



PROMPT ENGINEERING IN MEDICINE: HARASSING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN CLINICAL PRACTICE, EDUCATION, AND RESEARCH

Farrukh Musinovich Khurramov¹

Dostonjon Samat Nurmamatov²

¹Tashkent State Medical University, Tashkent, Uzbekistan

²Turon University, Karshi, Uzbekistan

<https://doi.org/10.5281/zenodo.19691334>

ARTICLE INFO

Received: 15th April 2026

Accepted: 20th April 2026

Online: 21st April 2026

KEYWORDS

Prompt engineering, large language models, clinical decision support, medical education, artificial intelligence in healthcare, evidence-based medicine, digital literacy.

ABSTRACT

Large language models (LLMs) are transforming approaches to medical information processing, but the quality of their output critically depends on the structure of the input query. Prompt engineering—a methodology for the purposeful construction of text instructions for LLMs—is becoming an essential tool for physician researchers, clinicians, and medical faculty. This article systematizes the basic principles of prompt engineering, including defining a role, context, output format, and reasoning chain. Specific application scenarios are considered in three domains: scientific research (cohort data analysis, systematic literature review, manuscript editing), medical education (problem-based learning, clinical case and MCQ generation), and clinical practice (decision support, differential diagnosis, preoperative planning). Ethical considerations are discussed, including the risk of hallucinations, data privacy, and the essential role of expert verification. Development prospects are linked to the development of digital literacy as a core competency of medical professionals.

PROMPT ENGINEERING В МЕДИЦИНЕ: УПРАВЛЕНИЕ ПОТЕНЦИАЛОМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КЛИНИЧЕСКОЙ ПРАКТИКЕ, ОБРАЗОВАНИИ И НАУЧНЫХ ИССЛЕДОВАНИЯХ

Хуррамов Фаррух Мусинович¹

Нурмаматов Достонжон Самат угли²

¹Ташкентский государственный медицинский университет,
г. Ташкент, Узбекистан

²Турон Университет, г. Карши, Узбекистан

<https://doi.org/10.5281/zenodo.19691334>

ARTICLE INFO

Received: 15th April 2026

Accepted: 20th April 2026

Online: 21st April 2026

KEYWORDS

ABSTRACT

Большие языковые модели (Large Language Models, LLM) трансформируют подходы к обработке медицинской информации, однако качество их



Prompt engineering, большие языковые модели, клиническая поддержка принятия решений, медицинское образование, искусственный интеллект в здравоохранении, доказательная медицина, цифровая грамотность.

выходных данных критически зависит от структуры входного запроса. Prompt engineering – методология целенаправленного конструирования текстовых инструкций для LLM – становится необходимым инструментом для врачей-исследователей, клиницистов и преподавателей медицинских вузов. В настоящей статье систематизированы базовые принципы prompt engineering, включая задание роли, контекста, формата вывода и цепочки рассуждений. Рассмотрены конкретные сценарии применения в трёх доменах: научные исследования (анализ когортных данных, систематический обзор литературы, редактирование рукописей), медицинское образование (проблемно-ориентированное обучение, генерация клинических кейсов и MCQ) и клиническая практика (поддержка принятия решений, дифференциальная диагностика, предоперационное планирование). Обсуждены этические ограничения, включая риск галлюцинаций, конфиденциальность данных и незаменимую роль экспертной верификации. Перспективы развития связаны с формированием цифровой грамотности как базовой компетенции медицинского профессионала.

Введение. Стремительная интеграция больших языковых моделей (LLM) в биомедицинское пространство обозначила принципиально новую парадигму взаимодействия врача с информационными системами. Модели семейства GPT-4, Claude, Gemini и их открытые аналоги продемонстрировали способность решать задачи медицинских экзаменов на уровне, сопоставимом с квалифицированными специалистами [1, 2]. Singhal и соавторы показали, что LLM, обученные на клинических корпусах, кодируют объём медицинских

знаний, достаточный для прохождения квалификационных тестов USMLE Step 1-3 с результатами выше порогового значения [3]. Эти достижения радикально сместили дискуссию от вопроса «способен ли ИИ работать с медицинскими данными?» к более прагматичному: «каким образом клиницист может извлечь из LLM максимально релевантный и безопасный ответ?» Именно здесь пролегает предметная область prompt engineering – дисциплины, изучающей принципы конструирования входных текстовых инструкций (промпов),



IF = 9.2

определяющих поведение генеративной модели [4, 5].

Термин «prompt engineering» охватывает совокупность приёмов, позволяющих структурировать запрос к LLM таким образом, чтобы минимизировать семантическую неопределённость ответа и направить генерацию в заданное предметное русло [6]. В отличие от классического программирования, где инструкции формализованы на языке кода, промпт представляет собой директиву на естественном языке, и его эффективность определяется не синтаксической корректностью, а прагматической точностью формулировки. Thirunavukarasu и соавторы подчёркивают, что даже незначительные вариации в структуре промпта могут приводить к качественно различным выходным данным, варьирующимся от клинически адекватных рекомендаций до фактически ошибочных утверждений [7]. Цель настоящей статьи – систематизировать методологические основы prompt engineering применительно к трём ключевым доменам медицины: научным исследованиям, образовательному процессу и клинической практике, а также определить этические рамки использования LLM в здравоохранении.

