ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Using artificial intelligence to create diverse and inclusive medical case vignettes for education

This is the submitted version (pre peer-review, preprint) of the following publication:

Published Version:
Bakkum, M.J., Hartjes, M.G., Piët, J.D., Donker, E.M., Likic, R., Sanz, E., et al. (2024). Using artificial intelligence to create diverse and inclusive medical case vignettes for education. BJCP. BRITISH JOURNAL OF CLINICAL PHARMACOLOGY, 90(3), 640-648 [10.1111/bcp.15977].

Availability:
This version is available at: https://hdl.handle.net/11585/958166 since: 2024-02-15

Published:

DOI: http://doi.org/10.1111/bcp.15977

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

14 November 2026

Received: 18 August 2023 | Revised: 27 October 2023 | Accepted: 28 October 2023

DOI: 10.1111/bcp.15977

THEMED ISSUE ARTICLE

Using artificial intelligence to create diverse and inclusive medical case vignettes for education

Michiel J. Bakkum[1,2] | Mariëlle G. Hartjes[1,2] | Joost D. Piët[1,2] | Erik M. Donker[1,2] | Robert Likic[3,4] | Emilio Sanz[3,5] | Fabrizio de Ponti[3,6] | Petra Verdonk[3,7] | Milan C. Richir[1,2] | Michiel A. van Agtmael[1,2,3] | Jelle Tichelaar[1,2,3]

[1]Department of Internal Medicine, Section Pharmacotherapy, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, HV, The Netherlands

[2]Research and Expertise Centre in Pharmacotherapy Education (RECIPE), Amsterdam, HV, The Netherlands

[3]European Association for Clinical Pharmacology and Therapeutics (EACPT) Education Working Group, Amsterdam, The Netherlands

[4]Unit of Clinical Pharmacology, University of Zagreb School of Medicine and Clinical Hospital Centre Zagreb, Zagreb, Croatia

[5]Universidad de La Laguna, school of Health Sciences, Tenerife, Spain and Hospital Universitario de Canarias. La Laguna, Tenerife, Spain

[6]Department of Medical and Surgical Sciences, Pharmacology Unit, Alma Mater Studiorum, University of Bologna, Bologna, Italy

[7]Department of Ethics, Law & Humanities, APH research institute, Amsterdam UMC-VU University, Amsterdam, The Netherlands

Correspondence
Michiel J. Bakkum, Department of Internal Medicine, Section Pharmacotherapy Amsterdam UMC, Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands.
Email: m.bakkum@amsterdamumc.nl

Funding information
European Union under Erasmus+ grant, Grant/Award Number: 2020-1-NL01-KA203-083098

Abstract

Aims: Medical case vignettes play a crucial role in medical education, yet they often fail to authentically represent diverse patients. Moreover, these vignettes tend to oversimplify the complex relationship between patient characteristics and medical conditions, leading to biased and potentially harmful perspectives among students. Displaying aspects of patient diversity, such as ethnicity, in written cases proves challenging. Additionally, creating these cases places a significant burden on teachers in terms of labour and time. Our objective is to explore the potential of artificial intelligence (AI)-assisted computer-generated clinical cases to expedite case creation and enhance diversity, along with AI-generated patient photographs for more lifelike portrayal.

Methods: In this study, we employed ChatGPT (OpenAI, GPT 3.5) to develop diverse and inclusive medical case vignettes. We evaluated various approaches and identified a set of eight consecutive prompts that can be readily customized to accommodate local contexts and specific assignments. To enhance visual representation, we utilized Adobe Firefly beta for image generation.

Results: Using the described prompts, we consistently generated cases for various assignments, producing sets of 30 cases at a time. We ensured the inclusion of mandatory checks and formatting, completing the process within approximately 60 min per set.

Conclusions: Our approach significantly accelerated case creation and improved diversity, although prioritizing maximum diversity compromised representativeness to some extent. While the optimized prompts are easily reusable, the process itself demands computer skills not all educators possess. To address this, we aim to share all created patients as open educational resources, empowering educators to create cases independently.

The authors confirm that the principal investigator for this paper is Michiel J. Bakkum. Clinical responsibility is not applicable to this research.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2023 The Authors. *British Journal of Clinical Pharmacology* published by John Wiley & Sons Ltd on behalf of British Pharmacological Society.

KEYWORDS
artificial intelligence, ChatGPT, diversity and inclusivity

1 | INTRODUCTION

Medical case vignettes are widely used for educational purposes, but they often fall short in representing the diverse patient population encountered in clinical practice.[1,2] The process of clinical reasoning begins the moment healthcare professionals access a patient's electronic health record. The patient's name may lead us to make assumptions about their ethnicity, while their prior medical history provides information about their condition and healthcare utilization. Even during the short walk back from the waiting room, we may form subconscious assumptions about the patient's gender identity, mobility, health literacy and more. In the best scenarios, these assumptions can enhance patient care by allowing personalized medical approaches. However, in less favourable situations, they can contribute to implicit provider bias.[3] To accurately reflect this reality and ensure that students gain experience with and an awareness of these complexities, it is vital that simulated case scenarios incorporate patients who mirror the diversity commonly encountered in real clinical settings. This includes considering factors such as gender, sexual orientation, race (viewed as a societal construct, determined by the patient's self-identification) ethnicity, disabilities and other aspects that shape patient identity (social identities). However, it is important to acknowledge that patients rarely fit neatly into predefined categories and conveying the complexity of patient characteristics within text-based cases can be challenging.

