Annika Meyer*, Johannes Ruthard and Thomas Streichert

Dear ChatGPT – can you teach me how to program an app for laboratory medicine?

https://doi.org/10.1515/labmed-2024-0034
Received February 26, 2024; accepted April 18, 2024;
published online May 15, 2024

Abstract

Objectives: The multifaceted potential of ChatGPT in the medical domain remains underexplored, particularly regarding its application in software development by individuals with a medical background but limited information technology expertise.
Methods: This study investigates ChatGPT's utility in creating a laboratory medicine application.
Results: Despite minimal programming skills, the authors successfully developed an automated intra-assay, inter-device precision test for immunophenotyping with a shiny user interface, facilitated by ChatGPT. While the coding process was expedited, meticulous oversight and error correction by the authors were imperative.
Conclusions: These findings highlight the value of large language models such as ChatGPT in code-based application development for automating work processes in a medical context. Particularly noteworthy is the facilitation of these tasks for non-technically trained medical professionals and its potential for digital medical education.

Keywords: ChatGPT; digital skills gap; programming; laboratory medicine

Introduction

Historically, the fields of laboratory medicine and information technology have been closely intertwined [1]. Digital progress has been a driving force behind the digital transformation towards Lab 4.0 [2]. With an increasing focus on digital solutions, the need for expertise extending beyond the core competencies of laboratory medicine becomes increasingly prominent [3].

Thus, according to the literature, future laboratory professionals require not only a fundamental understanding of technology and basic digital skills but also programming abilities [4]. Correspondingly, the Association for Diagnostics and Laboratory Medicine issued to 'Embrace the R Programming Language' as a 'Gateway to Laboratory Medicine's Digital Future' as early as 2020 [5].

However, among the 'digital native' 'Young Scientists' in laboratory medicine, only 23 % program daily in R and a mere 6 % in Python. Despite 96 % of young laboratory medicine scientists recognizing the necessity of digital education, only 20 % receive it. This reflects a disparity between the availability and demand for digital education in this field, with a call for learning resources tailored to various knowledge levels [4]. In this context, ChatGPT's capabilities in code generation, optimization, debugging assistance, code documentation, and review, coupled with its accessibility and ease of use [6], present a potential opportunity for novices to develop programming skills [7, 8]. This case study, to the best of the authors' knowledge, is the first to explore the application of ChatGPT in programming a laboratory medicine application by non-programming-experts, aiming to further bridge the 'digital skills gap' in laboratory medicine [4].

Methods

The problem

At the Institute of Clinical Chemistry at the University Hospital Cologne, the reproducibility of immunophenotyping is regularly assessed. This involves evaluating intra-assay, inter-device precision tests using Pearson correlation of multicolor flow cytometry from 10 patient samples across two different BD FACSCanto™ II. Traditionally, this process involved file sorting and manual data transfer from PDF to Excel, consuming several hours of medical staff time.

In pursuit of process optimization, reducing input errors, and more effective time utilization, the development of an R-based alternative solution was appealing. However, implementing such a solution with rudimentary knowledge of R,

***Corresponding author: Annika Meyer**, Institute of Clinical Chemistry, University Hospital Cologne, Cologne, Germany,
E-mail: annika.meyer1@uk-koeln.de. https://orcid.org/0000-0002-8411-8799
Johannes Ruthard and Thomas Streichert, Institute of Clinical Chemistry, University Hospital Cologne, Cologne, Germany. https://orcid.org/0000-0002-6588-720X (T. Streichert)

Open Access. © 2024 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

focused primarily on data management and statistics, did not seem feasible without external help.

The solution approach

Inspired by the utilization of ChatGPT by experienced programmers, we decided to employ ChatGPT (GPT-4 version) to address this challenge and explore its potential advantages and disadvantages. For this purpose, we initially developed a procedural plan for our experiment. The objective was to write an application in the R programming language, without the involvement of external human experts, that would automate the following steps:

- Reading all file names in a folder and including the names in a list.
- Removal of file names with erroneous measurements from the list.
- Grouping of file names according to the sample tested.
- Grouping of file names based on the measuring instrument.
- Removal of file names with incomplete measurements from the list.
- Extraction of comparative parameters (e.g., CD3+ values measured by the different devices) from the sorted PDF files in the list into juxtaposed tables.
- Calculation of analysis-specific Pearson correlation coefficients and their graphical representation.

In addition, this program was supposed to have a shiny user interface to lower the inhibition threshold for use in everyday laboratory work. The interface was intended to allow the user to choose the file folder, select parameters, and export individual graphs and tables. With this plan in mind, we began our adventure with the relatively unspectacular phrase: 'I have a PDF document and would like to read it into R as text, how can I do that?' (original prompt: ‚Ich hab ein PDF dokument und möchte dieses gerne als text in R einlesen, wie kann ich das machen?'). The R-Code developed in this project is available as Supplementary Material.

