companies. The algorithmic black box can be understood on two levels. One dimension is the technical level of the algorithmic black box, which refers to the comprehensibility of the algorithm. Correspondingly, there is a concept of algorithmic white box, which refers to the algorithmic model that can be understood. The other dimension is the accessibility of the algorithm, which refers to the cognitive barriers set by the platform for the protection of state secrets, trade secrets, or other motivations.

4.2.1 The comprehensibility of inscrutable evidence

The comprehensibility of an algorithm is a problem that technical practitioners generally experience the most. Theoretically, when data are used as the basis for generating conclusions for algorithmic decisions, it implies a necessary connection between the data and the conclusions, a connection which should in theory, be transparent. Even when it is not obvious and cannot be directly observed, we should be able to understand the connection through more effective or further explanatory efforts. The puzzling nature of algorithms reduces this comprehensibility, and we must confront the challenges generated by algorithmic black boxes. In our semi-structured in-depth interviews, Interviewees 07, 08, and 10 all discussed this issue. According to the current state-of-the-art, algorithm engineers have not yet found ways to white box each step of the algorithm, which further hinders them from effectively optimizing the entire algorithm model and affects the quality of algorithmic decisions.

> Since we do all this work based on deep neural networks, these multi-layer, deep-neural network-based models are a black box in themselves. Since this black box model involves a lot of computation, it is very difficult to find a definitive explanation for how the final results are derived, in many cases (Interviewee 08, M. Henry).

4.2.2 The accessibility of inscrutable evidence

The second dimension of inscrutable evidence is its accessibility. The accessibility problem is mainly relevant for the platform user group rather than the platform and algorithm engineers who control the algorithm. Specifically, this issue is relevant for unspecified groups without the necessary control over the specific cause of the problem when algorithm output is found to be faulty. In this case, when people cannot understand a certain function by directly linking the design model of the algorithm application, they are likely to invoke their own life experience to find a specific connection and use this as the basis for understanding the algorithm and determining the answer.

Case 3 (Sina - 2018) in our study shows an algorithmic black box phenomenon at the accessibility level in its content operations. According to experience and original cognition in Case 3 (Sina - 2018), users believed that gender inequality persisted in society. Therefore, when the users found that the ratio of male to female users participating in the sweepstakes was 1:1.2 and the ratio of male to female winners was 1:112, they first made the empirical decision that:

> The Weibo sweepstakes platform conducted the sweepstakes based on random sampling because they could not directly see the sweepstakes platform's prize pool settings through the Sina Weibo client. Given the obvious gender ratio imbalance in the raffle result, users naturally associate this non-random sampling with gender discrimination (@jiangningpopo, 2018).

The reason users directly attribute the imbalance in gender ratio in the Weibo raffle results to gender discrimination is that they do not understand the permission-setting mechanism of the Weibo raffle. This lack of understanding reflects the inaccessible black-box nature of the algorithm. In most cases, the public cannot know the 'goals' and 'intentions' of the algorithm, nor do they have access to information about the algorithm's designers and controllers, the responsibility of the machine-generated content, or the ability to accurately judge and monitor it.

To address this problem and enhance the public's understanding of AI and algorithms, tech companies have developed the strategy of explaining these logics to the general public, for example, by publicizing model ideas for algorithmic decisions in the form of reports or by promoting and introducing software they have developed. However, the puzzling nature of algorithms is not limited to model design and program debugging for engineers; the improvement of algorithm accessibility also encounters obstacles in terms of interpretability. In our semi-structured in-depth interviews, Interviewee 08, who works as an algorithm engineer, spoke about the difficulties he and his colleagues confronted while communicating with the public:

> They often do not quite understand how the model has predicted the outcome, and they would like to know if they can determine the cause of the error if the model predicts it wrong. They may analyze the logic chain of human mistakes: if a person makes a mistake, then it is possible to work backward through every step of his behavior to determine exactly why he made a mistake. However, that is not the logic of algorithms (Interviewee 08, M. Henry).