Comparable to medicine in general, medical case vignettes often oversimplify the relationship between patients' social identities, such as race or sexual orientation, and their corresponding diseases.[4] Instances vary from relatively benign scenarios, such as a forester visiting a healthcare provider's office, leading students to consider tick-borne illnesses, to more harmful situations where HIV cases are disproportionately associated with men who have sex with men. While it is important to educate students about relative risk factors, these oversimplified and stereotypical connections can inadvertently reinforce essentialist biases and foster the mistaken belief that these characteristics have a direct one-to-one association with disease.[5,6] Such an approach can have detrimental effects on real patient care, resulting in delayed diagnosis and suboptimal treatment.[7] Additionally, the persistence of race-based medicine, which considers race as an essential factor in diagnostic algorithms and clinical practice guidelines, remains prevalent. In our previous publication, a clinical pharmacology and therapeutics teachers' guide, we provided insights into effectively teaching about such guidelines without oversimplifying the relationship between race and illness.[8] Awareness of these biases is crucial during the creation of cases. However, attempting to eliminate them may inadvertently introduce new biases. Achieving the delicate balance between promoting diversity and mitigating stereotyping proves challenging, as human creators are inherently constrained in their capacity to develop truly diverse cases.

The process of manually creating high-quality cases, even without attention for diversity, is a labour-intensive task. This is particularly true for (objective structured) clinical examinations, where a substantial number of cases must be generated to prevent fraudulent practices. Policies that require publishing these cases after their initial use may restrict their reusability in subsequent years. During a study focused on determining the specific needs of clinical pharmacology and therapeutics teachers in terms of open educational resources (OERs), it was observed that a repository of reusable clinical case scenarios was the most frequently desired resource on the European Open Platform for Prescribing Education (EurOP^{2}E; http://prescribingeducation.eu).[9]

Introducing computer-generated case generation with randomized patient characteristics holds promise in theoretically enhancing diversity while simultaneously simplifying the case-creation process. However, our previous attempts at computer-generated case generation highlighted certain challenges. Completely randomized cases often led to the emergence of implausible scenarios, such as ethnically inappropriate names and biologically impossible combinations of diversity (eg, a 70-year-old individual being pregnant). While introducing extreme forms of diversity can be beneficial in encouraging students to question preconceived notions, we recognized the need to strike a balance that avoids the creation of scenarios that would undermine credibility or risk unintended ridicule.

The emergence of user-friendly artificial intelligence (AI), such as ChatGPT, has presented an opportunity to explore whether it can address the issues of limited diversity and stereotype-driven content found in human-generated cases, as well as the limitations of completely random computer-generated scenarios.[10] This innovative approach aims to utilize an AI-powered chatbot (OpenAI ChatGPT) to randomize patient characteristics in a manner that ensures broad inclusivity and maintains ethnical and biological appropriateness. Furthermore, it seeks to convey a significant portion of this information without relying on predefined categories, while preserving nuance similar to real-life clinical scenarios. This is achieved through the integration of AI image generation techniques (Adobe Firefly). The primary objective of this approach is to create a significant quantity of authentic and diverse cases that can be reused and customized as OERs within local educational contexts. Furthermore, it strives to enhance the accessibility of case creation using AI for all medical educators, particularly those teaching clinical pharmacology and therapeutics who are connected via EurOP^{2}E.

2 | METHODS

To assess the feasibility of this approach, we conducted initial tests to determine if ChatGPT could generate medical cases. Initially, we

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

prompted ChatGPT to create complete medical cases using a single prompt. While ChatGPT demonstrated the ability to generate cases, the results lacked diversity and customizing the output proved challenging. We observed that longer prompts often led ChatGPT to disregard certain parts of our instructions. Based on these findings, we determined that generating cases through small, incremental portions of patient characteristics at a time yielded more favourable outcomes.

We utilized ChatGPT to generate a table of fictive patients by prompting it to first generate their names and ethnicities (to reflect the diversity of Amsterdam). Subsequently, we added additional columns, one at a time, for characteristics such as age, body mass index (BMI), lifestyle and medicine use. This method yielded improved results as it allowed us to introduce more diversity and provide specific limitations in the prompts. For instance, we discovered that not specifying the desired age range led ChatGPT to predominantly create relatively young patients. However, this was not ideal when aiming to create a hypertension case where a broader age range was needed. We encountered two challenges when employing this approach. First, regenerating the tables after each prompt became time-consuming due to their increasing size, resulting in longer waiting times, and occasionally caused ChatGPT to lose track of the original prompt. Second, combining prompts in this manner occasionally led to the creation of stereotypical combinations of diversity. Although these examples were relatively harmless, our intention was to counteract stereotypes rather than reinforce them. Notably, we observed associations such as the occupation 'cook' being linked to a high BMI and French ethnicity being associated with the occupation 'wine connoisseur'.

