# RESEARCH



# Al and inclusion in simulation education and leadership: a global cross-sectional evaluation of diversity



Joana Berger-Estilita<sup>1,2†</sup>, Mia Gisselbaek<sup>3,4,5\*†</sup>, Arnout Devos<sup>6</sup>, Albert Chan<sup>7</sup>, Pier Luigi Ingrassia<sup>8</sup>, Basak Ceyda Meco<sup>9,10</sup>, Odmara L. Barreto Chang<sup>11</sup>, Georges L. Savoldelli<sup>3,4,5</sup>, Francisco Maio Matos<sup>12,13,14</sup>, Peter Dieckmann<sup>15,16,17</sup>, Doris Østergaard<sup>15,18</sup> and Sarah Saxena<sup>19,20</sup>

# Abstract

**Background** Simulation-based medical education (SBME) is a critical training tool in healthcare, shaping learners' skills, professional identities, and inclusivity. Leadership demographics in SBME, including age, gender, race/ethnicity, and medical specialties, influence program design and learner outcomes. Artificial intelligence (AI) platforms increasingly generate demographic data, but their biases may perpetuate inequities in representation. This study evaluated the demographic profiles of simulation instructors and heads of simulation labs generated by three AI platforms— ChatGPT, Gemini, and Claude—across nine global locations.

**Methods** A global cross-sectional study was conducted over 5 days (November 2024). Standardized English prompts were used to generate demographic profiles of simulation instructors and heads of simulation labs from ChatGPT, Gemini, and Claude. Outputs included age, gender, race/ethnicity, and medical specialty data for 2014 instructors and 1880 lab heads. Statistical analyses included ANOVA for continuous variables and chi-square tests for categorical data, with Bonferroni corrections for multiple comparisons: *P* significant < 0.05.

**Results** Significant demographic differences were observed among AI platforms. Claude profiles depicted older heads of simulation labs (mean: 57 years) compared to instructors (mean: 41 years), while ChatGPT and Gemini showed smaller age gaps. Gender representation varied, with ChatGPT and Gemini generating balanced profiles, while Claude showed a male predominance (63.5%) among lab heads. ChatGPT and Gemini outputs reflected greater racial diversity, with up to 24.4% Black and 20.6% Hispanic/Latin representation, while Claude predominantly featured White profiles (47.8%). Specialty preferences also differed, with Claude favoring anesthesiology and surgery, whereas ChatGPT and Gemini offered broader interdisciplinary representation.

**Conclusions** Al-generated demographic profiles of SBME leadership reveal biases that may reinforce inequities in healthcare education. ChatGPT and Gemini demonstrated broader diversity in age, gender, and race, while Claude skewed towards older, White, and male profiles, particularly for leadership roles. Addressing these biases through ethical AI development, enhanced AI literacy, and promoting diverse leadership in SBME are essential to fostering equitable and inclusive training environments.

<sup>†</sup>Joana Berger-Estilita and Mia Gisselbaek contributed equally and share first authorship.

\*Correspondence: Mia Gisselbaek Mia.gisselbaek@gmail.com Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/A.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Trial registration Not applicable. This study exclusively used AI-generated synthetic data.

**Keywords** Simulation-based medical education, Artificial intelligence (AI), Diversity and inclusion, Gender representation, Healthcare education, Bias, Ethical AI design

# Background

Simulation-based medical education (SBME) has emerged as a cornerstone in training healthcare professionals, providing a potentially safe, controlled environment for skill acquisition, decision-making, and reflective learning [1, 2]. This method can enhance technical competencies and foster deeper self-awareness and interpersonal growth. The dual process of experiential learning—combining episodes of hands-on practice with reflective thinking—shapes not only *what* participants do but also *how* they perceive themselves and others within the healthcare ecosystem [3, 4].

In the current climate where diversity, equity, and inclusion policies are questioned, [5] the cultural diversity of simulation instructors remains fundamental, as they directly engage with learners and guide the debriefing process, where cultural competence is essential to avoid potential harm [6-9]. Instructors with diverse backgrounds may better connect with learners' varied cultural experiences, ensuring more inclusive simulation sessions. Moreover, the demographic characteristics of simulation lab leaders also significantly influence this process. As role models, leaders embody traits and behaviors that learners may internalize, shaping their professional identity and sense of belonging [10]. This might be especially relevant for younger generations who are interested in finding their professions and disciplines. A lack of diversity in leadership can implicitly signal exclusionary norms, discouraging individuals from underrepresented groups from envisioning themselves in similar roles [11]. Conversely, diverse leadership fosters inclusivity, offering relatable role models and perspectives that resonate with a broader range of learners [12].

Leaders' demographics may shape decisions about which scenarios are prioritized, how they are designed, and whose perspectives are centered—factors critical to ensuring that simulation programs address culturally sensitive care, health equity, and interdisciplinary collaboration [13, 14].

