Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 11 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i11.62797

SHORT PAPER

Investigation of Sources of Bias in Healthcare AI in South Korea

Haeun Eun, Adamantios Koumpis (✉)

University Hospital Cologne, Cologne, Germany

adamantios.koumpis@uk-koeln.de

ABSTRACT

South Korea is expanding the use of artificial intelligence (AI) in healthcare through large-scale healthcare data initiatives. As AI increasingly relies on medical datasets, concerns have emerged regarding potential bias. This study aims to examine how existing inequalities in South Korea may contribute to potential bias in healthcare AI and to identify the major factors that influence the representativeness of medical datasets. The research was conducted using academic articles, government reports, and news sources related to healthcare, digital technology, and AI in South Korea. The analysis focuses on five potential sources of healthcare AI bias: age and digital literacy, gender, socioeconomic status, regional healthcare disparities, and structural limitations of healthcare databases. Existing social and structural inequalities may influence the representativeness of healthcare data used to develop AI systems. Older adults may be underrepresented due to lower digital literacy and reduced use of digital health technologies. Gender differences in disease prevalence, symptom presentation, healthcare utilization, and medication use may contribute to unequal data representation. Socioeconomic and regional disparities may further affect healthcare access and data generation, while limitations of national health insurance databases may influence data quality and completeness.

KEYWORDS

healthcare bias, social and healthcare inequalities, gender bias, racial bias, age and digital literacy bias, bias in artificial intelligence (AI) systems

1 INTRODUCTION

To advance precision medicine and digital healthcare, South Korea's Ministry of Health and Welfare [1] launched the National Integrated Bio Data Infrastructure Project. This project securely stores participant data and biological samples in dedicated data banks and biobanks, opening access to university and hospital researchers starting in 2026. However, as large-scale medical datasets drive artificial intelligence (AI) advances, risks of algorithmic bias arise, particularly when training data

Eun, H., Koumpis, A. (2026). Investigation of Sources of Bias in Healthcare AI in South Korea. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(11), pp. 205–212. https://doi.org/10.3991/ijoe.v22i11.62797

Article submitted 2026-07-14. Revision uploaded 2026-08-08. Final acceptance 2026-08-08.

© 2026 by the authors of this article. Published under CC-BY.

lacks equal representation across population groups. For example, Chung et al. [5] demonstrated that a COVID-19 severity prediction model trained on single-gender data performed poorly when applied to the other gender.

To address this issue, this study investigates potential biases within South Korean healthcare data that could compromise the fairness and accuracy of future AI systems.

2 APPROACH AND METHODOLOGY

To examine how structural inequalities in South Korea shape bias in healthcare AI systems, this study adopted a narrative review approach enriched by a mixed-methods framework. Recognizing that AI bias stems from both statistical skewedness and underlying social contexts, our strategy integrated quantitative evidence (e.g., model accuracy metrics, demographic statistics, and prevalence rates) with qualitative insights (e.g., diagnostic practices, policy frameworks, and socioeconomic barriers).

Relevant sources were identified through systematic searches across academic databases (e.g., PubMed, Web of Science, RISS, DBpia), government archives, and institutional repositories. Literature selection was guided by specific eligibility criteria:

- Inclusion Criteria: (1) Peer-reviewed empirical studies, national survey datasets, and official policy reports focused on South Korean healthcare, digital health, or AI applications; (2) studies providing disaggregated demographic data (e.g., by age, sex, income, or geography); and (3) investigations detailing national database structures or clinical diagnostic disparities.
- Exclusion Criteria: Editorials, non-peer-reviewed commentaries, and studies lacking direct applicability to the South Korean healthcare context or data representativeness.

The collected materials—comprising peer-reviewed literature, government publications, and policy documents—were synthesized through thematic analysis. Combining quantitative data with qualitative contextualization enabled a multidimensional examination of five primary drivers of healthcare AI bias: (1) age and digital literacy, (2) gender and sex differences, (3) socioeconomic status, (4) regional healthcare disparities, and (5) structural limitations of national claims databases. By evaluating both numerical disparities and structural mechanisms, this approach illustrates how existing social inequalities directly influence the quality, availability, and representativeness of healthcare data used to train AI systems. The results of this investigation are detailed in the following section.

