Spoken Term Detection of Zero-Resource Language Using Posteriorgram of Multiple Languages

doi: 10.4036/iis.2022.A.04

Interdisciplinary Information Sciences Vol. 28, No. 1 (2022) 35–43

ISSN 1340-9050 print/1347-6157 online
DOI 10.4036/iis.2022.A.01

Epistemic Injustice as a Philosophical Conception for Considering Fairness and Diversity in Human-centered AI Principles

Mariko NIHEI[1,2,*]

[1]*Faculty of Economics, Matsuyama University, Matsuyama, Ehime 790-8579, Japan*
[2]*Advanced Institute for Yotta Informatics, Tohoku University, Sendai 980-9577, Japan*

The sheer quantity of information in the modern world has increased significantly, which exceeds the volume that can be managed using human power. Although information is necessary for decision making, excessive information is not beneficial for proper decision making. Therefore, data mining conducted using machine learning and artificial intelligence (AI)-assisted decision-making systems are increasingly being used in our society. However, problems, such as discriminatory decisions and the promulgation of injustice by AI, have been exposed recently. In response to this, numerous countries and organizations have recently announced a set of AI principles based on the concept of human-centered AI that fosters human values. The principles call for understanding diversity, ensuring fairness, and eliminating discrimination in the use of AI. To implement these values in AI systems, having a philosophical understanding of the structure of injustice in human knowledge production is essential. The problems of injustice and discrimination in knowledge production have recently been categorized as "epistemic injustice" in philosophy and epistemology, and the theories explaining these phenomena are becoming more sophisticated. This paper aims to contribute to the understanding of "human-centered" AI by connecting the philosophical concept of "epistemic injustice" to the discussion of AI ethical principles. It further points out that the issue of injustice and unfairness in AI use is not only a social–ethical as well as an epistemic concern.

KEYWORDS: epistemic injustice, diversity, human values, human-centered AI, ethical principles of AI

1. Introduction

The volume of digital data generated and reproduced has been increasing in this century with the improvement in computer performance since the 1980s and 1990s, the development of the Internet, and the increase in the number of Internet of Things (IoT) sensors since 2000 [1, 2]. Furthermore, the Covid-19 pandemic since 2020 has been a major factor in the rapid acceleration of the digitalization of our work and daily lives. Therefore, the large amount and variety of digital data generated daily in our lives, that is, Big Data, is expected to become a new and powerful source for human decision making and knowledge production.

The arrival of the big data era seems to be good news for human knowledge production activities because information is essential for humans to make decisions and produce new knowledge. However, a large amount of data may not necessarily improve the accuracy of our decision making. Humans can process a limited amount of information. Human confusion regarding the management of excessive information has already appeared in the first-century literature [3]. Furthermore, the existence of a confirmation bias is also widely known, such as human judgments tend to be qualitative and philistine and that we tend to make judgments and evaluations based on our beliefs.

To process large amounts of data, which is far beyond what humans can manage, for decision making and knowledge production, many organizations and companies are now adopting systems, such as data analysis using machine learning and artificial intelligence (AI)-assisted decision making. In addition to the advantages of processing large amounts of data, AI decision making seems to have the advantage of being free from the qualitative nature of individual human judgments and the influence of the confirmation bias.

However, in recent years, several reports have raised questions regarding AI judgment and decision making, where AI has made discriminatory decisions. For example, in 2017, Amazon's plan to introduce an AI-assisted hiring evaluation system was scrapped because the system evaluated female candidates less favorably than male candidates [4]. Furthermore, some research institutions have shown that the accuracy of the identification of facial recognition systems using pattern recognition, which many major companies and organizations in North America have already implemented, fluctuates greatly depending on race and gender.

Accordingly, the accuracy of facial recognition is lower for women than for men and lower for black and Asian

Received October 25, 2021; Accepted January 19, 2022
This work was supported by JSPS KAKENHI Grant Number 19K13047 and Yotta Informatics Project by MEXT, Japan.
The author would like to thank the two anonymous reviewers for their comments, which helped to clarify the points made in this paper.
The author would also like to thank Enago (www.enago.jp) for the English language review.
*Corresponding author. E-mail: mnihei@g.matsuyama-u.ac.jp

people than for white people [5, 6]. As it became apparent that AI could create discrimination and bias, a surge in the publication of ethical principles for AI was observed from around 2017 to 2020. Many of these principles call for the cultivation of human values in the use of AI and human intervention in AI decision-making systems. Concepts, such as "human-centered" and the "Human-in-the-loop," have become trends in AI ethics over recent years.