Technical and disciplinary barriers are the two main obstacles in the process of algorithm black-box tuning. Although the rules underlying the logic of algorithms are relatively simple in terms of the current level of technological development and application, algorithm types are highly diverse. Algorithm program design is usually an overlay of multiple algorithm types, including, but not limited to, priority ranking, classification, correlation, and filtering algorithms (Zhang, 2018). This overlay becomes a technical barrier that must be faced to crack the black box.

4.3 Misguided evidence

The misguided evidence used by algorithms means that they inevitably make biased decisions (Mittelstadt et al., 2016). The misleading nature of algorithms is projected into algorithmic products as systematic and repetitive expressions of specific preferences, beliefs, or values in the algorithmic decision-making process, which is also known as algorithm bias (Friedman & Nissenbaum, 1996; Chander, 2017). Our research finds that there is a dual dimension of misguided evidence of algorithms. The first dimension lies in the data dimension, where the bias originally present in the dataset is amplified by the algorithm, manifesting the actual effect of algorithmic bias. The second dimension lies in the model, which is given to the algorithm by the engineer during design through parameter configuration and other means.

4.3.1 The data dimension of misguided evidence

The misleading nature of algorithms is largely caused by the quality of datasets used in machine learning. With the diversification of data types, content production and distribution have also adopted new requirements for data type identification. However, the diversification of data types is not accompanied by the improvement of data quality, which further deepens the influence of misleading algorithms on the reliability of algorithm output.

The automated decision-making of algorithms relies on the development of machine learning, an algorithmic procedure based on a dataset made up of an immense volume of data obtained from cell phones. Although the dataset is static, its characteristics can fundamentally affect the behavior of the model. When models are deployed in environments that do not match their training or evaluation datasets, or if these datasets reflect certain social biases, the output of the corresponding models is also likely to be biased (Gebru et al., 2021). Machine learning is a process of repeating the relationships between samples in a dataset. If the data used for learning is inherently biased, then these biases will be replicated or amplified an infinite number of times, eventually presenting themselves as algorithmic bias. Our semi-structured in-depth interviews with algorithm engineers suggest that the low-quality data used to train algorithms is a primary reason for the misleading results.

> The two main shortcomings that we currently experience most in the use of machine learning algorithms are that the algorithms are not very explanatory and the requirements for training data quality are high. Nowadays, AI algorithms are not perfect and often require a large amount of accurately labeled data. In actual production, such data are generally difficult to obtain (Interviewee 06, M, Franke).

Algorithmic programs can now operate independently, automatically distributing messages to users; however, it is precisely this independence that creates space for algorithmic bias to emerge. When algorithms are directly involved in information production and processing, the deep-seated biases that may be implicit in them are also exposed in practice and may algorithmically intensify biases already inherent to human cognition.

4.3.2 The model dimension of misguided evidence

The design and functionality of algorithms project the mental expectations and values of the algorithm designer. cambiar por Kevin Macnish (2012) pointed out that the values of the algorithm designer are necessarily frozen in the code. In other words, algorithm designers institutionalize their values through code. Algorithmic decision-making is based on the logic of statistics to build the model. The larger the sample size, the more accurate its overall description will be. However, in practical applications, a large amount of data does not guarantee the validity and reliability of the results. The strength of the effect of algorithmic bias may even increase with the amount of raw data (Kim et al., 2016).

Being frozen might be intentional on the part of algorithm designers. For platforms that use the 'manual recommendation + algorithm recommendation' content delivery model, the decision regarding what important content is, is delegated to the staff in the manual recommendation process. The staff will then deliver news to the users based on their own judgment regarding content importance. In the manual recommendation process, the decision as to important content is in the hands of the staff. In Case 1 (Byte Dance - 2017), Byte Dance's programmers manually set up the content distribution standard of the personalized recommendation function to create so-called eye-catching news and attract customers. This ultimately led to the increasingly vulgar and kitsch content received by users.