3 RESULTS

3.1 Age and digital literacy bias

In 2025, South Korea's population aged 65 and older reached 21.21% [2], marking its transition into a super-aged society. As a result, digital healthcare is rapidly expanding into routine health management and preventive care, positioning it as a key future industry [3]. However, severe digital access disparities persist among older adults. Many struggle to use mobile applications and wearable devices, leaving

them digitally vulnerable. Although the benefits of senior-focused digital healthcare are clear, policy support for digital literacy and accessibility only gained serious momentum in 2024 [3].

Digital literacy directly impacts health outcomes: seniors with higher digital skills are more likely to seek health information online, adopt healthier habits, and report better self-perceived health [4]. Yet, overall digital engagement remains low among older adults compared to younger demographics [11]. For instance, while smartphone and app usage is widespread among Korean seniors, wearable device adoption sits at just 7.1% [12]. Furthermore, many rely on family members to install and recommend apps, highlighting lingering barriers to independent technology use [12].

3.2 Gender bias

Training AI on single-gender datasets creates significant algorithmic bias. For instance, Chung et al. [5] demonstrated that when an early-stage COVID-19 severity prediction model was trained exclusively on one sex and applied to the other, its overall accuracy, sensitivity, and specificity dropped significantly. Beyond AI training data, gender bias persists in clinical diagnosis; Lee and Hong [6] note that conditions like ADHD are frequently underdiagnosed or delayed in females because they present with less visible symptoms (e.g., inattentiveness) rather than hyperactivity, leading clinicians to misinterpret them as mere personality traits.

Despite growing efforts to include women in clinical trials, research consistently fails to provide sex-stratified data analyses, ignoring key physiological differences [13]. Park et al. [13] highlighted these gaps in cardiovascular drug adverse event (AE) reporting. Women reported higher overall AE rates, particularly for gastrointestinal, psychiatric, and musculoskeletal disorders, yet lower rates of *serious* AEs compared to men. Higher gastrointestinal AEs in women were linked to greater stress susceptibility, anxiety, and higher baseline rates of conditions such as Crohn's disease and IBD. Meanwhile, musculoskeletal AEs in women were driven by postmenopausal fracture risks and cancer therapies (e.g., for breast and cervical cancers) [13].

Expanding on these physiological discrepancies, Rhee et al. [14] documented widespread sex differences in disease prevalence, medication usage, and healthcare utilization across South Korea:

- Disease Prevalence: Men showed significantly higher rates of obesity, hypertension, COPD, and hypertriglyceridemia. In contrast, women experienced higher rates of hypercholesterolemia, anemia, cataracts, osteoporosis, and osteoarthritis, with prevalence patterns shifting over their life course [14].
- Medication Use: Women used more medications overall across most therapeutic categories—most notably hormonal preparations, musculoskeletal drugs, antidepressants, diuretics, and gastrointestinal medications. While women's medication use rose steadily with age, men's peaked and then slightly declined [14].
- Healthcare Utilization: Outpatient analysis across 47 major diseases revealed female predominance in 33 conditions (such as thyroid, mental health, and connective tissue disorders) and male predominance in 14 conditions (predominantly cancers of the oral cavity, stomach, liver, and lung, as well as liver and stone diseases) [14].

3.3 Socioeconomic bias

Socioeconomic disparities in health status and healthcare utilization remain a major social challenge in South Korea [7]. Key factors, such as income, education, occupation, and employment stability, directly shape individual health, medical access, and healthcare utilization. Individuals with lower socioeconomic status consistently experience poorer health outcomes, higher disease prevalence, and elevated hospitalization rates. Moreover, limited access to timely care among low-income and precariously employed groups reinforces a cycle of poor health and economic disadvantage [7].