Responsible human intervention is necessary for the realization of trustworthy AI. However, why does AI, which is supposed to be free from the bias of human qualitative judgment, generate discriminatory decisions? The latent structure of prejudice and deviation in human knowledge activities has been introduced into AI. Therefore, to realize human-centered and trustworthy AI, it is necessary first to understand the type and process of distortions, injustices, and inequities that occur in human knowledge activities. However, there is a lack of discussions promoting the understanding of these factors in the AI ethical principles. The topic of injustice in human knowledge activities or epistemic domains, that is, the epistemology of "epistemic injustice," has become a trend in both philosophy and epistemology in recent years. This paper aims to make a small contribution to the realization of "human-centered" AI by connecting the philosophical concept of "epistemic injustice" to the discussion of AI ethical principles.

We first discuss instances of "discriminatory" decisions by AI that have arisen from 2017 in Sect. 2 and confirm that the discrimination therein is same root as the discrimination and distortions that arise in our activities in the real world. In Sect. 3, I will discuss the recent surge in the proposal of ethical principles regarding AI. There is a common trend, that is, the emphasis on the keywords "human-centered" and "human values." However, concrete measures to cultivate human-centered values, such as diversity or fairness, have not yet been sufficiently presented. In Sect. 4, I review the definition and occurrence of the concept of epistemic injustice and then outline the structure of several types of injustice. I also point out that further injustices could arise if epistemic injustices were reconstructed in the digital data world. Finally, I will also point out that the discrimination and bias by AI is not only an ethical problem but also an epistemic problem that directly affects our knowledge activities.

2. Discrimination by AI?

In October 2018, Reuters published an article titled "Amazon scraps secret AI recruiting tool that showed bias against women" [4]. According to the report, Amazon had been developing an AI system to automate hiring assessments since 2014. However, the prototype system had a problem, that is, the AI gave lower ratings to female candidates, thereby making decisions that discriminated against women. The AI made such "discriminatory" decisions because of the learning training material given to it. Developers gave the AI a large number of resumes sent to Amazon over the past 10 years as training material, most of which were from men. Therefore, the AI seems to have learned that it should hire men rather than women and formed an algorithm that gives lower ratings to female candidates or candidates associated with the keyword "female." Amazon gave up on the recruitment AI because of this problem and terminated the development project in early 2017.

The "discrimination" against women by AI is because AI has "learned" from the gender disparity and gender discrimination in the real world. That is, it is ultimately observed because of the discrimination and prejudice existing in human society. There is a report that an AI-based grade prediction system introduced by the UK's Office of Qualifications and Examinations Regulation (Ofqual) was found to give unfairly low grades to working-class and racial minority examinees [7]. This case may be attributed to AI learning and extracting from the instances of socioeconomic disparities, racial discrimination, and prejudice in the real world.

3. Human-centered AI Ethical Principles

Against the backdrop of the AI "discrimination" scandal mentioned above, there has been an increase in the publication of new ethical principles and guidelines for AI in the past few years. According to the 2021 AI Index Report by Stanford Institute for Human-Centered Artificial Intelligence, a total of 117 ethical principles related to AI were published by countries, organizations, and companies worldwide during 2015–2020 [8]. In particular, an increase in the number of ethical principles published during 2018–2019 was observed. In 2018, 45 ethical principles were published, including those from large companies, such as IBM, Google, and Facebook. Furthermore, 28 ethical principles were published in 2019, including those from the EU, the Institute of Electrical and Electronics Engineers (IEEE), and others discussed below.

The common thread among these recently released AI ethics principles is the emphasis on "human-centeredness" and human control and responsibility. In addition to the security, traceability, and privacy issues that have been present in data quality standards for some time, human control, diversity, and well-being are now clear to see in the principles and guidelines. Figure 1 summarizes the representative AI ethical principles published during 2017–2019.

The Future of Life Institute's (FLI) Asilomar AI Principles (2017) declare that the AI targeted for development and research should be beneficial intelligence and not undirected intelligence; and the principles include 13 principles related to values and ethics [9]. These principles call for AI to be in harmony with human values, such as human dignity, freedom, and diversity, and for humans to take responsibility and control to realize such an AI. The FLI principles have become the forerunner of the recent AI ethics principles, with human-centered AI and human-

Title	Asilomar AI Principles (2017) [9].	IEEE Ethically Aligned Design, first edition (2019) [10].	Ethics Guidelines for Trustworthy AI (2019) [11].	Principles of Human-centric AI society (2019) [12].
by	The Future of Life Institute	The IEEE Global Initiative /USA	High-Level Expert Group on Artificial Intelligence /EU	Council for Social Principles of Human-centric AI /Japan
Components	5 research issues + 13 Ethics and Values + 5 Longer-term Issues	8 General principles	4 General ethical principles + 7 requirements	3 General philosophical principles for society + 7 social principles of AI
Principles	Ethics and Values 1. Safety 2. Failure Transparency 3. Judicial Transparency **4. Responsibility** **5. Value Alignment** 6. **Human Values** 7. Personal Privacy 8. Liberty and Privacy 9. Shared Benefit 10. Shared Prosperity **11. Human Control** 12. Non-subversion 13. AI Arms Race	**1. Human Rights** **2. Well-being** 3. Data Agency 4. Effectiveness 5. Transparency 6. Accountability 7. Awareness of Misuse 8. Competence	• **Respect for Human Autonomy** • Prevention of Harm • **Fairness** • Explicability 1. Human Agency and Oversight 2. Technical Robustness and Safety 3. Privacy and Data Governance 4. Transparency **5. Diversity, Non-discrimination and Fairness** 6. Environmental and Societal Well-being 7. Accountability.	• Dignity: A society that has respect for **human dignity** • **Diversity & Inclusion**: A society where people with diverse backgrounds can pursue their well-being • Sustainability: A sustainable society **1. Human-Centric** 2. Education/Literacy 3. Privacy Protection 4. Ensuring Security 5. Fair Competition 6. **Fairness,** Accountability, and Transparency 7. Innovation