Artificial intelligence (AI) is increasingly integrated into healthcare systems, supporting decision-making, problem-solving, and even generating educational content [15]. However, AI's algorithmic biases can perpetuate harmful stereotypes related to gender, race/ ethnicity, and age, exacerbating systemic inequities [16– 21]. These biases are particularly concerning when AI is used to depict or inform leadership demographics, as misrepresentations could reinforce exclusionary norms in simulation-based education and beyond.

Recent discussions highlight the importance of diversity, equity, accessibility, and inclusion (DEAI) in healthcare leadership, including in simulation settings [22–25]. Understanding how AI describes the age, gender, race/ ethnicity, and professional specialties of simulation (lab) leaders is critical to identifying potential biases. This can provide insights into gaps in representation and the implications for simulation practices, such as learners' ability to identify with leaders and receive emotional support.

This study investigates the AI description of simulation instructors' and lab heads' demographic and professional profiles. By analyzing these characteristics, it explores the relationship between diversity and stereotypes—similar to those seen in media—which impact identity formation and belonging. We choose one of myriads of possible areas and constellations to investigate the principle, which may apply in many other areas.

# Methods

# Ethics

For this project, no approval from an ethics committee was required, as it exclusively involves the use of artificial intelligence-generated or synthetic data. No human participants, personally identifiable information, or sensitive real-world data were involved in the study. The nature of the data ensures that ethical considerations related to human subject research do not apply. This study adhered to the Declaration of Helsinki, and researchers followed the Data Protection Acts of their respective academic institutions [26, 27]. The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline [28].

### Study design and setting

A global cross-sectional web-based study design was used in this study. Data collection was conducted during five consecutive days (November 8–13, 2024). AI models investigating diversity (age, race/ethnicity, gender, and medical specialty) in simulation labs generated tables in nine locations (Angola, Belgium, Brazil, Caribbean, Hong Kong, Italy, Switzerland, Turkey, USA). Different countries were included to represent real-world coverage; differences in results could arise due to country-specific linguistic and cultural biases in AI models used in different locations, leading to diverse and adaptive responses [29].

## AI model data generation

The following two prompts were used to generate tables in the most commonly used AI models: ChatGPT 4 (available at https://openai.com/blog/chatgpt from OpenAI, San Francisco, CA, USA), Gemini (available at https://gemini.google.com/ from Alphabet Inc., Mountain View, CA, USA), and Claude (available at https:// claude.ai/ from Anthropic, San Francisco, CA, USA). Each request was entered individually in a new dialogue box: "A table with 100 times the age/gender/race/medical specialty of a simulation instructor" and "A table with 100 times the age/gender/race/medical specialty of the head of a simulation lab" In total, 45 tables were generated by entering prompts into three large language model(s) (LLM) ChatGPT4, Gemini, and Claude, in each of the 9 locations (9 countries, 3 systems, and 2 queries). The responses generated by each LLM were collected in a Google document file (Alphabet Inc., Mountain View, CA, USA). As tables were generated with each demographic variable, no further interpretation or classification was needed before statistical analysis.

# Statistical analysis

Statistical analysis considered both continuous variables, such as age, and categorical variables, including gender, race, and specialty preferences. Continuous data were summarized using means and standard deviations (SD), while categorical variables, including gender, race, and specialty preferences, were expressed as frequencies and percentages. Comparisons of mean age across platforms and roles were performed using analysis of variance (ANOVA), and differences in the distribution of categorical variables were assessed with chi-square tests of independence. Statistical significance was p < 0.05, with highly significant results reported as p < 0.001. Given the number of comparisons made, a Bonferroni correction was applied to adjust for multiple analyses, reducing the risk of type I error. For age comparisons, the adjusted threshold for significance was set to p < 0.017 (three pairwise comparisons), and for categorical variables, adjustments were made based on the number of categories compared. All statistical analyses were performed using IBM SPSS Statistics version 27 (IBM Corp., Armonk, NY, USA).

# Results

Forty-five ratings were collected between November 8 and 13, 2024, representing 3894 entries. Due to ethical issues, Claude AI refused to produce tables in Turkey

and Switzerland (Additional File 1). In several countries, Gemini provided only partial datasets. The original tables obtained from those countries are available in the Additional File 2.

For simulation instructors (Table 1), ChatGPT and Claude's outputs were younger compared to Gemini outputs (41.4 and 40.5 versus 47.9 years; p < 0.001). Women represented about half of the simulation instructor profiles in all three models. Gemini showed increased gender diversity when compared with the other two AI models and represented "non-binary" (6.8%) and "other" (4.3%) genders in higher proportions. Racial diversity is higher among ChatGPT- and Gemini-generated profiles, while Claude-generated profiles were predominantly "White"/"Asian" (34.4%) (Table 1, Fig. 1).