Results

Over the course of two weeks, we exchanged 135,615 words with ChatGPT, which facilitated the development of an application that met our intended specifications.

ChatGPT compensated for our lack of skills in areas such as extracting data from PDF files and creating user interfaces (Figure 1). It provided multiple solutions to open-ended questions. Knowledge gaps were bridged by the easy-to-understand code annotations and explanations provided by ChatGPT, concurrently improving our programming skills.

However, implementation without basic knowledge of programming languages and the structure of R would not have been feasible, even with the support of ChatGPT.

ChatGPT frequently produced erroneous code in response to vaguely formulated user input ("prompts"). We often had to experiment with the wording and language, switching from German to English, until ChatGPT correctly interpreted our intentions and translated them into code. Even with an adjusted input formulation, ChatGPT's execution of solution strategies was often inadequate, requiring the development of new strategies for sub-problems within the process. In general, the code snippets written by ChatGPT

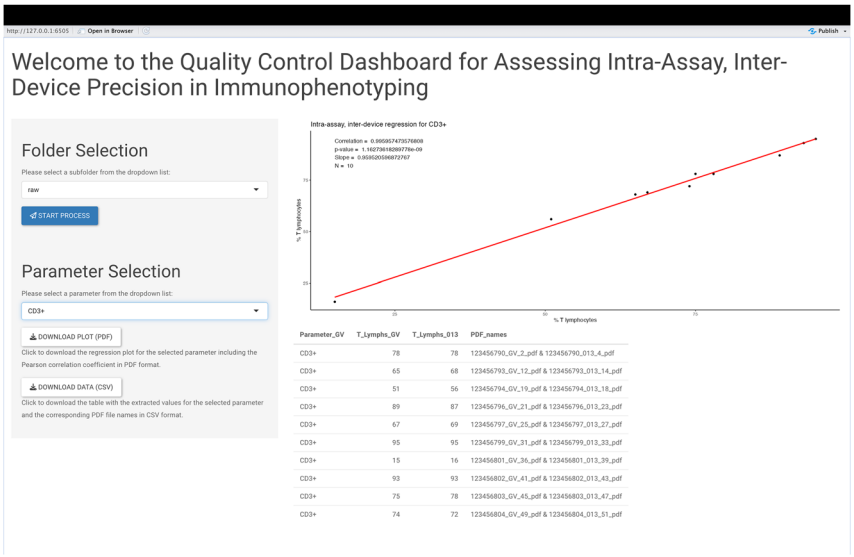

Figure 1: User interface of the developed application for assessing the intra-assay, inter-device precision of immunophenotyping, Cologne 2023. Users can select folders and parameters for analysis through a drop-down menu, download extracted values in CSV format, and obtain inter-assay precision results as a PDF containing a regression plot with a correlation coefficient.

Table 1: Evaluation of ChatGPT's implications on coding in laboratory medicine.

Attributes	Observation	Link reference	Implication
Positive	Code generation by ChatGPT	https://chat.openai.com/share/a6a29e6e-30fa-40bc-a89c-c5d23d1dae26	ChatGPT demonstrates the potential to assist non-experts in generating applications for medical laboratories, streamlining the coding process
	Explained solutions to error messages	https://chat.openai.com/share/93c9bf39-e123-4836-b54b-b59522989fd2	ChatGPT provides actionable solutions for resolving programming errors, potentially reducing downtime in error spotting
	Simplified explanations of code	https://chat.openai.com/share/51675e4e-5555-4d2f-8f91-6ce9fd5a9f1b	ChatGPT contributes to the enhancement of medical professionals' programming capabilities through clear and detailed explanations
	Provision of code annotations	https://chat.openai.com/share/1c13a0ea-9296-4618-b14f-1c2f330668e7	ChatGPT's annotations for code facilitate a deeper understanding of programming constructs, improving code literacy among medical researchers
Negative	Missing reproducibility of ChatGPT output	ChatGPT's first output: https://chat.openai.com/share/14c57cdf-ff99-4ba2-a527-8fe4b23e834d ChatGPT's second output: https://chat.openai.com/share/8d866f54-9ec5-4371-ac0e-b5bb291f311c	ChatGPT's responses are missing reproducibility despite identical user inputs, indicating the need for cautious interpretation of automated code suggestions
	Complex solutions without further examples	https://chat.openai.com/share/377dbaca-39b6-493e-8dc2-94038e7fc8f7	Understanding and implementing the solutions offered by ChatGPT requires basic coding skills, highlighting the importance of basic coding skills for healthcare professionals

often generated error messages, which in most cases could be resolved with the help of ChatGPT and further research into the vignettes (Table 1).

The authors are currently in communication with the IT department of the University Hospital Cologne to enable routine use of the shiny app in the future.

