BRIEF REPORT

Liver INTERNATIONAL   WILEY

Artificial replication cohort: Leveraging AI-fabricated data for genetic studies

Wiktor Smyk[1] | Krystian Adrych[1] | Marcin Krawczyk[2,3]

[1]Department of Gastroenterology and Hepatology, Medical University of Gdansk, Gdansk, Poland

[2]Department of Medicine II, Saarland University Medical Center, Saarland University, Homburg, Germany

[3]Laboratory of Metabolic Liver Diseases, Center for Preclinical Research, Department of General, Transplant and Liver Surgery, Medical University of Warsaw, Warsaw, Poland

Correspondence
Marcin Krawczyk, Department of Medicine II, Saarland University Medical Center, Saarland University, Kirrberger Str. 100, Homburg 66421, Germany.
Email: marcin.krawczyk@uks.eu

Handling Editor: Luca Valenti

Abstract

Recent advancements in artificial intelligence (AI) present both opportunities and challenges within the scientific community. This study explores the capability of AI to replicate findings from genetic research, focusing on findings from prior work. Using an AI model without exposing any raw data, we created a dataset that closely mirrors the results of our original study, illustrating the ease of fabricating datasets with authenticity. This approach highlights the risks associated with AI misuse in scientific research. The study emphasizes the critical importance of maintaining the integrity of scientific inquiry in an era increasingly influenced by advanced AI technologies.

KEYWORDS
artificial intelligence, genomic data, language model

1 | INTRODUCTION

In the rigorous peer review process of genetic association studies, reviewers underscore the necessity of validating findings in replication cohorts which aligns with STREGA recommendations.[1] Replication highlights the role of confirming the validity of initial results, particularly when establishing significant genotype–phenotype correlations.[2] The general purpose of replication is to carefully test and verify results from candidate-gene or genome-wide association studies. This process is crucial in lending credibility to observed genetic associations. Typically, it involves reanalysing the same genetic association in an independent cohort, ensuring sufficient sample size and confirming that the populations and phenotypes studied are comparable.[3] Ideally, the replication must demonstrate consistency in the direction and magnitude of effects with the original findings. Despite these challenges, replication remains a vital and indispensable part

of genetic research, ensuring that the scientific community can build upon only the most robust and reliable findings.[4]

In a recent publication, Taloni et al.[5] utilized ChatGPT-4 and highlighted the potential concerns associated with its involvement in science. They used the latest version of this AI language model in conjunction with Advanced Data Analysis (ADA), which integrates Python for statistical analysis and data visualization. The study demonstrated AI's capability to rapidly create a fabricated clinical trial dataset in support of unverified scientific claims. This investigation demonstrated the AI's capacity to generate misleading but apparently credible scientific data, emphasizing the critical need for rigorous quality assessments in research publications.[5]

In this context, we uncover another potential risk associated with the creation of a fake replication cohort intended to authenticate our previous research findings from the manuscript we previously published in LIVER INTERNATIONAL.[6]

Abbreviations: ADA, advanced data analysis; AI, artificial intelligence; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HSI, hepatic steatosis index; MAF, minor allele frequency; PCOS, polycystic ovary syndrome; SHBG, sex hormone binding globulin.

2 | METHODS

The process of fabricating data was initiated by providing a comprehensive prompt, which outlined the specific criteria for constructing the desired database. Additionally, we submitted the entire results paragraph from our previous genotype–phenotype study published in Liver International[6] with the explicit intention of replicating it using the ChatGPT ADA system (OpenAI). Of note, the language model did not have access to the original database and was not exposed to the raw data. Our primary goal was to generate a dataset comprising $N = 500$ fabricated patient records, which included genotyping results of *MTARC1* rs2642438 variant, liver function tests (i.e., alanine aminotransferase, (ALT) and aspartate aminotransferase (AST)), sex hormone binding globulin (SHBG) and hepatic steatosis index (HSI) as previously reported in our study in PCOS patients.[6] ADA received instructions (i.e., a desired number of cases, genotype frequency and minor allele frequency (MAF)) to fabricate data that would reproduce the statistically significant results as specified in our previous research. The language model was programmed to externally generate database and figures directly into a downloadable file of choice. Subsequently, we analysed the synthetic database using IBM SPSS Statistics (software ver. 29.0). Kolmogorov-Smirnov test was used to determine whether data were normally distributed. Associations between the *MTARC1* variant and quantitative variables were analysed using the Mann-Whitney U test or Kruskal–Wallis test. Clinical variables in two cohorts were compared using the Mann–Whitney U test. *MTARC1* genotype frequencies were compared between the published and artificial cohort using Armitage's trend test. p-values < .05 were regarded as statistically significant. Of note, we were working with an AI-fabricated database, which is why ethical approval and informed consent were not applied.

3 | RESULTS

Note: The new data presented in this section is entirely generated by the AI system (ChatGPT & ADA integration) and does not represent actual clinical research findings.