Regarding specialty preferences, all 3 models represented 27 different specialties. ChatGPT represented surgery as the top specialty, while Gemini and Claude showed more emergency medicine physicians. A complete description of specialty results can be found in Additional File 3.

For heads of simulation labs (Table 2), Claude's outputs were significantly older.

Tab	le	1	ΑI	profile	es of	simu	lation	instru	uctors
-----	----	---	----	---------	-------	------	--------	--------	--------

Sim instructor Al profile	ChatGPT ( <i>n</i> = 850)	Gemini ( <i>n</i> = 644)	Claude ( <i>n</i> = 520)	<i>p</i> -value
Age in years, mean (SD)	41.4 (8.9)	47.9 (8.3)	40.5 (9.7)	< 0.001
Gender (as per LLM output)				< 0.001
Female	425 (50.0)	329 (51.1)	264 (50.8)	
Male	424 (49.9)	243 (37.7)	256 (49.2)	
Non-binary	1 (0.1)	44 (6.8)	0 (0.0)	
Other	0 (0.0)	28 (4.3)	0 (0.0)	
Race/ethnicity				< 0.001
White	246 (28.9)	196 (30.4)	179 (34.4)	
Asian	211 (24.8)	197 (30.6)	145 (27.9)	
Black	207 (24.4)	108 (16.8)	103 (19.8)	
Hispanic/Latin	175 (20.6)	99 (15.4)	82 (15.8)	
Undetermined	11 (1.3)	44 (6.8)	11 (2.1)	
Specialty				< 0.001
Anesthesiology (n = 180)	62 (7.3)	59 (9.2)	59 (11.3)	
Emergency medicine (n = 265)	73 (8.6)	127 (19.7)	65 (12.5)	
Surgery (all types) (n = 203)	86 (10.1)	54 (8.4)	63 (12.1)	
Internal medicine (n = 188)	73 (8.6)	53 (8.2)	62 (11.9)	
Pediatrics ( $n = 201$ )	83 (9.8)	59 (9.2)	59 (11.3)	
All others (n = 979)*	473 (55.6)	292 (45.3)	214 (40.9)	

\* Specialties mentioned, with less than 7% of representation: radiology, cardiology, obstetrics and gynecology, neurology, psychiatry, family medicine, dermatology, orthopedics, ophthalmology, pathology, critical care, gastroenterology, pulmonology, oncology, infectious diseases, endocrinology, rheumatology, trauma surgery, nursing, otolaryngology (ENT), and urology



Fig. 1 Gender and racial/ethnic diversity across AI platforms for simulation instructors

Gender representation differed across platforms (see Table 2, Fig. 2).

A total of 36 specialties were represented. Specialty preferences highlighted a concentration among Claude outputs in surgery (22.4%), anesthesiology (22.7%), and emergency medicine (20.7%). In contrast, specialties in ChatGPT and Gemini were more broadly distributed (Additional File 3).

When comparing both profiles (simulation instructor vs head of simulation lab), the Claude model demonstrated the most significant age gap, with the heads of simulation labs being much older than simulation instructors. As for gender, ChatGPT and Gemini maintained similar gender trends across roles, but Claude shifted from gender balance in simulation instructors to male dominance among heads of simulation labs.

Regarding racial/ethnic diversity, Claude consistently showed a lack of diversity in both roles, with the "White" majority increasing for heads of simulation labs. ChatGPT and Gemini maintained consistent diversity across roles. Claude showed decreased diversity for its heads of simulation labs.

The specialty preferences showed Claude shifting towards anesthesiology as head of simulation labs,

while ChatGPT and Gemini consistently exhibited a broader specialty distribution across roles.

# Discussion

This study highlights notable differences in demographic and specialty patterns across AI-generated profiles of simulation instructors and sim lab directors. ChatGPT showed consistent diversity, with balanced gender representation and broad racial diversity. Gemini also maintained gender balance but exhibited less racial diversity. Claude showed a significant older demographic in leadership roles and exhibited less racial diversity, with a predominant "White" majority among heads of simulation labs. Gender dynamics also shifted with Claude, transitioning from balance among simulation instructors to male predominance in heads of simulation labs.

Specialty preferences also varied. Claude emphasized anesthesiology, surgery, and emergency medicine, while ChatGPT and Gemini favored multidisciplinary specialties, reflecting a broader approach to simulation education.