These socioeconomic divides are equally reflected in health literacy. A 2023 national survey revealed that only 60.4% of Korean adults possessed adequate health literacy, with the lowest levels observed among older adults and low-income, less-educated populations [15]. As Moon et al. [16] highlight, low health literacy hinders individuals from navigating complex health information and identifying when or where to seek care, ultimately driving distinct, often sub-optimal, healthcare utilization patterns.

3.4 Regional bias

Regional healthcare disparities pose a critical challenge in South Korea [8]. The current compensation system relies on uniform reimbursement rates based on utilization volume, offering no adjustments for regional differences. Consequently, rural and non-metropolitan areas, which have smaller populations and lower medical demand, face structural unprofitability and resource shortages. This problem is further compounded by a heavy reliance on private healthcare providers, who naturally concentrate facilities and specialists in economically lucrative metropolitan centers, primarily the Seoul capital region [8, 17].

As a result, medical infrastructure has become deeply centralized, severely restricting access to quality and specialized care in smaller regions. Highlighting this stark imbalance, Jeong et al. [17] reported that as of 2018, 152 non-metropolitan municipalities had either no general hospital or only a single one.

3.5 Structural bias

South Korea operates a universal healthcare system managed by a single insurer, the National Health Insurance Corporation (NHIC), giving it access to comprehensive, population-wide health data [9]. However, because this database was designed for insurance claims and reimbursements rather than research, its analytical utility is limited [9]. The data lacks coverage for non-insured treatments and makes establishing causality difficult. Additionally, Kyoung and Kim [10] note that disease severity can easily be distorted, as diagnostic coding is often tied directly to insurance benefit requirements.

4 DISCUSSION

Our investigation suggests that several existing inequalities in South Korea may influence the development of healthcare AI systems by affecting the representativeness of healthcare data. The findings of our work indicate that bias may already be

embedded in the data collection process before AI models are developed. In Table 1 we have provided an overview of all bibliographic sources that has been used for the scope of our work.

Table 1. Bibliographic sources considered in the conducted investigation

Study	Population/Data Source	Bias Category
Kim and Kang [3]	Older adults in South Korea	Age and Digital Literacy
Lee et al. [4]	Adults aged 65 years and older in South Korea	Age and Digital Literacy
Hwang et al. [11]	Smartphone-owning older adults living alone in South Korea	Age and Digital Literacy
Lee et al. [12]	Adults aged 65 years and older in South Korea	Age and Digital Literacy
Chung et al. [5]	COVID-19 patient records	Gender
Lee and Hong [6]	Patients with ADHD in South Korea	Gender
Park et al. [13]	AE reporting database	Gender
Rhee et al. [14]	National statistics of the Korean population	Gender
Gim [7]	Korean adults aged over 25/Korea National Health and Nutrition Examination Survey (KNHANES)	Socioeconomic
Choi et al. [15]	Korean adults aged 19 or older/National health survey/KNHANES	Socioeconomic
Moon et al. [16]	Korean adults aged 19 or older/Korea Health Panel	Socioeconomic
Jung et al. [8]	Domestic and international healthcare policy documents, incentive programs, and expert consultations	Regional
Jeong and Jun [17]	The Korea Transport Institute transportation accessibility indicators	Regional
Kim et al. [9]	The Korean National Health Information Database (NHID)	Structural
Kyoung and Kim [10]	Korean National Health Insurance Service (NHIS) and Health Insurance Review and Assessment (HIRA) claims databases	Structural

The findings show that age-related differences in digital literacy may affect participation in digital healthcare technologies and the generation of digital health data. As digital healthcare services become increasingly integrated into healthcare delivery, populations with limited digital access may contribute less data to healthcare databases. Consequently, AI systems trained on these datasets may be less representative of older adults despite their growing proportion within the Korean population.

The investigation also identified substantial gender differences in disease prevalence, symptom presentation, medication use, adverse drug reactions, and healthcare utilization. Datasets may capture substantially different health profiles for male and female patients, potentially contributing to sex-related bias in healthcare AI systems. These findings also suggest that healthcare datasets do not simply reflect biological differences but also reflect social and clinical practices that influence diagnosis and treatment. If such differences are not adequately considered during AI development, healthcare AI systems may produce unequal performance across genders.