The ossification of the designer's values in the algorithm may also be unconscious. The algorithm is designed to passively reflect the value choices inherent in the social environment in which it operates. In Case 3 (Sina - 2018), Weibo's weight setting for platform users is related to the composition of users' social networks (number of Weibo follows, number of followers, number of accounts that follow each other) and users' online behaviors (including original posts, retweets, comments, likes, etc.). In this case, Sina's algorithm engineers set up criteria for identifying social bots based on their understanding of robot accounts. These criteria include sending a few original Weibo(s), retweeting without comment, and sending Weibo(s) with no or fewer pictures than normal users. User accounts with such characteristics were classified into the robot account category. Similarly, their user weights were uniformly reduced. In other words, it is the cognitive bias of the algorithm designer that leads to the biased results of algorithmic decision-making.

In addition to the biased results caused by the active behaviors of the algorithm designers, as shown in the above cases, algorithm bias may also be caused by the passive behaviors of the algorithm designers. In Case 2 (Meta - 2017–2021), Meta presented a lack of employees who understood special languages and cultural background. Given this skill gap, the company could not neither identify hate speech, extreme emotion, or terrorist content on Facebook in a timely manner, or adjust the algorithm model using keyword settings to monitor and screen out undesirable content on the platform in real-time. This led to the rapid and uncontrolled spread of hate speech, extreme emotions, and terrorist content.

The current datasets used for machine learning rely heavily on manual labeling efforts, and the labelers' cognitive biases are inadvertently projected into the labeling results (Diakopoulos, 2015). In subsequent machine learning and algorithmic decision-making sessions, these biases become embedded in the generated outputs and models (Romei & Ruggieri, 2014) and end up as algorithmic biases. A conscious process of being frozen is subjectively accorded to algorithms by people, which is easier to circumvent through technical means. To avoid the unconscious frozen phenomenon and data bias, higher-level technical finetuning and institutional settings, as well as improving internet literacy, are needed to achieve optimal results.

5. Algorithm ethical principles in intelligent communication

5.1 Principle of certainty

Technical uncertainty is one of the eight inherent uncertainty factors in the innovation process (Jalonen, 2012). The inconclusive evidence of algorithms is problematic because it redefines the connections between things using a logic of relevance that is not entirely certain. It, along with other modern technologies, dissolves traditional determinism. The traditional sense of certainty based on the accumulation of factual experience over time suffers when uncertain algorithms begin to act as a determining rule instead of human factual experience, which one can use to arrive at conclusions about the world.

Therefore, it is necessary to introduce the ethical principles of certainty to combat the multitude of problems caused by the inconclusive evidence of algorithms. In the information environment of intelligent communication, if we want to overcome the effects of inappropriate conclusions, such as echo chambers brought about by non-determinism, and return to a pluralistic, interactive ecological environment that enables benign information flow, we must first break through the information barriers erected by the inconclusive evidence of algorithms.

5.2 Principle of interpretability

Algorithms should have interpretability, but this feature is difficult to achieve because of the opaque evidence of algorithms. Transparency is often seen as the primary means of addressing algorithm opacity (Crawford, 2016; Neyland, 2016). At the same time, the impediment to oversight represented by algorithmic opacity (Burrell, 2016) intensifies the social outcry for greater algorithmic transparency.

The public shows considerable curiosity regarding algorithms, which places a demand on accessibility. In addition to technical means, the principle of interpretability also requires direct scientific explanation for users. Providing users with detailed explanations about the principles of AI techniques can enhance their understanding of how the entire algorithmic system works (Kulesza et al., 2012; Baron & Musolesi, 2020). In practice, the accessibility of algorithms can be achieved through a certain degree of information disclosure.

5.3 Principle of reliability

Reliability implies invariance and stability, as well as a return to order. Gebru et al. (2021) proposed to manage datasets by building datasheets for datasets used in machine learning. On one hand, this approach helps dataset creators eliminate potential risks or hazards during the process; on the other, it can ensure that algorithm engineers select and obtain the information they need more efficiently.

6. Conclusion

The development and application of machine learning (especially deep learning) and other algorithms in news content production and dissemination have initiated a new era of intelligent communication. Despite the variety of algorithms and applications, their basic technical logic is straightforward, and they all achieve similar retrieval through multidimensional fitting. In this study, inconclusive, inscrutable, and misguided evidence of algorithms were found to determine the existence of uncertain, opaque, and biased tendencies in algorithms operating in intelligent communication. Our paper focused on news content production and dissemination, extending the applicability of the algorithm ethics conceptual map (Mittelstadt et al., 2016) from Western academic discourse to practical applications in China, and systematically testing the scientific quality and rationality of the epistemic and ethical concerns of the conceptual map.