These varied cultural nuances in the outputs of AI models such as ChatGPT, Gemini, and Claude likely stem from their distinct training methodologies. Such biases

Table 2 Al prof	les of head (	of simul	lation l	abs
-----------------	---------------	----------	----------	-----

Head of sim lab Al profile	ChatGPT ( <i>n</i> = 848)	Gemini ( <i>n</i> = 624)	Claude ( <i>n</i> = 406)	<i>p</i> -value
Age in years, mean (SD)	47.5 (7.45)	42.0 (8.69)	57.0 (5.28)	< 0.001
Gender (as per LLM outputs)				< 0.001
Female	420 (49.5)	333 (53.4)	148 (36.5)	
Male	442 (49.8)	252 (40.4)	258 (63.5)	
Non-binary	6 (0.7)	39 (6.3)	0 (0.0)	
Other	-	-	-	
Race/ethnicity				< 0.001
White	238 (28.1)	204 (32.7)	194 (47.8)	
Asian	220 (25.9)	166 (26.6)	110 (27.1)	
Black	186 (21.9)	105 (16.8)	48 (11.8)	
Hispanic/Latin	171 (20.2)	100 (16.0)	40 (9.9)	
Undetermined	33 (3.9)	49 (7.9)	14 (3.4)	
Specialty				< 0.001
Anesthesiology (n = 187)	47 (5.6)	48 (7.7)	92 (22.7)	
Emergency medicine (n = 265)	60 (7.2)	121 (19.4)	84 (20.7)	
Surgery (all types) (n = 221)	87 (10.4)	43 (6.9)	91 (22.4)	
Internal medicine (n = 175)	73 (8.7)	58 (9.3)	44 (10.8)	
Pediatrics (n = 156)	74 (8.9)	56 (9.0)	26 (6.4)	
All others $(n = 861)^*$	494 (59.2)	298 (47.8)	69 (17.0)	

\* Specialties mentioned, with less than 5% of representation: radiology, cardiology, obstetrics and gynecology, neurology, psychiatry, family medicine, dermatology, orthopedics, ophthalmology, pathology, critical care, gastroenterology, pulmonology, oncology, infectious diseases, endocrinology, rheumatology, trauma surgery, urology, allergy and immunology, geriatrics, pain management, palliative care, public health, plastic surgery, otolaryngology (ENT), nursing, medical education, simulation technology, and hematology

have been shown to stem from the predominantly Euro-American-centric data used in training, which can overlook or misrepresent local contexts in other regions. This highlights the necessity for developing regionally adapted large language models that better capture and reflect diverse global perspectives [30, 31].

Interestingly, when asked to generate gender demographics, Gemini and ChatGPT used the terms "male," "female," and "non-binary." While this represents a step towards inclusivity, it does not fully align with current guidelines, which recommend focusing on gender identity and avoiding terms suggesting a strictly biological or binary framework [32].

AI-generated descriptions of simulation lab leaders revealed an overrepresentation of specific medical specialties, particularly those emphasizing procedural skills and crisis management, such as surgery, anesthesiology, and emergency medicine [33, 34]. These fields align closely with the traditional focus of SBME on technical skills as well as life-and-death scenarios. They also represent pioneering people and disciplines in the field of simulation [35]. However, specialties such as psychiatry, family medicine, or pediatrics were less frequently highlighted [36, 37]. This disparity raises essential questions about representation and the influence of SBME leadership on shaping priorities and practices within the field. Representation in this context refers to who leads and how their leadership influences the design and priorities of simulation programs [38]. However, the extent and nature of this influence depend significantly on the leader's approach, their receptiveness to diverse viewpoints, and the specific context in which they operate.

Practical leadership skills taught in SBME encompass technical expertise, strong decision-making, and interpersonal skills [39]. Sim lab instructors manage complex team dynamics, enhance collaboration, and create an inclusive environment. Integrating AI into this landscape introduces opportunities and challenges, making ethical considerations paramount. Virtual reality scenarios offer a powerful tool for training on racial sensitivity and inclusivity, providing immersive experiences that help build empathy and understanding [40]. Similarly, voiceinteractive technologies can facilitate dynamic, reactive scenarios, allowing participants to engage in real-time conversations that closely simulate real-world challenges [41]. With this broad and growing application of AI systems, it is all the more important to reflect on where they guide attention and what they focus on or hide.

The integration of AI in SBME presents another evolving area of academic inquiry, particularly as a premise for promoting equity [42]. Incorporating tools and training focused on diversity, equity, accessibility, and inclusion (DEAI) is a powerful way to counteract previously described biases. By intentionally designing AI systems to highlight underrepresented specialties and demographics, these tools could challenge stereotypes and broaden perspectives on leadership [43]. This could influence perceptions of who "belongs" in leadership roles, potentially promoting a culture of inclusivity and mitigating harmful stereotypes [44]. Moreover, training in AI technologies could equip leaders and designers to develop fair, inclusive simulations. By embedding bias awareness into AI-driven practices, these efforts foster a culture of inclusivity and actively challenge harmful stereotypes.

The EU Artificial Intelligence Act underscores the importance of AI literacy, raising significant questions about equipping educators and practitioners with the knowledge required to use AI responsibly and effectively. Comprehensive "user manuals" that include detailed explanations of algorithms, training data, intended use cases, known limitations, and ethical considerations could be a foundational step towards this goal. Additionally, strategies such as country-specific prompting



Fig. 2 Gender and racial/ethnic diversity across AI platforms for heads of simulation labs

to address cultural and linguistic biases warrant further investigation to evaluate their efficacy in reducing disparities in AI applications [29].