Fig. 1. Ethical AI principles (**Emphasis (bold)** is by the author.).

participatory AI development as core ideas. The following are some principles related to human values in the FLI principles.

- Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications. (FLI-4)
- Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation. (FLI-5)
- Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity. (FLI-6)

In 2019, the IEEE first published *Ethically Aligned Design* comprising eight general principles for "the ethical and values-based design, development, and implementation of autonomous and intelligent systems" [10]. The following are some principles related to human values in the IEEE principles.

- Human Rights: AI systems shall be created and operated to respect, promote, and protect internationally recognized human rights. (IEEE-1)
- Well-being: AI systems' creators shall adopt increased human well-being as a primary success criterion for development. (IEEE-2)

The EU Ethics Guidelines for Trustworthy AI (2019) comprises four general ethical principles and seven requirements to achieve them [11]. In particular, the principles of fostering diversity, eliminating discrimination in AI, and employing fairness in the use of AI are more specific than the other principles.

- Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle. (EU-5)

In Japan, the *Social Principles of Human-centric AI* was published by the Council for Social Principles of Human-Centric AI in 2019 [12]. It comprises three basic principles and seven social principles of AI. The basic principles include "diversity and inclusion: A society where people with diverse backgrounds can pursue their well-being." The social principles of AI include "human-centeredness," which calls for AI to harmonize with the human rights and well-being of diverse people, and "fairness, accountability, and transparency," which calls for the elimination of discrimination by AI.

- The Human-Centric Principle: The utilization of AI must not infringe upon the fundamental human rights guaranteed by the Constitution and international standards. AI should be developed, utilized, and implemented in society to expand the abilities of people and allow diverse people to pursue their own well-being. In a society using AI, we should introduce appropriate mechanisms for literacy education and the promotion of the proper use of AI to ensure that people do not become over-dependent on AI or misuse AI to manipulate other people's decision making. (Japan-1)
- The Principle of Fairness, Accountability, and Transparency: In an "AI-Ready Society," it is necessary to ensure fairness and transparency in decision making, establish appropriate accountability for the results, and trust in the technology, so that people using AI are not subject to undue discrimination with regard to personal background or unfair treatment in terms of human dignity.
  Under AI's design concept, all people are treated fairly without unjustified discrimination on the grounds of backgrounds, such as race, gender, nationality, age, political beliefs, religion. (Japan-6)

Many other principles claim to be consistent with human values. Moreover, "fostering diversity" and "eliminating discrimination and injustice" are considered specific issues intended to address the substance of human values. In the background, of course, there is a reflection on many cases of "discrimination by AI" reported in this period, including the case of Amazon's human resource evaluation system.

Although the low ratings for women were derived from Amazon's AI system, it may be inaccurate to describe the cases as discrimination "by AI." We should understand it as an example of AI reproducing discriminatory decisions by learning from existing discrimination and disparity in human society. If there is any bias in the given existing data, it will be "learned" by AI without being corrected. This aspect has been pointed out in a research report on big data published in 2016. Thus, "approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment." [13]. In their co-authored book, Daugherty and Wilson describe this situation as "Bias In, Bias Out," in analogy to "Garbage in, garbage out."

Avoiding "Bias In, Bias Out" requires appropriate human intervention, including human review of the data used to train the AI for preventing the introduction of existing biases into the AI and human evaluation of the AI's output to ensure that it does not reflect discrimination or injustice. In recent years, human involvement in the development and design of AI, including collecting input data, training, evaluation and verification of output, and feedback, has become increasingly crucial. Ethical sensitivity around humans interacting with AI as trainers and developers is essential to avoid AI making discriminatory judgments and to prevent the expansion and reproduction of biased judgments by AI in society.