The inconclusive evidence of algorithms determines the uncertainty of algorithms in nature. The core of algorithms in intelligent communication is machine learning, including supervised learning (SL), semi-supervised learning (SSL), and unsupervised learning (UL) methods. SL refers to the discovery, construction, and evaluation of mathematical models by algorithms in the context of pre-determined human classification of a finite dataset (Nasteski, 2017). From SL to SSL to UL, while algorithms get 'smarter,' their uncertainty also increases. In the presence of a powerful AI that can learn and set goals on its own, the information that humans edit, disseminate, and store through technology is at risk of uncertainty. The 'true designers' of AI systems are always multiple and in constant flux, and the multiplicity of agents contributes to diverse unforeseeable outcomes. This fluidity and uncertainty are the main basis for exonerating algorithm designers when ethical problems arise.

There is also a process whereby the algorithm increases its inscrutability and decreases its opacity. The inscrutable evidence of algorithms and the contrived black boxes lead to uncertainty in the content and procedures of algorithmic practices. The embedding of AI in multiple aspects of journalistic practice makes algorithm and human roles overlap and become indistinguishable from each other. Meanwhile, factors such as high-dimensional data, complex codes, and variable decision logics lead to a lack of interpretability of automated algorithmic decisions. In recent years, the crisis of trust in journalism, the decline of journalistic objectivity, and the changing relationship between journalism and its audience have made transparency an increasingly important issue in journalistic discourse and, ultimately, one of the key elements of ethical journalism norms. As such, the trend toward increasing opacity stemming from algorithmic developments clearly counteracts the public's calls for greater transparency in terms of journalistic practice.

The ethical demand on technology is an inherent requirement for its development (Guo & Wen, 2011). Human behavior guided by instrumental rationality is purposively rational and emphasizes maximizing outcome benefits (Chen & Shi, 2017). However, an instrumental maximization of benefits is not directly related to a maximization of just outcomes. Algorithm bias refers to the uncertainty of the results of algorithm production practice. Algorithm biases stem from misguided evidence. In the machine learning process, these biases are not only reflected in the bias of the original dataset but also in the design biases of the algorithm model, which can be both conscious and unconscious. Since it is difficult for humans to be aware of their presence, integrating them into the algorithm in the form of a technical detection model is challenging. Under the guidance of instrumental rationality, allowing the algorithmic model and its derived technology to continue to influence human society through interpellated biases runs counter to the original goal of promoting the development of human society when algorithms were first developed.

The discussion in this study was limited to the cases selected for analysis. It should be mentioned that the research on algorithmic risk falls within the domain of positivism, while research on algorithmic ethics falls under critical research. There are fundamental differences in the research paradigms between the two. Using multi-case studies and semi-structured in-depth interviews, we attempted to link the study of algorithmic risk and the discussion of algorithmic ethics in the context of intelligent communication. When employing different research paradigms, disputes inevitably arise over the choice of method. However, interdisciplinary research depends on the meeting of diverse theories, methods, and modes of thought. In conducting this research, we sought to make the connection between intelligent communication and the epistemic concerns of algorithm ethics explicit and we hope that our study will influence future related research.

Due to case study limitations, this paper was unable to provide a deeper theoretical consideration of the ethical principles presented. Future research that could trace their cultural and historical roots would be valuable. Additionally, the ethical principles of algorithms proposed in this paper are only theoretical ideas. Further research and reflection are needed to make them acceptable to internet companies.

7. Authors' contribution statement

Jialin Lin: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft. Changfeng Chen: Validation, Writing – Review & Editing, Supervision, Project Administration.