To mitigate stereotypical associations, we next aimed to randomize patient characteristics independently. However, it was not feasible to randomize all characteristics independently, as certain attributes were inherently interconnected. For instance, ensuring ethnically and gender-appropriate names required simultaneous consideration of the name, gender and ethnicity. Nonetheless, characteristics such as BMI, lifestyle and occupation could be randomized independently. To systematically organize the desired characteristics and their interrelationships, we initially created an overview (Figure 1) as a reference. This overview served as a guide to ensure comprehensive coverage of relevant patient attributes and facilitated the identification of interconnected characteristics that needed to be randomized together. Not all patient attributes were intended to be directly displayed as text within the case scenarios due to the risk of oversimplification and the loss of nuanced information. To address this, we employed the concept of 'dummies' for certain attributes, which served as underlying variables influencing other characteristics. For instance, ethnicity was treated as a dummy variable that influenced the selection of appropriate names and the AI image.

In the end, we were able to generate all relevant patient information in eight short ChatGPT prompts (Table 1). These prompts can be further tailored and expanded to suit local contexts and the specific focus of the assignment. For instance, if the aim is to simulate a scenario in a different city or country, the requirement to match the ethnicities of the fictional individuals with that of Amsterdam can be substituted accordingly. Moreover, when exploring a case study on type 2 diabetes mellitus, adjustments might be needed to accommodate different age ranges and weight distributions, and for a case study on urinary tract infections, the addition of a randomization prompt for urinalyses may be beneficial. The patient presentation and physical examination findings can be randomized as desired, and the manner in which they present their symptoms can be customized, as was previously demonstrated by Benoit.[11] However, for the purpose of this example, we deliberately ensured that each patient was designed to exhibit a highly similar manner of presentation. This placed emphasis on the diverse patient characteristics, allowing learners to focus on the intricacies and complexities of individual patient profiles. The process of generating AI images involves utilizing the eight prompts to generate a textual description of the patient. This description is then copied and pasted into Adobe Firefly, which produces four photographs based on the provided details. From these generated images, the most appropriate one can be manually selected for use.

Through our observations, we have found that generating 30 cases strikes a balance between diversity and loading times. Creating too few cases may lead to an overrepresentation of certain ethnicities and gender identities, while generating too many can unnecessarily increase loading time and occasionally lead to malfunctions. To capture the outputs, we recommend making ChatGPT output them as .csv files (as explained in the footnote to Table 1) and copying the responses into one spreadsheet. While transforming this spreadsheet into nicely formatted cases can be done manually, we suggest exploring the mail merge option in MS Word as a way to automate the process. By utilizing mail merge, data from the spreadsheet can be seamlessly merged into predefined templates, resulting in time and effort savings by eliminating the need for individual case formatting.

While the authors have access to a ChatGPT Plus subscription, which proved helpful during peak usage times for nonpaying users, for the sake of repeatability, all prompts were conducted using GPT-3.5, the freely accessible version of OpenAI's ChatGPT (https://chat.openai.com/). All images were generated using Adobe Firefly's (beta) free text to image generator (https://firefly.adobe.com/).

3 | RESULTS

Using this approach, we have successfully expedited the process of creating cases, while taking steps to improve the representation of diversity within these cases. Figure 2 presents four case vignettes about cellulitis and Supporting Information Data S1 shows the 30 unedited vignettes generated for urinary tract infections. The optimization of prompts required considerable time and effort, but having completed this process, the generation of 30 cases, including formatting, manual checks for appropriateness and other necessary tasks, could be accomplished within approximately 1 h.

Our objective was to implement a trial using these cases in our educational setting, where students will evaluate and provide