4. Epistemic Injustice and Human Values of Diversity and Fairness

The discussion so far has confirmed that to realize human-centered AI, humans must be aware of the prejudices, injustices, and disparities occurring in reality and be sensitive to the diversity in society. However, the prejudice, injustice, and discrimination that we engage in and form are both conscious and latent as well as individual and structural. In particular, there has been a lack of concepts that adequately recognize and depict discrimination and injustice related to knowledge activities for a long time. However, in recent years, a new area of epistemological debate has begun to emerge, namely the concept of "epistemic injustice." Unfortunately, there have not been many discussions directly associating the concept of epistemic injustice to AI ethics or development. Although epistemic injustice, in its origins, is a discussion of injustice that occurs in our human epistemic activities, it can appear in the epistemic environment where AI and humans intersect. I would also argue that the cases of discriminatory judgments brought about by AI are a new form of epistemic injustice.

Understanding the conceptual tool of epistemic injustice, which linguistically visualizes the biases, deviations, and injustices that can occur in the realm of information and knowledge, would be useful for humans to evaluate the

fairness of the information given to AI and the decisions it subsequently makes. Alternatively, it has also been argued that the phrases of principles do not explain the significance or necessity of the phrases and thus are not effective. However, the following discussion should clarify the significance of the call for diversity and fairness, or rather, why it is necessary. These demands are not only social and ethical but also epistemic.

4.1 Beginning of the epistemic injustice concept and its various types

The epistemology of "epistemic injustice" is a philosophical and epistemological area that considers "how epistemic practices and institutions may be deployed and structured in ways that are simultaneously infelicitous toward certain epistemic values (e.g., truth, aptness, and understanding) and unjust with regard to particular knowers" [15]. The term "epistemic injustice" is a relatively new concept in the history of philosophy and epistemology because it was first proposed in M. Fricker's 2007 book. However, this term explains a phenomenon experienced by us all whether we are aware of it or not.

According to Fricker's definition, epistemic injustice is "wrong done (to) someone specifically in their capacity as a knower" (Fricker 2007, 1) [16]. Fricker herself describes the typical patterns of epistemic injustice as "testimonial injustice" and "hermeneutical injustice." Testimonial injustice is when "prejudice causes a hearer to give a deflated level of credibility to a speaker's word" (Fricker 2007, 1) [16]. For example, she cites a white police officer who refused to listen to a black speaker. Another example is when a woman's testimony is evaluated as less reliable and persuasive than a man's testimony. In this case, the hearer's unfair devaluation of the speaker may be done consciously or unconsciously. Prejudices that already exist and permeate society, for example, biases against women, affect the evaluation of the ability of a person as a woman and as a speaker/epistemic agent. However, because being a woman is essentially irrelevant to one's evaluation as a speaker/epistemic agent, then such a low evaluation constitutes "injustice."

Hermeneutical injustice occurs when "a gap on collective interpretative resources puts someone at an unfair disadvantage when it comes to making sense of their social experience" (Fricker 2007, 1) [16]. When there are no "collective hermeneutical resources," which are typically linguistic and institutional conceptual resources to speak and understand the experience or situation of something, the people experiencing those experiences or situations are prevented or deprived of the opportunity to speak and understand their experiences. Fricker cites the concept of "sexual harassment" as an example: in the absence of this concept, the harassed person cannot recognize and talk about the experience as harassment. At the same time, the harassers cannot recognize their experience as harassment. Moreover, when the harassed person tries to talk about their experience, the hearers cannot understand the experience and the disadvantageous position the harassed individual is in.

Because justice and injustice have always been philosophical themes, there has been a certain amount of discussion about them. However, they are also "epistemic" injustice or evil, which is a new point that Fricker has clarified. Epistemic injustice is epistemically harmful because it is "an epistemic disadvantage to the individual hearer and a moment of dysfunction in the overall epistemic practice or system" (Fricker 2007, 43), and it "suppresses and insults people as epistemic subjects because of prejudicial stereotypes" (Fricker 2007, 44). Not being treated legitimately as an epistemic subject is "to be wronged in a capacity essential to human value" (Fricker 2007, 44) [16].

Moreover, suppose these injustices continue to arise, they will affect not only the day-to-day practices among us individually but also the epistemic institutions at the social and collective levels. For example, there is a danger that school curricula will become systematically biased or that academic disciplines will be structured in ways that ignore certain intellectual traditions. Therefore, our knowledge systems and histories will be fixed and reproduced as biased and unjust. As I have already pointed out, although the use of AI technology is becoming increasingly integral to our epistemic practices and institutions, it can also learn from, expand upon, and further entrench our biases if used carelessly. The discrimination against women in Amazon's AI recruitment system is a form of epistemic injustice.

Following Fricker's groundbreaking conceptualization, several philosophers have pointed out the various types of injustice involved in epistemic activities. For example, Hookway points out an example that is often observed within a community for intellectual practice, such as a scientific community, where a person or group is not treated as a legitimate participant in the intellectual practice [17]. According to Hookway, for a community member to be considered a legitimate epistemic subject and as a co-participant in the practice of knowledge inquiry, the individual being only heard is not sufficient. It is also necessary that one's claims and questions are taken seriously and that responses, such as explanation and criticism, are maintained. When such reciprocal response relationships are unjustly prevented or deprived, epistemic injustice occurs. This type of epistemic injustice was later termed "participatory epistemic injustice" by Grasswick [18].