Furthermore, socioeconomic and regional disparities may influence the quantity and quality of healthcare data available for different population groups. Individuals with lower health literacy, lower income, or limited healthcare accessibility may interact differently with healthcare services, resulting in gaps within healthcare datasets. Similarly, regional inequalities in healthcare infrastructure may lead to the underrepresentation of rural populations and medically underserved communities in AI training data.

Bias may correctly seem as unavoidable when considering real-world AI applications. While we can dramatically *mitigate* bias through careful data curation, algorithmic constraints, and continuous auditing, total elimination remains rather elusive. This stems from three fundamental realities:

- *Data reflect historical reality*: this practically means that all data 'fed' into AI reflects inequalities such as e.g. underrepresentation of minorities in clinical trials.
- *There is always a fairness trade-off one needs to face*: AI developers may prioritize one definition of fairness over another, so there will be always a decision that is inherently bias-laden and never neutral.
- *Human framing is also unavoidable*: AI systems do not exist in a vacuum, they are designed, labelled, and deployed by humans and reflect the Zeitgeist within they were given birth to.

Last but not least, a model that may be considered as 'unbiased' at one point or in one context or for one population may perform poorly or exhibit bias when transferred to another time setting or context or population.

5 CONCLUSIONS

For the area of medical AI bias is a major difficulty that affects several phases of data collection, processing, and model building [18]. Under the imperative of digital transformation in the field of health and medical education, it is also evident that AI technologies apart from enhancing learning and educational outcomes can also streamline other functions and practices [19], such as to name only one, the application of digital twins that can act as a game change in the field of clinical decision making [20].

The findings of our research highlight the importance of considering the characteristics of healthcare databases themselves. The purpose of these databases creates limitations regarding disease severity, uninsured treatments, and clinical detail. Therefore, even large-scale datasets may not necessarily provide a complete representation of healthcare experiences. Taken together, these findings suggest that healthcare AI bias should be understood as a consequence of broader social, economic, and healthcare inequalities. AI systems trained on existing healthcare data may unintentionally learn patterns that reflect historical disparities in healthcare access, diagnosis, treatment, and data collection. Therefore, improving algorithmic fairness alone may be insufficient if inequalities in healthcare data generation and collection remain unaddressed.

6 REFERENCES

[1] Ministry of Health and Welfare, "Launch of the national integrated bio big data project involving one million participants," 2024. [Online]. Available: https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&act=view&list_no=1483988

[2] H. Moon, "Nationwide population aged 65 and older surpasses 21%... Seoul also enters a super-aged society," 2026. [Online]. Available: https://www.joongang.co.kr/article/25394969

[3] C. Kim and I. Kang, "A study on policy directions for digital literacy to enhance the utilization of digital healthcare by the elderly," *JSEL*, vol. 27, no. 2, pp. 3–25, 2024.

[4] S. Lee, Y. Kim, B. Kim, S. Lee, S. Jang, and T. Kim, "Digital literacy and its association with subjective health status and healthy lifestyle behaviors among korean older adults: Cross-sectional study," *JMIR Aging*, vol. 8, Art. no. e64974, 2025. https://doi.org/10.2196/64974

[5] H. Chung, C. Park, W. Kang, and J. Lee, "Gender bias in artificial intelligence: Severity prediction at an early stage of COVID-19," *Front. Physio.*, vol. 12, Art. no. 778720, 2021. https://doi.org/10.3389/fphys.2021.778720

[6] J. Lee and S. Hong, "Mistaken for simple forgetfulness: Why ADHD diagnosis in women is often delayed," IT/Medicine, 2026. [Online]. Available: https://www.donga.com/news/Health/article/all/20260107/133106387/1 [Accessed: May 30, 2026].