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

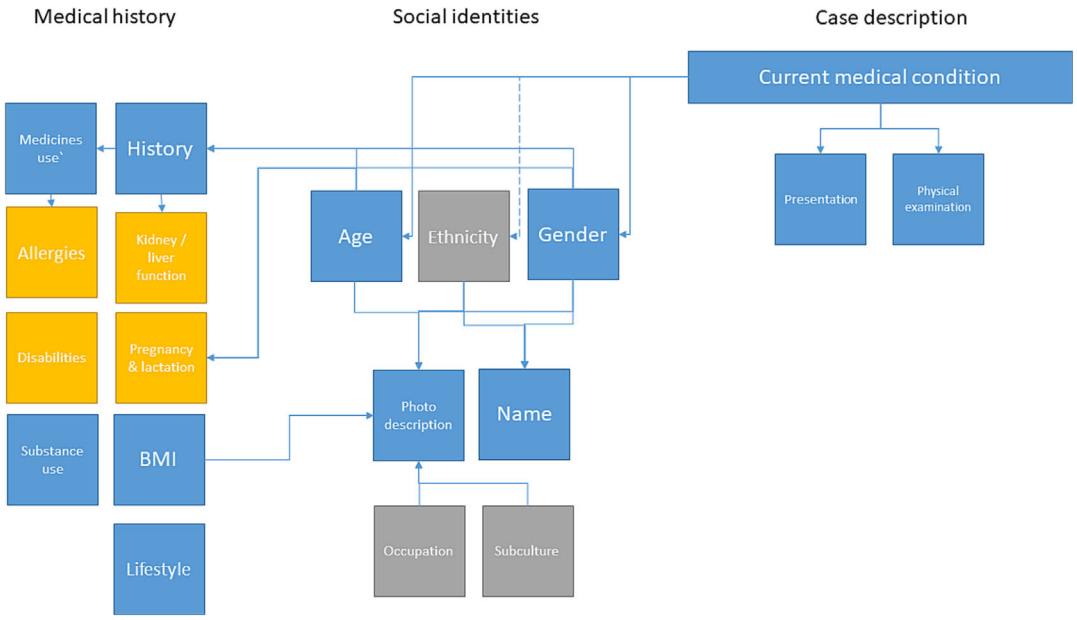

FIGURE 1 Overview of randomized characteristics and their interconnections. The patient characteristics generated by ChatGPT: blue, randomized characteristics displayed as text; grey, dummy variables used for influencing other characteristics; yellow, characteristics to be combined in medical alerts and disabilities.

feedback on them. This feedback will be utilized to further refine and enhance the case-generation process. We envision a lesson where each student is presented with a patient who presents in a similar manner, but with patients themselves demonstrating substantial diversity. Through team-based learning,[12] students will engage in discussions to identify and analyse the distinct factors that differentiate the patients and explore how these variations impact their diagnostic and treatment approaches. This interactive approach fosters critical thinking and promotes a deeper understanding of the influence of diversity on healthcare decision-making.

4 | DISCUSSION

Our approach has successfully achieved its goal of expediting the case-creation process, enhancing diversity through computer randomization of patient characteristics. Nevertheless, it is important to acknowledge a trade-off when it comes to representativeness. In the real world, patients do not possess a randomized set of characteristics; instead, they have multiple social identities that intersect and mutually shape one another (known as intersectionality).[13,14] While we believe ChatGPT to be quite capable of giving ethnically appropriate names, this was not verified through name databases. Moreover, it insufficiently takes other intersections into account. For instance, ChatGPT generated a case where a 61-year-old Moroccan patient had an Arabic first name and a Dutch last name (Mohamed van der Berg). While this combination is not impossible, it is uncommon, as most 61-year-old Moroccan males living in the Netherlands are first-generation migrants, and such name combinations are more likely in later generations. Another example is the case vignette of Fatima Rahman in Figure 2, who consumes an atypical amount of alcohol (1.8 units per day) considering her pregnancy. Our current approach aimed to minimize stereotypes, prioritizing diversity over representativeness. Whether this balance is optimal is a topic that should be explored in future research and subject to debate.

The same is true for the choice of what characteristics to include in the vignettes and how to display them. By incorporating a wide range of patient characteristics, students gain insights into the factors that influence their clinical decision-making, distinguishing between relevant and irrelevant information across different clinical scenarios. Case vignettes should avoid providing clues about relevance by selectively including or excluding certain details. Currently, dimensions of diversity such as religion, political beliefs and sexual preference were not encompassed in the cases. This decision was made to focus on the most commonly relevant factors in medical scenarios and to align with real-life patient care practices. For instance, parameters like BMI and smoking history hold relevance across various medical scenarios and are typically documented in medical records. Conversely, religion seldom holds relevance nor is it a routinely recorded parameter in such contexts. Ethnicity can be significant in relation to disease prevalence and therapeutic response.[15] However, capturing ethnicity as a single categorical label poses challenges due to its multifaceted nature, encompassing concepts such as ancestry, nationality and culture.[16,17] Moreover, in real clinical settings, ethnicity is seldom objectified but rather inferred from a patient's phenotype, name and cultural expressions.[18,19] To reflect this reality and encourage students to recognize and challenge their assumptions, we made the decision to convey information about ethnicity (as well as [sub]culture and occupation) through a photograph of the patient, rather than a label. Utilizing AI image generation to create lifelike photographs of

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 ChatGPT prompts to generate all relevant patient information for case vignettes.