A central aspect of this discourse is the role of AI literacy training for AI users. The EU Artificial Intelligence Act [45] underscores the importance of AI literacy, raising significant questions about the knowledge required to use AI responsibly and effectively. In medical education, particularly in SBME, AI literacy is vital in addressing bias. Proposals are underway to deploy a comprehensive "user manual," ensuring each user employs it fairly and with knowledge. This approach, rooted in the concept of AI literacy, is essential for democratizing the understanding of these complex technologies [46]. Providing comprehensive information about AI systems-such as their algorithms, training data, and limitations-opens the door for users, including healthcare professionals, educators, and the general public, to make more informed decisions about their use. This approach raises interesting questions about how best to address potential biases inherent in AI systems. For instance, some researchers have suggested strategies like country-specific prompting to reduce cultural and linguistic biases[29]. Exploring these and other methods could help determine how AI can adapt to diverse contexts, fostering greater fairness and inclusivity while maintaining effectiveness.

Finally, integrating agentic AI into SBME opens a new avenue for research, offering opportunities to explore how autonomous systems can transform how healthcare professionals are trained. Agentic AI refers to artificial intelligence systems designed to act as autonomous agents, capable of perceiving their environment, making decisions, and performing tasks independently to achieve specific goals. Integrating agentic AI into SBME can significantly enhance the realism and adaptability of training scenarios [41]. For instance, the development of systems like "AIPatient," which utilizes a knowledge graph derived from electronic health records and a reasoning retrievalaugmented generation workflow, enables the creation of advanced simulated patients that closely mimic realworld clinical conditions [47]. Additionally, the MEDCO framework employs a multi-agent approach to emulate complex medical training environments, facilitating more comprehensive and interactive learning experiences for healthcare professionals [48]. However, as agentic AI systems make autonomous decisions, they also bring ethical considerations, including accountability for errors, the transparency of their decision-making processes, and the potential to inadvertently reinforce biases embedded in their programming or training data [49]. Incorporating agentic AI into SBME requires deliberate efforts to align these systems with principles of fairness and inclusivity. When thoughtfully implemented, with previous AI literacy training, ethical principles included, agentic AI can expand the scope of simulation scenarios, improve accessibility to training, and foster the development of critical competencies in healthcare professionals.

Our study has limitations. First, we used only two English-language prompts, limiting the results' generalizability to other languages and cultural contexts. AI outputs often reflect biases inherent in language and culture, and this narrow linguistic focus may fail to capture these nuances in different settings. Second, our prompt was deliberately general and did not specifically instruct the AI models to consider gender, race/ethnicity, or specialty diversity. While intentional prompt engineering could potentially influence the output, our previous research has shown that even when bias is explicitly addressed, the response may not change meaningfully. For example, when we asked ChatGPT's DALL-E 2 why all generated images of department heads were White and male, the model acknowledged the lack of diversity but continued to produce similarly homogeneous outputs thereafter [18]. Additionally, the study used a cross-sectional design, evaluating three AI models at a single time point. Given that AI systems undergo continuous updates and iterations, the findings may not represent these models' latest advancements or improvements.

Another significant limitation lies in the exclusive reliance on AI-generated demographic profiles, which may not accurately align with real-world leadership demographics in SBME. This reliance creates a disconnect between the theoretical representations produced by AI and the actual diversity of SBME leadership. Our efforts to bridge this gap by scoping global demographic, racial, and gender data from major SBME societies were hindered by incomplete datasets or unavailability, as some organizations did not collect such data, did not respond to requests, or unfortunately did not want to hand over this data. This lack of real-world data restricts the findings to simulated or theoretical observations, limiting their practical applicability and continues to obscure or underrepresent existing inequities.

Furthermore, while the study included an analysis of speciality representation, it did not explore the complexities of how biases in AI-generated speciality profiles might influence perceptions of leadership or program design in SBME. For instance, AI's representation of certain specialities may inadvertently perpetuate stereotypes or misalign with real-world practices, potentially affecting educational outcomes or leadership development. Although we cannot report how the data generated in the system compares to the "real world," the differences generated are appalling.

At a time when DEI policies are being questioned, the lack of representation in AI platforms can have realworld consequences. A diverse healthcare workforce is known to be better equipped to address patient safety, our ultimate goal [5, 50].

## Conclusion

This cross-sectional study reveals that commonly used AI platforms may exhibit significant biases in representing the demographics of instructors and heads of labs within simulation-based medical education, mirroring systemic inequities. While ChatGPT and Gemini showed broader diversity in age, gender, and race, Claude's outputs leaned towards older, predominantly White, and male profiles, particularly for leadership roles.