Alternatively, Coedy proposed the concept of "distributive epistemic injustice," which focuses on the distribution of epistemic goods in society [19]. This occurs when epistemic resources and epistemic goods are distributed in ways that unfairly restrict access to and opportunities for education and information or violate the right of epistemic agents to know.

4.2 Epistemic injustice and its dangers in information resources

With the establishment of the concept of "epistemic injustice" as described previously, it became possible to

consciously recognize phenomena and situations that had already appeared in our intellectual lives and the domain of information resources as epistemically unjust. (This means that before the advent of the concept of "epistemic injustice," a state of "hermeneutical injustice" arose in response to various epistemic injustices that had already occurred). Epistemic injustice has become one of the hot issues in philosophy of science and the theory of knowledge as well as in various feminist theories that have dealt with gender inequality and gender prejudice. Access to this concept has allowed us to conceptualize the bias in our understanding of the world and to understand the significance of diversity in knowledge inquiry. Recently, library and information science, which has long been concerned with the management and preservation of information resources, has also begun to propose pioneering arguments that introduce the concept of "epistemic injustice." By referring to the results of these studies, in this paper, I will extract the epistemic injustice that appears in information resources and their use, which is deeply related to data mining using AI. I will also reconfirm the critical impact of such epistemic injustice on our knowledge system and the history of knowledge.

Patin and her collages, library and information scientists, formulate the dangers that the epistemic injustices that emerge in the information resources and our knowledge practices can pose as three levels of harm to our knowledge [20]. The first harm is the damage experienced by the disadvantaged individual by prejudice and the absence of hermeneutical resources. This is at the individual level when the experience of not being recognized as an epistemic subject accumulates, and the person loses confidence in themself as an epistemic subject.

Secondary harms are suffered by the society and community contemporaneous with those individuals, including those suffering from epistemic injustice. Such societies or communities are not willing to hear the narratives of disadvantaged people. Furthermore, they lack the hermeneutical resources that express the experiences of disadvantaged people. Therefore, they are unable to learn from the experiences, which is epistemic damage in the sense that they miss out on potential knowledge.

The third harm associates with the future, which Patin *et al.* find most problematic. "The third harm considers missing iterations of knowledge transfer caused by generations of epistemic injustice" [20]. The first and second harms mean the loss of potential knowledge in their time; however, this loss of knowledge will be passed on from generation to generation, thereby resulting in the loss of a tremendous amount of potential knowledge in the future. In a future where tertiary damage is occurring, "learners cannot build on work that has not been legitimized/archived" [20]. Suppose experiences that should have been collected and preserved are not archived and lost because of the first and second harms. In that case, in the future, we will be deprived of the opportunity to learn from those past experiences and lose potential knowledge. Although "the third harm, and the influence it has intergenerationally, will always resist datafication or quantification," [20] it is easy to imagine the severity of this harm. The third harm will also produce future distributive epistemic injustices, such as the entrenchment of unfair educational systems and curricula and the unfair distribution of information opportunities desired by epistemic agents.

The discussion by Patin *et al.* is not directly tied to the use of AI. However, at all the levels pointed out above, AI technology is already being used, and thus it can be complicit in any level of danger. Moreover, there concerns that the spread of AI technology could accelerate the scale of damage. In the following, relying on the three-level framework of Patin *et al.* and pointing out the connection with the use of AI technology, I will examine examples of epistemic injustice at each level.

In the field of feminism, the phenomenon of doctors disregarding or refusing to understand the symptoms of female patients has long been established as a part of clinical practice. This is one of the manifestations of testimonial injustice because of prejudice which holds that women are overly emotional or lack rational explanatory power. It may also be a manifestation of hermeneutical injustice due to the lack of hermeneutical resources, which makes it difficult for women patients to recognize and communicate their own experiences of discomfort and pain and for the doctors who listen to them to understand their experiences. (Fricker herself cites an example of hermeneutical injustice due to the absence of the concept of "postpartum depression." [16]) Epistemic injustice between doctors and patients can lead to biased healthcare data for certain patient categories (e.g., women) and affect the data used for training in AI-based medical diagnostic support systems. This epistemic injustice creates harm at the first individual level, i.e., the negative impact on the health status of individual patients whose symptoms are not understood, as well as the second harm in contemporaneous society, i.e., society's lack of understanding and potential knowledge of the health status of certain categories of people. Furthermore, suppose the absence or bias of data on specific categories of patients continues to accumulate without the existence of this type of injustice being recognized. In that case, the third level of harm may occur: the long-term fixation of bias and ignorance toward specific categories of people through the fixation of biased information in the educational and diagnostic systems. As a result of continuing to use biased data influenced by epistemic injustice as learning materials, AI diagnostic systems may amplify and immobilize or authorize incomprehension and ignorance toward, for example, female patients.