Characteristic	Prompt	Notes
Name Gender Ethnicity	Make 30 fictive people for use in medical case vignettes. Make their ethnicity reflect the diversity of Amsterdam. Make four people have a trans <use transman or transwoman> or nonbinary gender. Output in table with columns first name, last name, gender, ethnicity.	Limiting ethnic diversity to reflect Amsterdam helped create a more representative population. By specifying gender diversity, the 30 cases included a diverse range of gender identities. Without this specification, there would be fewer nonbinary patients and no mention of transgender individuals.
Age Date of birth Occupation Subculture	For use in medical case vignettes randomize 30 birthdates, so that the fictive patients are between 18 and 98 years old. Also randomize their occupation. Be inclusive of pensioners and jobless people. Also randomize their subculture. Be inclusive of all possibilities, including none. Output as table with columns age, date of birth (in dd-mm-yyyy), occupation and subculture.	Specifying the age limit ensured a more accurate representation of disease prevalence (in this example, infectious diseases). Without this, ChatGPT primarily generated responses related to individuals aged 20-60. Both occupation and subculture served as dummy variables in the photograph, showcasing relatively diverse examples such as student, bus driver, barista, graphic designer, as well as punk, gamer and sports fan subcultures.
BMI Length Weight Alcohol intake Tobacco use Drug use Lifestyle	For use in medical case vignettes randomize 30 BMIs to reflect a diverse population of individuals. Be inclusive of thin, healthy weight, obese and morbidly obese individuals. Also randomize lifestyle, alcohol (in average units/day), tobacco use (as either active, with number of packs/per day; stopped with the number of pack-years smoked; or never) and recreational drug use (with explanation of what drugs and the frequency). Output as table with columns length (in metres), weight (in kilograms) and BMI (one decimal), lifestyle, alcohol intake, tobacco use and recreational drug use.	Specifying body metrics is crucial for representing diversity accurately. Without it, ChatGPT tends to generate responses predominantly depicting individuals with a healthy weight. However, it can be further customized by prompting to make it typical for patients with a specific condition, such as diabetes type II.
Medical history	For use in medical case vignettes randomize medical history for the following fictive patient's age and gender: <paste list of patient age and gender> Make it include between two and four of the following chronic conditions. Be inclusive of all options. Pain <replace with specific example, like ankle fracture, osteoarthritis and more>, hypertension, diabetes type 2, contraceptive use, acne, eczema, allergies <replace with specific example like allergic rhinitis, conjunctivitis>, hypothyroidism, acid reflux, gastric ulcer disease, atrial fibrillation, ischemic heart disease, heart failure, depression Also include between two and four of the following minor ailments. Be inclusive of all options. Headache, common cold, sore throat, cough, fever, nausea, upset stomach, diarrhoea, heartburn, indigestion, muscle ache, backache, minor cuts <use example like cut finger>, bruises, sprains, minor burns <use example like burned hand> etc.* Avoid conditions that are untypical for age and gender. In total the lists must be between three and seven conditions long. Randomize the order in the format <year of onset>—<condition>. Output as a table with a new column, with linebreaks between the conditions.	Addressing medical history can be challenging since it intertwines with various biological factors. However, we discovered that providing comprehensive lists of potential conditions is necessary. Without such specificity, ChatGPT tends to rely on a limited range of options. Differentiating between chronic and minor illnesses is vital for accuracy and relevance, especially considering the diseases that students have previously learned to diagnose and treat. The 'avoid conditions that are untypical for age and gender' prompt feature may occasionally experience malfunctions, therefore we highly recommend manually reviewing the lists to ensure accuracy and appropriateness. *Table shortened for presentation; actual list of minor conditions is longer.
Current medicine use	For use in medical case vignettes. Randomize the current medicine lists for patients with the following conditions. Output as a table, one fictive patient per row as below. indicate linebreaks within a cell. <paste list of medical histories> Use generic drug names, not brand names, nor group names. Use only drugs available in the Netherlands. Minor conditions <2021 must not be treated.	Creating accurate medical lists can be challenging for ChatGPT. While the prompt to use generic names and output in the given format is effective, it may not account for local availability of medicines, contraindications and drug-drug interactions. Due to these limitations, it is crucial to manually review and verify the lists for accuracy.

(Continues)

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Characteristic	Prompt	Notes
	Avoid clinically relevant drug-drug interactions and medicine contraindicated by comorbidty. Use format: drug name, dosage, daily use (eg, lisinopril, 10 mg, 1 tablet per day) Output as a table, with a new column with linebreaks between the medicines.	
Pregnancy Lactation Allergies	For use in medical case vignettes, randomize the allergies, pregnancy and lactations status for fictive patients with the following ages and genders. <paste list of patient ages and genders> 40% of female, transman (not transwoman) or nonbinary patients aged 18-46 should be pregnant or lactating (not both). Write pregnancies including duration in weeks + days notations (eg, pregnant 34 + 5). Write lactating as 'lactating'. Leave empty if not applicable. 25% of all patients should have one or more allergies to medicinal products (specific drugs, latex, bandages etc). Make it diverse. Other allergies should not be listed (leave empty). The most common medicinal allergies are for penicillin, Non-steroidal anti-inflammatory drugs and Angiotensin-converting enzyme inhibitors, but use others too. Output as table with a new columns for allergies and pregnancy/lactation.	Randomizing medical contraindications poses challenges, particularly when it comes to specifying pregnancy and lactation. ChatGPT may not fully comprehend transgender pregnancies or reproductive age limits, therefore it becomes necessary to explicitly define who can be pregnant and lactate to ensure accurate and inclusive information. The pregnancy/lactation rate was set higher to offer students ample training opportunities. However, it can be adjusted for a more representative experience in line with real-world scenarios. Specifying allergies is essential to include only medically relevant ones. We extensively experimented with combining the prompts so that ChatGPT avoids allergic reactions to medications already present in the current medications list, but unfortunately we were unable to make this feature work reliably. As a result, manual correction is necessary to ensure accuracy in handling allergies and medications.
Disability Renal insufficiency	For use in medical case vignettes randomize fictive patients. <paste list of patient ages> Make one in 10 of these people have a disability, consider prevalence in relation to age. Leave empty if not applicable. Pick from visual impairment, hearing loss, mobility impairment, cognitive impairment, autism spectrum disorder, speech and language disorder, developmental delay, attention deficit hyperactivity disorder, learning disability, intellectual disability, dyslexia, down syndrome, cerebral palsy, amputee (right foot). Make one in 10 of these people have a renal insufficiency warning. Consider prevalence in relation to age. Write as renal insufficiency (eGFR <randomly generated 20-60>). Leave empty if not applicable. Output as a table with new columns for disability and kidney insufficiency.	Randomizing disability, along with the last medical contraindication of renal insufficiency, can be treated as separate characteristics during the generation process. However, combining these factors with the previously generated ages may enhance the ability to generate more age-appropriate results.
AI image prompt	Provide a visual description for each of the following fictive patients: Use format: A <age> year old <gender description> who is <BMI description>, <pregnancy description>, <disability description>, <occupation> <subculture description> <paste table with aforementioned parameters> Output as new column in the table	To generate four images, copy these prompts into Adobe Firefly and manually select the best one. For optimal results, we suggest using 'photo' settings rather than 'art' or 'graphic' settings to ensure a more realistic representation and avoid caricature-like images.