Discussion

Programming expertise is becoming increasingly vital in the daily practice of future medical laboratory professionals. Thus, to bridge the "digital skills gap" in laboratory medicine, the development and research of potential learning resources for individual skill acquisition are essential [4].

In the wider landscape of software development, social media users are already leveraging ChatGPT's capabilities for developing and debugging code across 10 different programming languages [9]. This utility underscores ChatGPT's superior problem-solving capabilities, as in Python, where it outperforms counterparts such as Bard and Claude [10], and demonstrates robustness in addressing tasks of varying complexity in Java and C++ [11].

This research further extends this narrative by demonstrating the suitability of ChatGPT for aiding the development of functional laboratory medicine applications by non-programmers. This is consistent with existing literature on the utilization of ChatGPT for code generation in pharmacometrics [7] and medical statistics [8], where it has enabled users with minimal programming knowledge to write code and develop operational programs [7, 8], indicating its potential applicability across various disciplines.

Furthermore, this study underscores ChatGPT's cross-linguistic ability to generate code, a capability further emphasized by the diverse linguistic focus in prior studies [7, 8]. However, the observations indicate a preference for English inputs, a tendency probably stemming from the predominance of English in ChatGPT's training dataset [12]. This linguistic inclination is consistent with trends noted in other domains [13], including medical exams [14], suggesting a broader pattern of language bias.

In harmony with findings from multiple studies [7, 8, 10, 11, 15], the necessity of repeated input refinement is critical for deriving accurate and operational code from ChatGPT also within the domain of laboratory medicine. Moreover, observations regarding the inconsistency in ChatGPT's output for identical prompts are reaffirmed, highlighting a deficiency in reproducibility [7]. In addition to concerns regarding reproducibility, reliability, and accuracy, ChatGPT's code-generation capabilities are also not immune to the more general criticisms of this type of Artificial Intelligence, including ethical concerns, over-reliance, and security risks [6]. For instance, over-reliance on ChatGPT's programming capabilities might inhibit critical thinking and hinder the development of individual programming skills [6, 16], thus leading to "deskilling" and "automation-bias" [17].

In turn, inadequate proficiency in code reading and interpretation present a security hazard [6], with the potential for sensitive data leakage [18] and discriminatory algorithms [9] to pass undetected. This concern is especially pronounced in sensitive domains such as healthcare, where the integrity of patient data and network security are critical [19]. Consequently, instead of relying solely on automated code generation, adopting a cautious strategy alongside a basic grasp of programming principles is essential for ensuring safe usage.

Contrary to these concerns, research by Kazemitabaar et al. highlights that ChatGPT does not adversely affect novice programmers' ability to modify or generate code [20]. In fact, it has been shown to improve performance, boost self-efficacy, decrease frustration, and promote skill retention over time [20, 21], thereby positioning ChatGPT as a potential educational aid and companion for novice programmers [16].

Overall, while ChatGPT holds the promise to streamline code development and debugging, as well as enrich educational experiences for learners, its outputs must be rigorously monitored and evaluated, especially within data-sensitive fields like healthcare. Therefore, ChatGPT should only be used as a complementary tool in laboratory medicine by users with basic programming knowledge. Future studies are needed in this context to investigate possible integration into corresponding academic training programs.

Conclusions

Overall, it is evident that ChatGPT is capable of assisting individuals with limited coding expertise in writing laboratory medicine programs using R. However, in light of valid criticisms regarding its accuracy and reliability, as well as concerns pertaining to security, over-reliance, and ethical implications, the outputs generated by ChatGPT should be subjected to rigorous scrutiny.

The easily comprehensible explanations and annotations provided by ChatGPT underscore its potential to support digital education in the field of laboratory medicine. Therefore, future research focusing on the successful integration of ChatGPT into academic digital education programs appears to be a worthwhile endeavor.

Learning points

(1) ChatGPT can aid non-experts in programming medical laboratory applications.
(2) Through clear explanations and detailed connotations, ChatGPT can help develop the programming skills of medical professionals.
(3) ChatGPT's code often generates error messages that can only be partially solved by input reformulations.
(4) Due to justified criticism regarding reproducibility and accuracy as well as ethical and safety concerns, ChatGPT should only be used by trained personnel for programming support.

Acknowledgments: The authors thank Regine Meyer for proofreading the manuscript. DeepL as well as ChatGPT were utilized for linguistic and translation purposes. As described in the article, ChatGPT was also used for the programming part of the application development. All outputs from AI have been critically reviewed by the authors.
Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: AM, JR and TS designed this experiment. AM programmed the application with the assistance of ChatGPT and JR. AM wrote the manuscript. TS and JR critically reviewed the manuscript. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: TS and AM received support by the DFG (German Research Foundation) for article processing charges of other publications. TS received honoraria for lectures by Roche, Siemens and Werfen as well as travel support by the DGKL (German Society for Laboratory Medicine), DGLN (German Society for CSF/Neurology), ADLM (Association for Diagnostics and Laboratory Medicine) as well as ICFF (International Federation of Clinical Chemistry).
Research funding: The research was supported by the Institute for Clinical Chemistry at the University Hospital of Cologne.
Data availability: The underlying code is enclosed in the appendix.