These patterns emphasize the critical influence of AIgenerated perceptions on shaping professional identities and inclusivity in healthcare education. Addressing these challenges requires integrating ethical AI principles, enhancing AI literacy, and fostering diverse leadership within SBME. By leveraging AI thoughtfully, the medical education field can create equitable learning environments that reflect the diversity of modern healthcare and inspire underrepresented groups to pursue leadership roles.

#### Abbreviations

Analysis of variance						
Artificial intelligence						
Diversity, equity, accessibility, and inclusion						
Large language	Large language model					
Simulation-based medical education						
Standard deviat	ion					
Strengthening Epidemiology	the	Reporting	of	Observational	Studies	in
	Analysis of varia Artificial intellig Diversity, equity Large language Simulation-base Standard deviat Strengthening Epidemiology	Analysis of variance Artificial intelligence Diversity, equity, acce Large language mode Simulation-based me Standard deviation Strengthening the Epidemiology	Analysis of variance Artificial intelligence Diversity, equity, accessibility, and Large language model Simulation-based medical educat Standard deviation Strengthening the Reporting Epidemiology	Analysis of variance Artificial intelligence Diversity, equity, accessibility, and inclu Large language model Simulation-based medical education Standard deviation Strengthening the Reporting of Epidemiology	Analysis of variance Artificial intelligence Diversity, equity, accessibility, and inclusion Large language model Simulation-based medical education Standard deviation Strengthening the Reporting of Observational Epidemiology	Analysis of variance Artificial intelligence Diversity, equity, accessibility, and inclusion Large language model Simulation-based medical education Standard deviation Strengthening the Reporting of Observational Studies Epidemiology

# **Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s41077-025-00355-1.

Supplementary Material 1	
Supplementary Material 2	
Supplementary Material 3	

#### Acknowledgements

The authors would like to thank Dr. Marta Assuncao for her valuable input.

#### **Clinical trial registration**

Not applicable.

#### Declaration of generative AI in scientific writing

During the preparation of this work, the authors used ChatGPT 4.0 to improve readability. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

## Authors' contributions

JBE, MG, AD, PLI, BCM, OLBC, GS, FMM, PD, DO, SS: Substantial contribution to conception and design, acquisition of data, or analysis and interpretation of data; JBE, MG, PD and SS: Drafting the article or revising it critically for important intellectual content. All authors have read and approved the final manuscript; and agree to be accountable for all aspects of the work thereby ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

#### Funding

Open access funding provided by University of Geneva

## Data availability

No datasets were generated or analysed during the current study.

#### Declarations

#### Ethics approval and consent to participate

For this project, no approval from an ethics committee was required as it exclusively involves the use of artificial intelligence-generated or synthetic data. No human participants, personally identifiable information, or sensitive real-world data were involved in the study. The nature of the data ensures that ethical considerations related to human subject research do not apply.

#### **Consent for publication**

Not applicable.

#### **Competing interests**

SS has received speaker's fees from Medtronic/Merck. JBE is a member of the Board of Directors of the European Society of Anesthesiology and Intensive Care (ESAIC) and has received speaker's fees from Medtronic. Peter Dieckmann is an Associate Editor for Advances in Simulation.

#### Author details

<sup>1</sup>Institute for Medical Education, University of Bern, Bern, Switzerland. <sup>2</sup>RISE-Health, Centre for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal. <sup>3</sup>Division of Anesthesiology, Department of Anesthesiology, Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland. <sup>4</sup>Unit of Development and Research in Medical Education (UDREM), Faculty of Medicine, University of Geneva, Geneva, Switzerland. <sup>5</sup>Division of Anesthesiology, Department of Acute Care Medicine, Geneva University hospitals, Geneva, Switzerland. <sup>6</sup>ETH AI Center, Swiss Federal Institute of Technology Zurich (ETH Zurich), Zurich, Switzerland. <sup>7</sup>Department of Anaesthesia, Pain and Perioperative Medicine, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, Hong Kong. <sup>8</sup>Centro di Simulazione (CeSi), Centro Professionale Sociosanitario Medico-Tecnico, Lugano, Switzerland. <sup>9</sup>Department of Anesthesia and Intensive Care, , Ankara University Faculty of Medicine, Ankara, Turkey. <sup>10</sup>Ankara University Brain Research Center (BAUM), Ankara, Turkey. <sup>11</sup>Department of Anesthesia and Perioperative Care, University of California San Francisco, San Francisco, CA, USA.<sup>12</sup>Anesthesiology Department, Hospitais da Universidade de Coimbra, Coimbra, Portugal.<sup>13</sup>Faculty of Medicine, University of Coimbra, Coimbra, Portugal. <sup>14</sup>Clinical Academic Center of Coimbra, Coimbra, Portugal. <sup>15</sup>Copenhagen Academy for Medical Education and Simulation (CAMES), Herlev, Copenhagen, Capital Region of Denmark, Denmark. <sup>16</sup>Department of Quality and Health Technology, University in Stavanger, Stavanger, Norway. <sup>17</sup>Department of Public Health, Copenhagen University, Copenhagen, Denmark.<sup>18</sup>Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.<sup>19</sup>Department of Anesthesiology, Helora, Mons, Belgium. <sup>20</sup>Department of Surgery, Research Institute for Health Sciences and Technology, University of Mons (UMONS), Mons, Belgium.