From the relationship between AI and humans, concerns have been expressed about new testimonial injustices caused by the authority of data mining results and AI judgments over individual human testimony. Fricker's testimonial

injustice was considered something that occurs between the human hearer and the human speaker. However, Origgi and Ciranna point out that something similar can occur in the relationship between the claims of humans as epistemic agents and the decisions made by AI [21]. For example, which is considered more reliable, the prediction of a person's behavior derived by AI from Google's accumulated location data or the testimony based on the person's own memory? Today, many of us may only give low ratings to human testimonies because of our "prejudice" that human memories are poor and subjective or that information from Google or AI is, by contrast, totally accurate and objective. Moreover, the low evaluation may be given by the person who has the memory. In this case, the structure of testimonial injustice at the individual level is established because humans themselves unfairly underestimate their own epistemic abilities as speakers. Or such a relationship between AI systems and humans can also be considered a new form of participatory injustice in the sense that humans are not recognized as participants in epistemic activities. Such participatory injustice can be interpreted as the second level of harm, in the sense that the community discounts and values human testimony over AI or machines decisions, thereby missing the opportunity to learn from individual humans.

In addition, this situation could cause a third, severe harm. If AI decision-making continues to be given more weight than human testimony and decision-making, humans may gradually lose confidence in themselves as epistemic agents. Those who are disadvantaged by testimonial injustice continue to live in the face of their hearers' incomprehension and indifference. Moreover, they also continue to have their epistemic abilities as speakers questioned and disregarded. Finally, they may even lose the motivation to talk and understand their experiences. Dotson defines such situations as the "testimonial smothering" [22], wherein the epistemic subject, tired of being continually placed at a disadvantage, elects to remain silent about their experience. As we have pointed out earlier, there is a danger in always prioritizing AI decisions over human ones, because the data that serve as learning materials for AI can contain our human biases and prejudices. Nevertheless, suppose humans continue to rely on AI without awareness or responsibility. In that case, this is also a violation of human values because it abandons or denies the ability and right to narrate one's experience, communicate it to others, and be understood. If the use of AI by humans leads to the denial of human values, this would be the exact opposite of "human-centered AI." This is a bad scenario; however, it is not difficult to imagine.

In light of the above discussion, testimonial injustice and testimonial smothering teach us an important fact about existing information resources: the absence of data, narratives, statistics, and documents about a particular social event or phenomenon is not equivalent to the actual absence of the event or phenomenon. The absence of data may result from the choice of the victims or parties involved to testimonial injustice or smothering. In other words, we should be aware that existing data from the past can contain the results of individual and communal epistemic injustices when the data was collected. If we neglect this point and use the past data in our current data mining, we will end up making predictive judgments about the future by preserving the epistemic injustice for certain people or pretending that certain social experiences and events do not. Data management that ignores the effects of epistemic injustice may become institutionalized in educational curricula and public archives, which may lead to the fixation and expansion of false histories and information. This is the unconscious and irresponsible creation of the third level of harm.

In recent years, digital humanities, i.e., methods that use AI and machine learning to analyze large data sets (text, statistics, images, etc.) from the past, have been gaining attention in the humanities, including history and literature. One of the major advantages of digital humanities is that it allows us to compare and examine a much larger amount of data than is possible by human hands. At the same time, however, the danger of uncritically treating existing data and archives as neutral research objects is beginning to be pointed out [23]. This is socially problematic in the sense that it ignores past injustices and is epistemically problematic for our knowledge systems in the sense that data mining based on injustice structures may be fixed on the academic side. While appreciating that digital methods will be a powerful tool for the humanities in the future, historical researcher Kokaze points out that digital methods are only possible on the basis of "what is written (書かれているもの)," that is, the data that exists. According to Kokaze, a characteristic of humanities research conducted by humans is to "decipher the unwritten (書かれていないものを読み解く)" [24]. Human beings can examine the meaning of the absence of any data in a given period, domain, or society. To reiterate this point in approaching this paper, human researchers can and should be conscious of the effects of human-created epistemic injustice. Without sensitivity to epistemic injustice, data mining using AI technology may lead to discriminatory decisions, and if this situation continues over many generations, a great deal of information that should have been there will have "never been there," and humans will lose a lot of knowledge.

As for the problems I pointed out above, i.e., the epistemic injustices that can be contained in existing data or archives, some efforts to correct them are beginning to be made, and I will introduce one example at the end. Against the backdrop of growing protests against racism, such as Black Lives Matter, the United States has recently begun to examine the injustice, prejudice, and dismissal of diversity in the collections and archives of domestic institutions. With this trend, the State of Alabama's Archives and History Department announced in June 2020 that archival materials in its custody contained racist bias and declared a renewed commitment to fostering diversity. According to the statement, "for well over a half-century, the agency committed extensive resources to the acquisition of Confederate records and artifacts while declining to acquire and preserve materials document, declining to acquire and preserve materials documenting the lives and contributions of African Americans in Alabama" [25]. Such efforts to validate existing archives and rebuild archiving to foster diversity are advancing in North America and other countries and regions

worldwide. Digital Humanities, which can handle large-scale data, can produce socially and epistemically meaningful results if it is promoted in conjunction with efforts toward epistemic justice, such as critical examination of existing archives.