Note: To capture the outputs effectively, we propose employing a follow-up prompt: "Output as .csv with semicolon separators, preserving line breaks." This prompt will generate text that can be conveniently copied and saved as a .csv file using software like Notepad or pasted into a spreadsheet program, aided by the text-to-column wizard. Given that csv files do not support line breaks, it is crucial to substitute them with a concise code, such as in this case, which can be subsequently replaced with an actual line break using the find and replace function in your preferred text processing software.
Abbreviations: BMI, body mass index, eGFR, estimated glomerular filtration rate.

nonexistent individuals offers several advantages over using real photographs, eliminating the need for extensive searches for suitable images that match the fictional patient's profile while addressing privacy, and portrait and intellectual property law concerns. Images generated through Adobe Firefly can be freely reused for noncommercial purposes. In situations where a sexual history is relevant, such as in cases related to sexually transmitted diseases or potential pregnancy, we recognize that this cannot be simply captured through a

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

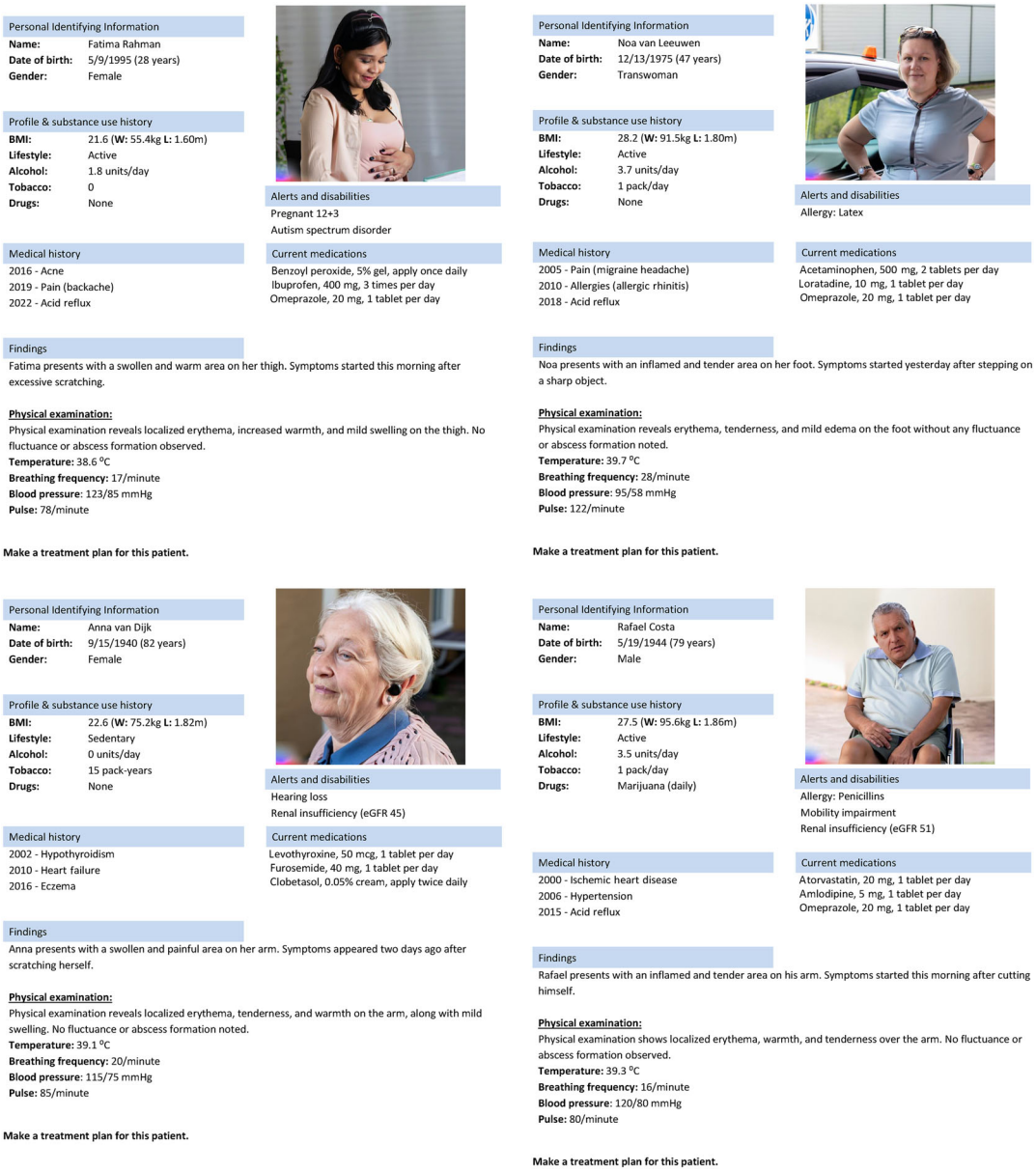

FIGURE 2 Examples of case descriptions. Four examples of cases generated for a case-based clinical pharmacology and therapeutics assignment about cellulitis.

single categorical label (eg, men having sex with men). Instead, it necessitates a comprehensive explanation within the patient presentation. This detailed information can also be generated by AI. If this information is included in a case, it is essential to approach it without normativity by ensuring equal detail in the display of information.[2,20] For example, the sexual history of patients with one primary sex partner should be described as thoroughly as that of patients with multiple sex partners, and the details of heterosexual relations should be presented with the same level of depth as that of bisexual and homosexual relations.

ChatGPT functions by accurately predicting word order, enabling it to generate responses that may appear knowledgeable. However, it is crucial to remember that ChatGPT does not possess actual medical, biological or social knowledge. Furthermore, it is essential to emphasize that ChatGPT was trained using a vast collection of texts sourced from the internet, including webpages, books and texts mostly from the Global North. Although the exact details of the training data set have not been publicly disclosed, it is improbable for the materials to be completely free from biases, therefore it is essential to acknowledge that ChatGPT inherently reflects the same biases that we are already familiar with, such as a tendency towards favouring young and white individuals.[21,22] Consequently, certain limitations arise when employing it for the development of medical cases, such as BMIs not being calculated accurately from length and weight, genders and ethnicities not generating according to the requested distribution, patients having allergies to currently prescribed medications and so