8. References

Acemoglu, Daron and Restrepo, Pascual (2020). The wrong kind of AI? Artificial intelligence and the future of labor demand. *Cambridge Journal of Regions, Economy, and Society, 13*(1), 25–35. https://doi.org/10.1093/cjres/rsz022

Ananny, Mike (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, *41*(1), 93–117. https://doi.org/10.1177/0162243915606523

Athey, Susan (2017). Beyond prediction: Using big data for policy problems. *Science, 355*(6324), 483–485. https://doi.org/10.1126/science.aal4321

Baldini, Gianmarco, Botterman, Maarten, Neisse, Ricardo and Tallacchini, Mariachiara (2018). Ethical design in the Internet of Things. *Science and Engineering Ethics*, *24*(3), 905–925. https://doi.org/10.1007/s11948-016-9754-5

Barocas, Solon (2014, August). Data mining and the discourse on discrimination. In *Data Ethics Workshop, Conference on Knowledge Discovery, and Data Mining* (pp. 1–4). http://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf

Baron, Benjamin and Musolesi, Mirco (2020). Interpretable machine learning for privacy-preserving pervasive systems. IEEE *Pervasive Computing*, *19*(1), 73–82. https://doi.org/10.1109/MPRV.2019.2918540

Burrell, Jenna (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Calders, Toon, Kamiran, Faisal and Pechenizkiy, Mykola (2009, December). Building classifiers with independency constraints. In *2009 IEEE International Conference on Data Mining Workshops* (pp. 13–18). IEEE. https://doi.org/10.1109/ICDMW.2009.83

Chander, Anupam (2016). The racist algorithm. *Michigan Law Review*, *115*, 1023. https://heinonline.org/HOL/Page?handle=hein.journals/mlr115&div=42&g_sent=1&casa_token=&collection=journals

Chen, Changfeng and Shi, Ze (2017). Jishu yu jiazhi de lixing jiaowang: Rengong zhineng shidai xinxi chuanbo. Suanfa tuijian zhong gongju lixing yu jiazhi lixing de sikao [Rational communication between technology and value: Information dissemination in the age of artificial intelligence. Reflections on tool rationality and value rationality in algorithm recommendation]. *Xinwen Zhanxian*, 17, 71-74. https://kns.cnki.net/kcms2/article/abstract?v=Y7KVjSR9dMIWG_PzDMCOYOQeHdMolQagIJJ2XTYFYHN1S5eT16jIz_h0UNn0KVnyuFQRBtuUaQFy59ZKQ4r_Q3GUtpNOXbxOBUaDt4N-u1K43Te-sROTGPEeuis2usH7&uniplatform=NZKPT&language=CHS

Crawford, Kate (2016). Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology, & Human Values*, *41*(1), 77–92. https://doi.org/10.1177/0162243915589635

Del Vicario, Michela, Bessi, Alessandro, Zollo, Fabiana, Petroni, Fabio, Scala, Antonio, Caldarelli, Guido, Stanley, Eugene and Quattrociocchi, Walter (2016). The spreading of misinformation online. In *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113

Diakopoulos, Nicholas (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Eisenhardt, Kathleen and Graebner, Melissa (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Friedman, Batya and Nissenbaum, Helen (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Gebru, Timnit, Morgenstern, Jamie, Vecchione, Briana, Vaughan, Jennifer-Wortman, Wallach, Hanna, Iii, Hal-Daumé and Crawford, Kate (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Guo, Zhiye and Wen, Chengwei (2011). Jishu De Sange Neizai Lunli Weidu [Three intrinsic ethical dimensions in technology]. *Ziran Bianzhengfa Yanjiu*, *05*, 41–46. http://doi:10.19484/j.cnki.1000-8934.2011.05.008

Hajian, Sara, Bonchi, Francesco and Castillo, Carlos (2016, August). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2125–2126). https://doi.org/10.1145/2939672.2945386

Jalonen, Harri (2012). The uncertainty of innovation: A systematic review of the literature. *Journal of Management Research*, *4*(1), 1–47. http://dx.doi.org/10.5296/jmr.v4i1.1039

Kim, Yoonsang, Jidong, Huang and Sherry, Emery (2016). Garbage in, garbage out: Data collection, quality assessment, and reporting standards for social media data use in health research, infodemiology, and digital disease detection. *Journal of Medical Internet Research*, *18*(2), e4738. https://doi.org/10.2196/jmir.4738