So far, using several examples, I have examined the relationship between AI use and epistemic injustice in our knowledge circumstances. If humans remain unaware of the epistemic injustices and their potential harms in knowledge practices, we will allow AI to create discrimination and prejudice. Discrimination and prejudice created by AI are not only socially and ethically wrong but also wrong for the future of our knowledge. Therefore, cultivating diversity and ensuring fairness in the use of AI, as required by AI ethics, is not only a social–ethical requirement but also an epistemic one. Those involved in AI development should be keenly aware of this point.

4.3 Toward redressing epistemic injustice in AI use

How can epistemic injustice be corrected? Unfortunately, compared with the conceptualization of epistemic injustice and pointing out examples of epistemic injustice that occur in various domains, the discussion of the resolution and correction of epistemic injustice is overwhelmingly underdeveloped and the task of proposing solutions and justifying them is also beyond the scope of the discussion in this report. Of course, the fact that the concept of "epistemic injustice" has been clarified and made available is a major first step toward rectification. Before this, these injustices existed without being interpreted or described and were not even noticed by anyone other than those who were so disadvantaged. I believe that this report also played a role in this "first step" in that it introduced the concept of "epistemic injustice" to the discussion of AI use and AI ethics.

Fricker herself argues that the solution to epistemic injustice is for the individual to acquire the virtues of epistemic justice and sensitivity to diversity. However, as Anderson points out, the correction of injustice at the societal level, such as in education, information systems, and the development of public archives, requires institutional efforts and the cultivation of individual virtues [26]. It is also necessary to address epistemic injustice in the use of AI technology at the institutional level.

The following is a brief list of issues that have been proposed to remedy epistemic injustice.

- Create and support organizational diversity: Create professional groups and organizational teams that include people from diverse backgrounds. Patin *et al.* also point out that it is important to first associate with our colleagues from diverse backgrounds without underestimating their testimonial abilities [20].
- Create open platforms: Samarzijia and Cerovac emphasize the importance of creating politically, culturally, and academically open spaces, where social groups with different experiences and baggage can interact on equal terms [27]. By creating a space where people with epistemically disadvantaged experiences can speak without being epistemically disrespected, we can expect to enrich our conceptual resources.
- Verify and continually check the distribution of epistemic resources: As in the case of the Alabama Archives, existing archives and statistical materials can contain unfairness in favor of or overrepresented by certain groups or diluted or lost in favor of certain groups. These biases should be examined and corrected to avoid passing them on to future generations. Alternatively, syllabus audits should be conducted to check whether the educational curriculum disadvantages certain groups in terms of their information resources.

These points raised here are nascent institutional design ideas that allow us to have a conscious and critical perspective on the epistemic injustices that can arise in data production or be contained in existing data. These are not necessarily ideas specific to the use of AI; however, there are also lessons to be learned in the practice of AI technology and AI development. For example, greater diversity within an AI development team will increase the number of different values, perspectives, and experiences contained within the team. The ability to scrutinize the data and AI-derived results from multiple, diverse perspectives may allow for more critical bias detection. It is also possible, for example, to work with an open platform where people with the various experiences mentioned above can come together in an attempt to do a technology assessment during the AI development process. The development of AI through trial and error in the context of human beings with diverse experiences and perspectives will prevent the spread of epistemic injustice to certain people by AI. At the same time, it will contribute to the awareness and maintenance of human epistemic agency in the relationship between AI and humans, that is, the awareness and maintenance of human responsibility for the use of AI. When AI technology is becoming indispensable for knowledge practice, to prevent the bad scenarios described in the previous section from becoming a reality, humans must use AI technology consciously while maintaining a critical perspective on the data used by AI and the decisions made by AI.

5. Conclusion

Based on the discussion in this paper, the principles of diversity and fairness in human-centered AI ethics require the following additions: Humans must take a responsibility not to allow AI technology to be complicit in epistemic injustice. Of course, philosophical discussions, such as this paper, may be some distance from providing direct

guidelines for development and use at the technical level. However, by introducing the concept of epistemic injustice and understanding its danger to human intellectual life, this paper may contribute to understanding the significance of human values, such as diversity and fairness, as something more concrete and necessary. Principles are not just abstract phrases but issues that will directly affect the future of our knowledge life and intellectual environment.