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

forth. Some of these limitations can be mitigated by including explicit specifications in the prompts or regenerating the response several times until it has the correct format and level of diversity. However, addressing other limitations necessitates manual refinement. To ensure precision and consistency, we highly recommend conducting comprehensive manual reviews of all cases to identify any potential inconsistencies.

While our optimized prompts have significantly facilitated the process of creating cases, we acknowledge that making these cases with ChatGPT still requires a certain level of computer skills. For instance, saving ChatGPT outputs to a .csv file and creating a template for mail merge in MS Word can be challenging. Exploring the possibilities of utilizing the ChatGPT API code to streamline the process is an avenue worth considering, although it may require expertise beyond our current capabilities. As an alternative approach, we intend to make the diverse fictional individuals we generate accessible on EurOP^{2}E as OERs. This will allow teachers to access the patient information and manually input the cases for various medical subjects themselves. Moreover, we have plans to develop a comprehensive guide on crafting these cases detailing the challenges and opportunities, and establish a discussion board where teachers can engage in conversations about the prompts, share their experiences and collectively enhance their quality.

A scientific evaluation of our efforts, including students' reactions and evaluations is not included in the current manuscript. We are committed to conducting such an evaluation and reporting the findings separately.

AUTHOR CONTRIBUTIONS

Michiel J. Bakkum: Conceptualization; formal analysis; writing draft; visualization. **Mariëlle G. Hartjes**, **Joost D. Piët**, **Erik M. Donker**: Conceptualization; formal analysis; writing—review and editing. **Robert Likic**, **Emilio Sanz**, **Fabrizio de Ponti**, **Petra Verdonk**, **Milan C. Richir**, **Michiel A. van Agtmael**: Conceptualization; writing—review and editing. **Jelle Tichelaar**: Conceptualization; writing—review and editing; supervision.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of ChatGPT in copy-editing the text of this article. It is important to note that ChatGPT was solely used as a tool for editing purposes (prompt: improve writing) and no content was generated by ChatGPT. This research was funded by the European Union under Erasmus+ grant 2020-1-NL01-KA203-083098.

CONFLICT OF INTEREST STATEMENT

Author Robert Likic is guest editor for the holidAI special of the *British Journal of Clinical Pharmacology* and should not be involved in any editorial decisions regarding this manuscript. All other authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

No data were gathered for this research.