This study demonstrates ChatGPT with ADA integration in replicating our original research findings on Polish women with PCOS.[6] Specifically, we focused on the *MTARC1* rs2642438 variant and its association with serum activities of ALT and AST, SHBG levels and its potential protective role against fatty liver disease. Using only 3 initial prompts and responding to two additional queries, followed by 3 more prompts for adjustments, we successfully utilized ChatGPT to generate a database draft within 10 minutes. This underscores the AI's capability to rapidly produce fake data aligning with previous research findings.

Initially, the AI was tasked to model the distribution of *MTARC1*, incorporating specific allele symbols and MAF details. This step set the stage for examining the interplay between *MTARC1* genotypes and the biochemical parameters of ALT and AST activities, SHBG

Key points

Artificial intelligence can mimic the results of genetic research, creating datasets that look just like the real ones without accessing any actual patient data. This skill brings new chances but also risks, showing why it is so important to have rigorous review and verification processes in science. This brief report serves as a reminder, stressing the critical role of validation to ensure that the data behind new medical insights is genuine and not the product of a sophisticated algorithm.

levels and HSI, aiming for a p-value close to .01. We adjusted the genotype frequencies for *MTARC1*, targeting approximately 40%–50% for both the wild-type [GG] and the heterozygous variant [GA] and 10%–15% for the homozygous variant [AA]. The language model proposed genotype frequencies of 43.6%, 42.8% and 13.6%, respectively. Minor allele frequency was .35 and stayed in line with the Polish PCOS cohort from our publications.[6]

To simulate the results for ALT, AST activities and SHBG levels, we input data from the first table detailing cohort characteristics.[6] The characteristics of the new, virtual replication cohort are shown in Table 1. The generated cohort exhibited a mean ALT activity of 17.8 ± 8.2 IU/L, AST activity of 22.3 ± 8.6 IU/L and SHBG concentration of 67.3 ± 30.9 nmol/L. The mean HSI was 34.9 ± 4.6, with 35.4% of individuals having an HSI ≥36, indicative of fatty liver. These characteristics were not significantly different ($p > .05$) from those reported in our initial study, however, as presented in Table 1, the differences in SHBG, HIS and AST between both cohorts were close to reaching the threshold of significance.

Further, the AI analysis was able to replicate the statistically significant association of the *MTARC1* rs2642438 allele with lower serum ALT ($p = .001$, Figure 1A) and AST ($p < .001$, Figure 1B) activities. For exact comparison, ALT and AST results from the previously published PCOS study[6] are presented in Figure 1C,D, respectively. As presented in Figure 1, there was no significant difference ($p = .084$) in genotype distribution between the cohorts. Importantly, the minor alleles of *MTARC1* were significantly linked to a reduced risk of fatty liver disease (OR = .72, 95%CI .55–.94; $p = .020$), as indicated by an HSI ≥ 36. This finding aligns with the data from our initial publication,[6] as all comparative analyses between cohorts yielded p-values > .05.

4 | DISCUSSION

We present a thought-provoking exploration of the capabilities of AI, specifically the ChatGPT ADA system, in replicating genetic association research. Our intention was to spotlight the potential misuse of AI in scientific research, particularly in the field of genetics, where replication of data is vital for validation. The ease and speed

with which the AI system generated a fabricated dataset mirroring the results of a previous genuine study should be seen as a sign of caution. Noteworthy, our study demonstrated the AI's capabilities

TABLE 1 Baseline characteristics of the real-life and virtual cohorts of patients with PCOS.

Variables	Polish cohort[6]	AI-generated cohort	p
Number of PCOS cases	143	500	
ALT (IU/L)	15.4 ± 9.2	17.8 ± 8.2	.387
AST (IU/L)	20.7 ± 7.2	22.3 ± 8.6	.074
SHBG	58.5 ± 35.2	67.3 ± 30.9	.061
HSI (points)	32.7 ± 7.7	34.9 ± 4.6	.052
Patients with HSI ≥ 36 points	54 (37.8%)	177 (35.4%)	.604

Note: Values are given as means ± SD.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate transaminase; HSI, hepatic steatosis index; SHBG, sex hormone binding globulin.

using only the results presented in our previous manuscript6 and parameters specification without applying the language model to raw data. This rapid generation of seemingly authentic data, capable of supporting unverified scientific claims, poses a profound challenge to the integrity of scientific research.