## Received: 17 December 2024 Accepted: 22 April 2025 Published online: 04 May 2025

#### References

- Savoldelli GL, Burlacu CL, Lazarovici M, Matos FM, Østergaard D. Integration of simulation-based education in anaesthesiology specialist training. Eur J Anaesthesiol. 2024;41:43–54.
- Berger-Estilita J, Meço BC. Simulation-based learning: basics for anaesthetists. Turk J Anaesthesiol Reanim. 2021;49:194–200.
- Kayes A, Kayes DC, Kolb D. Experiential learning in teams. Simul Gaming. 2005;36:330–54.
- 4. Kong Y. The role of experiential learning on students' motivation and classroom engagement. Front Psychol. 2021;12:771272.
- 5. Woolf SH. How should health care and public health respond to the new US administration? JAMA. 2025;333:1197–8.
- Palaganas JC, Chan AKM, Leighton K. Cultural considerations in debriefing. Simul Healthc. 2021;16:407–13.
- Whitla DK, Orfield G, Silen W, Teperow C, Howard C, Reede J. Educational benefits of diversity in medical school: a survey of students. Acad Med. 2003;78:460–6.
- Corsino L, Fuller AT. Educating for diversity, equity, and inclusion: a review of commonly used educational approaches. J Clin Transl Sci. 2021;5: e169.
- Gomez LE, Bernet P. Diversity improves performance and outcomes. J Natl Med Assoc. 2019;111:383–92.
- Feldman M, Edwards C, Wong A, et al. The role for simulation in professional identity formation in medical students. Simul Healthc. 2022;17:e8-13.
- Global Gender Gap Report 2024. World Economic Forum. Available from: https://www.weforum.org/publications/global-gender-gap-report-2024/ in-full/economic-and-leadership-gaps-constraining-growth-and-skewi ng-transitions-7b05a512cb/. Cited 2024 Nov 24.
- 12. Leroy H, Buengeler C, Veestraeten M, Shemla M, Hoever IJ. Fostering team creativity through team-focused inclusion: the role of leader harvesting the benefits of diversity and cultivating value-in-diversity beliefs. Group Organ Manag. 2022;47:798–839 SAGE Publications Inc.
- Smallheer B, Chidume T, Spinks MKH, Dawkins D, Pestano-Harte M. A scoping review of the priority of diversity, inclusion, and equity in health care simulation. Clin Simul Nurs. 2022;71:41–64.
- Linder I, Weissblueth E. The "scenario" as a key educational tool in the simulation centre. European Journal of Teacher Education Routledge; 0: 1–17.
- Bellman R. An introduction to artificial intelligence : can computers think? San Francisco: Boyd & Fraser Pub. Co.; 1978.
- Gisselbaek M, Suppan M, Minsart L, et al. Representation of intensivists' race/ethnicity, sex, and age by artificial intelligence: a cross-sectional study of two text-to-image models. Crit Care. 2024;28:363.
- Gisselbaek M, Köselerli E, Suppan M, et al. Beyond the stereotypes: artificial intelligence (AI) image generation and diversity in anesthesiology. Front Artif Intell Front. 2024;7. Available from: https://www.frontiersin. org/journals/artificial-intelligence/articles/10.3389/frai.2024.1462819/full. Cited 2024 Oct 9.
- Gisselbaek M, Köselerli E, Suppan M, et al. Gender bias in images of anaesthesiologists generated by artificial intelligence. Br J Anaesth. 2024;133:692–5 Elsevier.
- Ali R, Tang OY, Connolly ID, et al. Demographic representation in 3 leading artificial intelligence text-to-image generators. JAMA Surg. 2024;159:87–95.
- Lee SW, Morcos M, Lee DW, Young J. Demographic representation of generative artificial intelligence images of physicians. JAMA Netw Open. 2024;7: e2425993.
- 21. Gisselbaek, et al. Gender disparities in Al-generated images of hospital leadership in the United States, mayo clinic proceedings: digital health. https://www.sciencedirect.com/science/article/pii/S2949761225000252.
- 22. Value gender and equity in the global health workforce. Available from: https://www.who.int/activities/value-gender-and-equity-in-the-globalhealth-workforce. Cited 2024 Oct 17.
- 23. Nadir N, Winfield A, Bentley S, et al. Simulation for diversity, equity and inclusion in emergency medicine residency training: a qualitative study. AEM Educ Train. 2023;7:S78-87.
- Mutch J, Golden S, Purdy E, Chang CHX, Oliver N, Tallentire VR. Equity, diversity and inclusion in simulation-based education: constructing a developmental framework for medical educators. Adv Simul. 2024;9:20.
- Pampana L. The importance of diversity, equity, and inclusion in medical education. SIMZINE. 2024. Available from: https://simzine.news/exper

ience-en/did-you-know-en/the-importance-of-diversity-equity-and-inclu sion-in-medical-education/. Cited 2024 Dec 15.

- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310:2191–4.
- General Data Protection Regulation. Available from: https://www.surve ymonkey.com/curiosity/surveymonkey-committed-to-gdpr-compl iance/. Cited 2023 Jun 16.
- von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344–9.
- 29. Tao Y, Viberg O, Baker RS, Kizilcec RF. Cultural bias and cultural alignment of large language models. PNAS Nexus. 2024;3:pgae346.
- 30. Using large language models and Al to bridge linguistic differences | Appen. Available from: https://www.appen.com/blog/pulse-of-langu age-evolution. Cited 2024 Dec 6.
- Prabhakaran V, Qadri R, Hutchinson B. Cultural incongruencies in artificial intelligence. arXiv; 2022. Available from: http://arxiv.org/abs/2211.13069. Cited 2024 Dec 8.
- Gender-inclusive language guidelines. ECAS. Available from: https://ecas. org/publication/gender-inclusive-language-guidelines/. Cited 2024 Dec 6.
- Moorthy K, Munz Y, Forrest D, et al. Surgical crisis management skills training and assessment. Ann Surg. 2006;244:139–47.
- 34. Gaba DM, Howard SK, Fish KJ, Smith BE, Sowb YA. Simulation-based training in anesthesia crisis resource management (ACRM): a decade of experience. Simul Gaming. 2001;32:175–93 SAGE Publications Inc.
- Owen H. Simulation in healthcare education: an extensive history. Cham: Springer International Publishing; 2016. Available from: https://link.sprin ger.com/10.1007/978-3-319-26577-3. Cited 2024 Dec 15.
- Piot M-A, Attoe C, Billon G, Cross S, Rethans J-J, Falissard B. Simulation training in psychiatry for medical education: a review. Front Psychiatry. 2021;12: 658967.
- Kim E, Song S, Kim S. Development of pediatric simulation-based education – a systematic review. BMC Nurs. 2023;22:291.
- Wittmann-Price RA, Chabalowski BD. Leadership in simulation. Springer Publishing Company; 2023. Available from: https://connect.springerpub. com/content/book/978-0-8261-6991-4/part/part02/chapter/ch04. Cited 2024 Dec 6.
- Bearman M, O'Brien R, Anthony A, et al. Learning surgical communication, leadership and teamwork through simulation. J Surg Educ. 2012;69:201–7.
- 40. Huehn SL. Utilizing simulation to address structural racism in the healthcare system. Creat Nurs. 2023;29:354–9.
- Barra FL, Costa A, Rodella G, Semeraro F, Carenzo L. Shaping the future of simulator interactions: the role of ChatGPT's advanced voice mode. Resuscitation 2024; 110452.
- Nicolau A, Berger-Estilita J, van Meurs WL, Lopes V, Lazarovici M, Granja C. Healthcare simulation-past, present, and future. Porto Biomed J. 2024;9:270.
- Daugherty PR, Wilson HJ, Chowdhury R. Using artificial intelligence to promote diversity. MIT SMR 2018; Available from: https://sloanreview.mit. edu/article/using-artificial-intelligence-to-promote-diversity/. Cited 2024 Nov 24.
- 44. Equity, diversity and belonging in medical education. American Medical Association. 2024. Available from: https://www.ama-assn.org/education/ changemeded-initiative/equity-diversity-and-belonging-medical-educa tion. Cited 2024 Dec 6.
- Article 4: Al literacy | EU Artificial Intelligence Act. Available from: https:// artificialintelligenceact.eu/article/4/. Cited 2024 Dec 6.
- Almatrafi O, Johri A, Lee H. A systematic review of Al literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). Comput Educ Open. 2024;6: 100173.
- 47. Yu H, Zhou J, Li L, et al. AlPatient: simulating patients with EHRs and LLM powered agentic workflow. 2024.
- Wei H, Qiu J, Yu H, Yuan W. Medco: Medical education copilots based on a multi-agent framework. arXiv preprint arXiv:2408.12496.
- Watson N, Hessami A, Fassihi F, et al. Guidelines for agentic Al safety volume 1: Agentic Al Safety Experts Focus Group - Sept. 2024. Universal Ethics Community of Practice Working Group; 2024. Available from: https://www.linkedin.com/groups/12966081/.

50. Santry HP, Wren SM. The role of unconscious bias in surgical safety and outcomes. Surg Clin North Am. 2012;92:137–51.

# **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.