The idea of human-centered AI is not yet mature. To understand and embody this developing idea in a more meaningful way, accumulating discussions on human-centered values from a broad, interdisciplinary perspective, including philosophical directions such as this paper, and continuously updating our understanding of human values are crucial. This paper, although modest, was an attempt to do so.

REFERENCES

[1] Muraoka, H., Gyoba, G., Suzuki, Y., Shioiri, S., Nakao, M., Nihei, M., *et al.*, "Gigantic amount information and storage technology: Challenge to Yotta-byte-scale informatics," *IEICE Technical Report*, **116(441)**: 27–32 (2017).

[2] Nihei, M., "Trends and issues about increasing the data volume and an approach for information quality," *MORALIA*, **23**: 18–33 (2016).

[3] Blair, A., Information overload, the early years, *The Boston Globe*, Retrieved from http://archive.boston.com/bostonglobe/ideas/articles/2010/11/28/information_overload_the_early_years/ (2010 November 28).

[4] Dastin, J., Amazon scraps secret AI recruiting tool that showed bias against women, *Reuters*, Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (2018 October 11).

[5] Boutin, C., NIST study evaluates effects of race, age, sex on face recognition software: Demographics study on face recognition algorithms could help improve NIST, Retrieved from https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software (2019 December 19).

[6] Hao, K., Making face recognition less biased doesn't make it less scary, *MIT Technology Review*, Retrieved from https://www.technologyreview.com/2019/01/29/137676/making-face-recognition-less-biased-doesnt-make-it-less-scary/ (2019 January 29).

[7] Hao, K., The UK exam debacle reminds us that algorithms can't fix broken systems, *MIT Technology Review*, Retrieved from https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/ (2020 August 20).

[8] Stanford University Human-Centered Artificial Intelligence, 2021 AI Index Report, Retrieved from https://aiindex.stanford.edu/report/.

[9] Asilomar AI Principles, Future of life institute, Retrieved from https://futureoflife.org/ai-principles/ (2017).

[10] IEEE Ethically Aligned Design, first edition, The IEEE Global Initiative, Retrieved from https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined (2019).

[11] Ethics Guidelines for Trustworthy AI, Hight-level expert group on artificial intelligence, Retrieved from https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines, Retrieved from, pdf (2019).

[12] Principles of Human-centric AI Society, Council for social principles of human-centric AI, Retrieved from https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/aigensoku.pdf (2019).

[13] Barocas, S., and Andrew, D. S., "Big data's disparate impact," *California Law Review*, **104**: 671–732 (2016).

[14] Daugherty, P. R., and Wilson, H. J., HUMAN+MACHINE: Reimagining Work in the Age of AI, Harvard Business Review Press (2018).

[15] Pohlhaus, G., "Varieties of Epistemic Injustice," in Kidd, J., Medina, J., and Pohlhaus, G. (eds.), *The Routledge Handbook of Epistemic Injustice*, Routledge, 13–26 (2017).

[16] Fricker, M., *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press (2007).

[17] Hookway, C., "Some varieties of epistemic injustice: Reflections of fricker," *Episteme*, **7(2)**: 151–163 (2010).

[18] Grasswick, H., "Epistemic Injustice in Science," in Kidd, J., Medina, J., and Pohlhaus, G. (eds.), *The Routledge Handbook of Epistemic Injustice*, Routledge, 313–323 (2017).

[19] Coady, D., "Two concepts of epistemic injustice," *Episteme*, **7(2)**: 101–113 (2010).

[20] Patin, B., Sebastian, M., Yeon, J., Bertolini, D., and Grimm, A., "Interrupting epistemicide: A practical framework for naming, a practical framework for naming, identifying, and ending epistemic injustice in the information professions," *Journal of the Association for Information Science and Technology*, **72**: 1306–1318 (2021).

[21] Origgi, G., and Ciranna, S., "Epistemic Injustice the Case of Digital Environments," in Kidd, J., Medina, J., and Pohlhaus, G. (eds.), *The Routledge Handbook of Epistemic Injustice*, Routledge, 303–312 (2017).

[22] Dotson, K., "Tracking epistemic violence, tracking the practice of silencing," *Hypatia*, **26(2)**: 236–257 (2011).

[23] Yamanaka, M., "American history and digital history (<Features> digital humanities and American studies)," *Rikkyo American Studies*, **40**: 7–31 (2018).

[24] Kokaze, N., "A digital toolbox for historical researchers," *Clio: A Journal of European Studies*, **31**: 2–9 (2017).

[25] Murray, S., Statement of recommitment, Alabama Department of Archives and History, Retrieved from https://archives.alabama.gov/docs/ADAH_Statement_Recommitment_200623.pdf (2020 June 23).

[26] Anderson, E., "Epistemic justice as a virtue of social institutions," *Social Epistemology*, **26(2)**: 163–173 (2012).

[27] Samarzija, H., and Cerovac, I., "The institutional preconditions of epistemic justice," *Social Epistemology*, **35(6)**: 621–635 (2021).