Recent studies have indicated that AI language models can effortlessly generate scientific articles[7] as well as create entirely fake dataset.[5] Our study underscores the importance of detailed peer review and validation processes in scientific publishing. We demonstrated the AI model's capability to replicate results and highlighted the near-significant variation in HSI, SHBG and ALT between analysed cohorts. This observation does not necessarily reflect a limitation in the model's precision. Instead, it underscores the variability and complexity in biological data, which AI-generated cohorts can generate and mirror real-life situation. The ability of AI to fabricate data that appears statistically significant requires a robust review process, potentially one that incorporates AI detection tools, to ensure the authenticity of data submitted for publication.
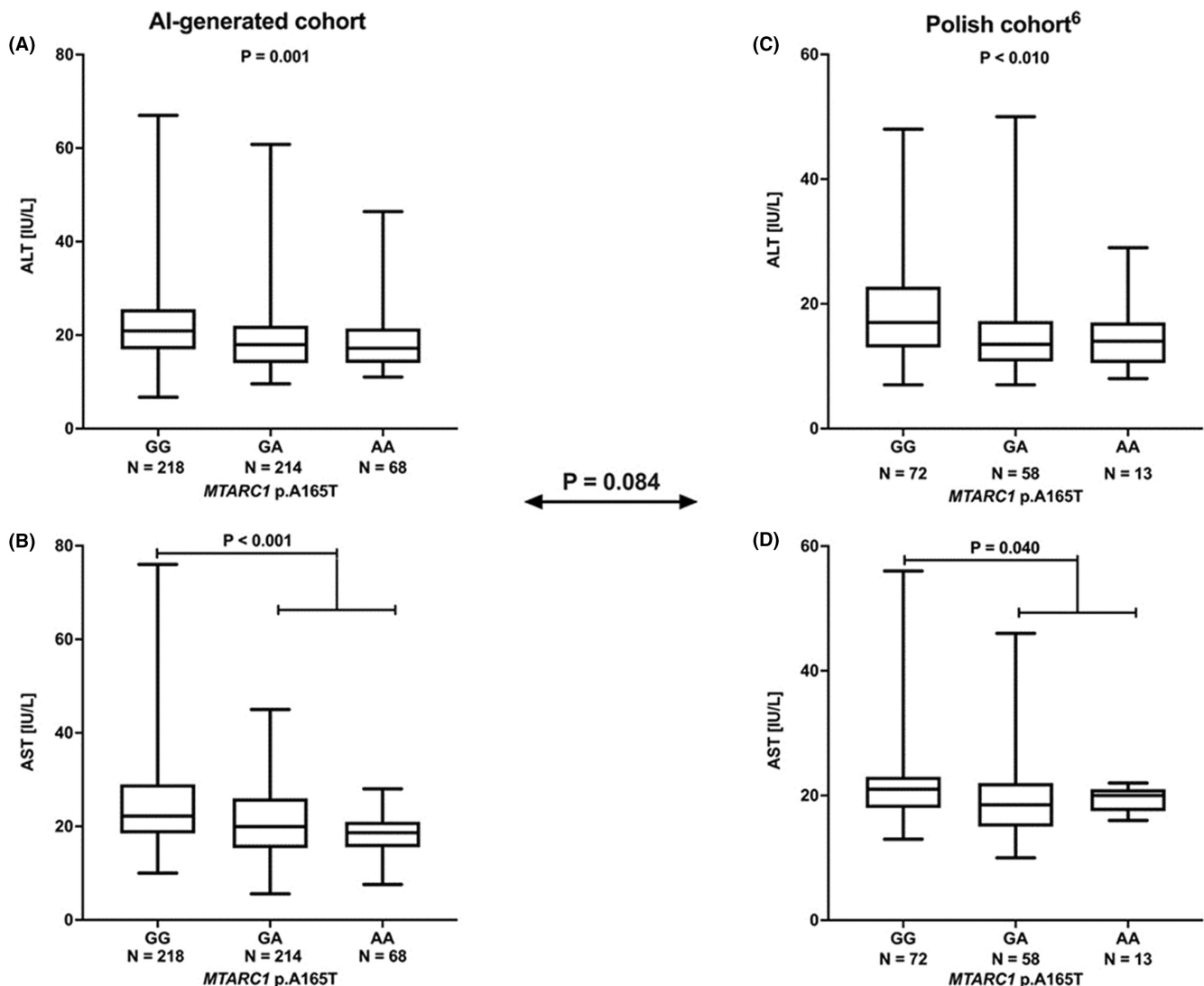

FIGURE 1 Associations between the *MTARC1* rs2642438 polymorphism and serum ALT and AST activities in the virtual (panel A and B) and previously published real-life[6] cohorts (panel C and D).

In conclusion, while AI presents remarkable opportunities for advancing scientific research but its potential for misuse in fabricating data, as demonstrated in this study, is a serious concern that requires immediate attention. This brief report should serve as a cautionary tale, highlighting the urgent need for the development of mechanisms to ensure the authenticity and integrity of science.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

ORCID

Marcin Krawczyk https://orcid.org/0000-0002-0113-0777

REFERENCES

1. Little J, Higgins JPT, Ioannidis JPA, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Hum Genet*. 2009;125:131-151.
2. Kraft P, Zeggini E, Ioannidis JPA. Replication in genome-wide association studies. *Stat Sci*. 2009;24:561-573.
3. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. *Nat Rev Methods Primers*. 2021;1:59.
4. NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype–phenotype associations. *Nature*. 2007;447:655-660.
5. Taloni A, Scorcia V, Giannaccare G. Large language model advanced data analysis abuse to create a fake data set in medical research. *JAMA Ophthalmol*. 2023;141:1174-1175.
6. Smyk W, Papapostoli I, Żorniak M, et al. Liver phenotypes in PCOS: analysis of exogenous and inherited risk factors for liver injury in two European cohorts. *Liver Int*. 2023;43:1080-1088.
7. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *J Med Internet Res*. 2023;25:e46924.