References

1. Queraltó Compañó JM, Bosch Ferrer MA, Bedini Chesa JL, Raventós Monjo J, Fuentes-Arderiu X. Computers in clinical laboratories. Chemistry Int – Newsmagazine for IUPAC 2008;30:5–8.
2. Jovičić SŽ, Vitkus D. Digital transformation towards the clinical laboratory of the future. Perspectives for the next decade. Clin Chem Lab Med 2023;61:567–9.
3. Desiere F, Kowalik K, Fassbind C, Assaad RS, Füzéry AK, Gruson D, et al. Digital diagnostics and mobile health in laboratory medicine: an International Federation of Clinical Chemistry and Laboratory Medicine Survey on current practice and future perspectives. J Appl Lab Med 2021;6:969–79.
4. Adler J, Lenski M, Tolios A, Taie SF, Sopic M, Rajdl D, et al. Digital competence in laboratory medicine. J Lab Med 2023;47:143–8.

5. Haymond S, Master S. Why clinical laboratorians should embrace the R Programming Language – a case for learning R as a gateway to laboratory. Medicine's Digital Future Clinical Laboratory News: Association for Diagnostics & Laboratory Medicine; 2020. Available from: https://www.myadlm.org/cln/articles/2020/april/why-clinical-laboratorians-should-embrace-the-r-programming-language#.
6. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 2023;3:121–54.
7. Cloesmeijer ME, Janssen A, Koopman SF, Cnossen MH, Mathôt RAA, consortium ftS. ChatGPT in pharmacometrics? Potential opportunities and limitations. Br J Clin Pharmacol 2024;90:360–5.
8. Loh BCS, Fong AYY, Ong TK, Then PHH. Deep learning in digital health with chatgpt: a study on efficient code generation. Eur Heart J 2023;44. https://doi.org/10.1093/eurheartj/ehad655.2937.
9. Feng Y, Vanam S, Cherukupally M, Zheng W, Qiu M, Chen H, editors. Investigating code generation performance of ChatGPT with crowdsourcing social data. In: 2023 IEEE 47th annual computers, software, and applications conference (COMPSAC), Torino, Italy, June 26–30, 2023. Torino, Italy: IEEE; 2023:876–85 p.
10. Coello CEA, Alimam MN, Kouatly R. Effectiveness of ChatGPT in coding: a comparative analysis of popular large language models. Digital 2024;4:114–25.
11. Bucaioni A, Ekedahl H, Helander V, Nguyen PT. Programming with ChatGPT: how far can we go? Mach Learn Appl 2024;15:100526.
12. Nicholas G, Bhatia A. Lost in translation: large language models in non-English content analysis. Center for Democracy & Technology; 2023. Available from: https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/.
13. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. Gpt-4 technical report. arXiv preprint arXiv:230308774; 2023.
14. Meyer A, Riese J, Streichert T. Comparison of the performance of GPT-3.5 and GPT-4 with that of medical students on the written German medical licensing examination: observational study. JMIR Med Educ 2024;10:e50965.
15. Rodriguez DV, Lawrence K, Gonzalez J, Brandfield-Harvey B, Xu L, Tasneem S, et al. Leveraging generative AI tools to support the development of digital solutions in health care research: case study. JMIR Hum Factors 2024;11:e52885.
16. Bringula R. ChatGPT in a programming course: benefits and limitations. Front Education 2024;9. https://doi.org/10.3389/feduc.2024.1248705.
17. Ethikrat D. Mensch und Maschine–Herausforderungen durch Künstliche Intelligenz. Vorabfassung der Stellungnahme Berlin: Geschäftsstelle der Deutschen Ethikrats; 2023. ethikrat org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine pdf.
18. Li J. Security implications of AI Chatbots in health care. J Med Internet Res 2023;25:e47551.
19. Olatunji IE, Rauch J, Katzensteiner M, Khosla M. A review of anonymization for healthcare data. Big Data 2022. https://doi.org/10.1089/big.2021.0169.
20. Kazemitabaar M, Chow J, Ma CKT, Ericson BJ, Weintrop D, Grossman T. Studying the effect of AI code generators on supporting novice learners in introductory programming. In: Proceedings of the 2023 CHI conference on human factors in computing systems. Hamburg, Germany: Association for Computing Machinery; 2023:Article 455 p.
21. Yilmaz R, Karaoglan Yilmaz FG. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. Comput Educ: Artif Intell 2023;4:100147.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/labmed-2024-0034).