Kulesza, Todd, Stumpf, Simone, Burnett, Margaret and Kwan, Irwin (2012, May). Tell me more? The effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 1–10). https://doi.org/10.1145/2207676.2207678

Macnish, Kevin (2012). Unblinking eyes: The ethics of automating surveillance. *Ethics and Information Technology*, *14*(2), 151–167. https://doi.org/10.1007/s10676-012-9291-0

Malkamäki, Arttu, Wagner, Paul, Brockhaus, Maria, Toppinen, Anne and Ylä-Anttila, Tuomas (2021). On the acoustics of policy learning: Can co-participation in policy forums break up Echo chambers? *Policy Studies Journal*, *49*(2), 431–456. https://doi.org/10.1111/psj.12378

Meng, Wei (2018). Xin chuanbo changjing zhong de lunli zhuiwen [Ethical inquiry in the new communication scene]. *Wangluo Chuanbo*, 4, 46–47. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7iXrgB0HQ3hJ0tlenCS92HVlBTHtAH1SRbSi6tRCh-jsffE8ARC2Q_Ji3NhhGtxYKo&uniplatform=NZKPT

Mittelstadt, Brent-Daniel, Allo, Patric, Taddeo, Mariarosaria, Wachter, Sandra and Floridi, Luciano (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Mittelstadt, Brent-Daniel, Russell, Chris and Wachter, Sandra (2019, January). Explaining explanations in AI. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 279–288). https://doi.org/10.1145/3287560.3287574

Miyazaki, Shintaro (2012). Algorhythmics: Understanding micro-temporality in computational cultures. *Computational Culture*, *2*, 1–16. http://computationalculture.net/algorhythmics-understanding-micro-temporality-in-computational-cultures/

Mullainathan, Sendhil and Spiess, Jann (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Mulvenna, Maurice, Boger, Jennifer and Bond, Raymond (2017, September). Ethical by design: A manifesto. In *Proceedings of the European Conference on Cognitive Ergonomics 2017* (pp. 51–54). https://doi.org/10.1145/3121283.3121300

Nasteski, Vladimir (2017). An overview of the supervised machine learning methods. *Horizons*, *4*, 51–62. DOI:10.20544/HORIZONS.B.04.1.17.P05

Neyland, Daniel (2016). Bearing account-able witness to the ethical algorithmic system. *Science, Technology, & Human Values*, *41*(1), 50–76. https://doi.org/10.1177/0162243915598056

Obermeyer, Ziad and Ezekiel, Emanuel (2016). Predicting the future–big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, *375*(13), 1216. https://www.nejm.org/doi/10.1056/nejmp1606181

Plantin, Jean-Christophe and De Seta, Gabriele (2019). WeChat as infrastructure: The techno-nationalist shaping of Chinese digital platforms. *Chinese Journal of Communication*, *12*(3), 257–273. https://doi.org/10.1080/17544750.2019.1572633

Romei, Andrea and Ruggieri, Salvatore (2014). A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review*, *29*(5), 582–638. https://doi:10.1017/S0269888913000039

Schermer, Bart (2011). The limits of privacy in automated profiling and data mining. *Computer Law & Security Review*, *27*(1), 45–52. https://doi.org/10.1016/j.clsr.2010.11.009

Taylor, Linnet (2020). The price of certainty: How the politics of pandemic data demand an ethics of care. *Big Data & Society*, *7*(2), 1–7. https://doi.org/10.1177/20539517209425

Zhang, Linghan (2021). *Regulation of algorithm in the age of artificial intelligence.* Shanghai People's Publishing House Press.

Zhang, Shuling (2018). Pojie heixiang: zhimei shidai de suanfa quanli guizhi yu touming shixian jizhi [Breaking the black box: Algorithm power regulation and transparent implementation mechanism in the era of intelligent media]. *Zhongguo Chuban*, *7*, 49–53. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7i0-kJR0HYBJ80QN9L51zrPx4WneRALTmmUdmiPx7Ni2I8EtMb9w_2PHB93ooBJjcl&uniplatform=NZKPT