ORCID

Michiel J. Bakkum https://orcid.org/0000-0003-0928-812X
Mariëlle G. Hartjes https://orcid.org/0000-0002-7595-2570
Joost D. Piët https://orcid.org/0009-0000-7380-4876
Erik M. Donker https://orcid.org/0000-0002-8169-0714
Robert Likic https://orcid.org/0000-0003-1413-4862
Emilio Sanz https://orcid.org/0000-0001-6788-4435
Fabrizio de Ponti https://orcid.org/0000-0002-0367-9595
Petra Verdonk https://orcid.org/0000-0003-0464-8210
Michiel A. van Agtmael https://orcid.org/0000-0002-7966-6934
Jelle Tichelaar https://orcid.org/0000-0002-7485-9219

REFERENCES

1. Lee CR, Gilliland KO, Beck Dallaghan GL, Tolleson-Rinehart S. Race, ethnicity, and gender representation in clinical case vignettes: a 20-year comparison between two institutions. *BMC Med Educ*. 2022; 22(1):585. doi:10.1186/s12909-022-03665-4
2. Muntinga M, Beuken J, Gijs L, Verdonk P. Are all LGBTQI+ patients white and male? Good practices and curriculum gaps in sexual and gender minority health issues in a Dutch medical curriculum. *GMS J Med Educ*. 2020;37(2):Doc22. doi:10.3205/zma001315
3. Oxtoby K. How unconscious bias can discriminate against patients and affect their care. *BMJ*. 2020;371:m4152. doi:10.1136/bmj.m4152
4. Ripp K, Braun L. Race/ethnicity in medical education: an analysis of a question Bank for Step 1 of the United States medical licensing examination. *Teach Learn Med*. 2017;29(2):115-122. doi:10.1080/10401334.2016.1268056
5. Nieblas-Bedolla E, Christophers B, Nkinsi NT, Schumann PD, Stein E. Changing how race is portrayed in medical education: recommendations from medical students. *Acad Med*. 2020;95(12):1802-1806. doi:10.1097/ACM.0000000000003496
6. Tsai J, Ucik L, Baldwin N, Hasslinger C, George P. Race matters? Examining and rethinking race portrayal in preclinical medical education. *Acad Med*. 2016;91(7):916-920. doi:10.1097/ACM.0000000000001232
7. Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. *The Lancet*. 2020;396(10257):1125-1128. doi:10.1016/S0140-6736(20)32076-6
8. Bakkum MJ, Verdonk P, Thomas EG, et al. A clinical pharmacology and therapeutics Teacher's guide to race-based medicine, inclusivity, and diversity. *Clin Pharmacol Ther*. 2023;113(3):600-606. doi:10.1002/cpt.2786
9. Bakkum MJ, Loobeek BJ, Richir MC, et al. Teaching resources for the European open platform for prescribing education (EurOP(2)E)-a nominal group technique study. *NPJ Sci Learn*. 2022;7(1):23. doi:10.1038/s41539-022-00141-y
10. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. doi:10.3390/healthcare11060887
11. Benoit JRA. ChatGPT for Clinical Vignette Generation, Revision, and Evaluation. medRxiv. 2023:2023.02.04.23285478.
12. Parmelee D, Michaelsen LK, Cook S, Hudes PD. Team-based learning: a practical guide: AMEE guide no. 65. *Med Teach*. 2012;34(5):e275-e287. doi:10.3109/0142159X.2012.651179
13. Muntinga ME, Krajenbrink VQ, Peerdeman SM, Croiset G, Verdonk P. Toward diversity-responsive medical education: taking an intersectionality-based approach to a curriculum evaluation. *Adv Health Sci Educ Theory Pract*. 2016;21(3):541-559. doi:10.1007/s10459-015-9650-9

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14. Verdonk P, Muntinga M, Leyerzapf H, Abma T. From Gender Sensitivity to an Intersectionality and Participatory Approach in Health Research and Public Policy in the Netherlands. In: Hankivsky O, Jordan-Zachery JS, eds. *The Palgrave handbook of intersectionality in public policy*. Springer International Publishing; 2019:413-432. doi:10.1007/978-3-319-98473-5_18
15. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight — reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/NEJMms2004740
16. Commission E, Justice D-gf, consumers, Farkas L. Analysis and comparative review of equality data collection practices in the European Union: data collection in the field of ethnicity: publications office, 2020.
17. Routen A, Akbari A, Banerjee A, et al. Strategies to record and use ethnicity information in routine health data. *Nat Med*. 2022;28(7):1338-1342. doi:10.1038/s41591-022-01842-y
18. Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. *Jama*. 2003;289(20):2709-2716. doi:10.1001/jama.289.20.2709
19. Boehmer U, Kressin NR, Berlowitz DR, Christiansen CL, Kazis LE, Jones JA. Self-reported vs administrative race/ethnicity data and study results. *Am J Public Health*. 2002;92(9):1471-1472. doi:10.2105/AJPH.92.9.1471
20. Krishnan A, Rabinowitz M, Ziminsky A, Scott SM, Chretien KC. Addressing race, culture, and structural inequality in medical education: a guide for revising teaching cases. *Acad Med*. 2019;94(4):550-555. doi:10.1097/ACM.0000000000002589
21. Singh S. Is ChatGPT biased? A review. 2023.
22. Stypinska J. AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies. *AI Soc*. 2023;38(2):665-677. doi:10.1007/s00146-022-01553-5

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bakkum MJ, Hartjes MG, Piët JD, et al. Using artificial intelligence to create diverse and inclusive medical case vignettes for education. *Br J Clin Pharmacol*. 2024;1-9. doi:10.1111/bcp.15977

13652125, 0, Downloaded from https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15977 by CochraneItalia, Wiley Online Library on [15/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License