[7] D. Gim, "A study on the effects of socioeconomic status on health level and medical service use," *IHSS*, vol. 16, no. 4, pp. 329–369, 2015. https://doi.org/10.15818/ihss.2015.16.4.329

[8] S. Jung, H. Lee, H. Shin, and H. Kim, "A study on improving compensation systems to reduce regional healthcare disparities," Researcj@KIHASA, 2025. https://doi.org/10.23060/kihasa.b.2025.04

[9] M. Kim, K. Han, and S. Lee, "Current trends of big data research using the korean national health information database," *DMJ*, vol. 46, no. 4, pp. 552–563, 2022. https://doi.org/10.4093/dmj.2022.0193

[10] D. Kyoung and H. Kim, "Understanding and Utilizing Claim Data from the Korean National Health Insurance Service (NHIS) and Health Insurance Review & Assessment (HIRA) Database for Research," *JLA*, vol. 11, no. 2, pp. 103–110, 2022. https://doi.org/10.12997/jla.2022.11.2.103

[11] M. Hwang, G. Kim, S. Lee, and Y. H. Park, "Digital health literacy and associated factors among older adults living alone in South Korea: A cross-sectional study," *Research in Community and Public Health Nursing*, vol. 35, no. 4, pp. 389–400, 2024. https://doi.org/10.12799/rcphn.2024.00766

[12] H. Lee *et al.*, "Digital health technology use among older adults: Exploring the impact of frailty on utilization, purpose, and satisfaction in Korea," *J. Korean Med. Sci.*, vol. 39, no. 1, Art. no. e7, 2024. https://doi.org/10.3346/jkms.2024.39.e7

[13] H. Park, J. Kim, D. Yoon, H. Lee, and J. Shin, "Gender differences in the adverse events associated with cardiovascular drugs in a spontaneous reporting system in South Korea," *IJCP*, vol. 43, no. 4, pp. 1036–1044, 2021. https://doi.org/10.1007/s11096-020-01212-z

[14] S. Rhee, B. Lee, and K. Kwon, "Investigation study on gender difference based on korean data related to drug use," *Kor. J. Cli. Pharm.*, vol. 23, no. 2, pp. 114–122, 2013.

[15] S. Choi, Y. Kang, H. Kim, and K. Oh, "Health literacy in Korea: Findings from the 2023 korea national health and nutrition examination survey," *Epidemiol. Health*, vol. 47, Art. no. e2025037, 2025. https://doi.org/10.4178/epih.e2025037

[16] S. Moon, J. Kim, Y. Choi, and H. Kim, "Influencing factors on health literacy and the influence of health literacy on healthcare utilization," *KIHASA*, vol. 44, no. 4, pp. 130–150, 2024. https://doi.org/10.15709/hswr.2024.44.4.130

[17] H. Jeong and H. Jun, "Spatial disparities and contributing factors in medical facility accessibility: A comparative analysis across different-sized medical facilities," *J. of Korea Plan. Assoc.*, vol. 59, no. 3, pp. 74–97, 2024. https://doi.org/10.17208/jkpa.2024.06.59.3.74

[18] A. P. Parate *et al.*, "Review of data bias in healthcare applications," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 12, pp. 124–136, 2024. https://doi.org/10.3991/ijoe.v20i12.49997

[19] D. Kazimova, G. Tazhigulova, G. Shraimanova, A. Zatyneyko, and A. Sharzadin, "Transforming university education with AI: A systematic review of technologies, applications, and implications," *International Journal of Engineering Pedagogy (iJEP)*, vol. 15, no. 1, pp. 4–24, 2025. https://doi.org/10.3991/ijep.v15i1.50773

[20] M. Knio, G. Kaur, P. Pujari, A. Kumar, and N. Chaudhary, "Digital twins in biomedical engineering: A systematic review of applications and future challenges," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 22, no. 6, pp. 91–105, 2026. https://doi.org/10.3991/ijoe.v22i06.61523

7 AUTHORS

Haeun Eun is student of medicine at the Medical Faculty of the University of Cologne. For her scientific project (WissPro1) she worked with the Biomedical Informatics Institute (E-mail: heun@smail.uni-koeln.de).

Adamantios Koumpis is Teaching and Research Officer at the Biomedical Informatics Institute of the University Hospital Cologne (E-mail: adamantios.koumpis@uk-koeln.de).