Базовые принципы конструирования промптов. Архитектура эффективного промпта, согласно современным рекомендациям, базируется на

четырёх структурных элементах: назначении роли, определении контекста, формулировке задачи и спецификации формата вывода [8, 9]. Назначение роли («Действуй как опытный хирург-онколог») активирует в латентном пространстве модели кластеры знаний, ассоциированные с указанной специализацией, что повышает релевантность терминологии и глубину клинического рассуждения. Определение контекста подразумевает предоставление модели анамнестических данных, лабораторных показателей или описания дизайна исследования – иными словами, всей информации, необходимой для генерации обоснованного ответа. Формулировка задачи должна быть максимально конкретной: вместо расплывчатого «расскажи о лечении» целесообразно указать «сравни эффективность неoadъювантной химиотерапии режимом FOLFIRINOX и гемцитабин/наб-паклитаксел при резектабельной аденокарциноме поджелудочной железы по критериям pCR и OS». Наконец, спецификация формата вывода («ответ в форме структурированной таблицы с колонками: препарат, механизм действия, уровень доказательности») предупреждает генерацию избыточного или неструктурированного текста [10].

Особую роль приобретает техника цепочки рассуждений (chain-of-thought prompting), впервые описанная Wei и соавторами [11]. Инструкция «рассуждай пошагово, прежде чем дать финальный ответ»



побуждает модель эксплицитно развернуть логическую последовательность от исходных данных к выводу, что существенно снижает частоту фактических ошибок в задачах клинического рассуждения [12]. White и коллеги предложили каталог шаблонов промптов (prompt patterns), включающий паттерн «flipped interaction», при котором

модели поручается задавать уточняющие вопросы пользователю, прежде чем формировать ответ, – приём, особенно полезный при сборе дифференциально-диагностических данных [13].

Таблица 1

Примеры структурированных промптов для различных медицинских задач

Домен	Задача	Пример промпта (фрагмент)	Ключевой принцип
Наука	Редактирование рукописи	«Действуй как рецензент журнала Lancet. Проверь абзац на логическую связность, грамматику и соответствие стилю АМА»	Роль + формат
Наука	Статистический анализ	«Предложи дизайн статистического анализа для сравнения 30-дневной летальности между двумя когортами (n=240 vs n=185). Укажи критерии выбора теста»	Контекст + задача
Образование	Генерация кейса PBL	«Создай клинический кейс для PBL-сессии: пациент 62 лет, механическая желтуха. Раскрой данные последовательно в 3 этапа»	Роль + цепочка рассуждений
Образование	Генерация MCQ	«Составь 5 вопросов MCQ с одним правильным ответом по теме патофизиология острого панкреатита. Уровень сложности: клиническая резидентура. К каждому вопросу добавь объяснение»	Задача + формат
Клиника	Дифф. диагностика	«Пациент 45 лет, острая боль в правом подреберье, лейкоцитоз $15 \times 10^9/\text{л}$, СРБ 82 мг/л. Предложи 10 дифференциальных диагнозов, ранжируя по вероятности. Обоснуй каждый»	Контекст + цепочка рассуждений



Клиника	Адаптация для пациента	«Переформулируй следующее заключение КТ на язык, понятный пациенту без медицинского образования. Сохрани клиническую точность»	Задача + ограничение
---------	------------------------	--	----------------------

Prompt engineering в научных исследованиях. Применение LLM в исследовательском процессе охватывает широкий спектр задач, от языковой редактуры рукописей до концептуального анализа дизайна клинических испытаний. Редактирование научных текстов представляет собой одну из наиболее зрелых областей: модели способны выявлять нарушения логической связности, стилистические несоответствия и грамматические ошибки при условии, что промпт содержит указание на целевой стиль (AMA, Vancouver, APA) и роль рецензента конкретного журнала [14]. Clusmann и соавторы отмечают, что использование LLM для предварительного рецензирования позволяет сократить время подготовки рукописи на 30-40%, хотя финальная экспертная вычитка остаётся обязательной [15].

Более сложным сценарием является анализ когортных исследований. Если врач-исследователь формулирует промпт, включающий характеристики основной группы и группы сравнения, первичные и вторичные конечные точки, тип распределения переменных и допустимый уровень ошибки первого рода, LLM способна предложить обоснованный

статистический план: выбор между критерием Стьюдента и U-критерием Манна-Уитни, необходимость поправки Бонферрони при множественных сравнениях или применение модели пропорциональных рисков Кокса для анализа выживаемости [16]. Sandmann и коллеги в систематическом анализе показали, что GPT-4 превосходит обычный поиск Google по точности рекомендаций для клинических решений при условии адекватной структуры промпта [17]. Тем не менее, для задач систематического обзора литературы LLM способны генерировать поисковые стратегии для PubMed и Embase, формировать матрицы «PICOS» и предлагать критерии включения/исключения, однако итоговый скрининг абстрактов и полнотекстовых статей по-прежнему требует участия двух независимых рецензентов в соответствии с методологией PRISMA [18].

Prompt engineering в медицинском образовании. Проблемно-ориентированное обучение (problem-based learning, PBL) составляет сердцевину современного медицинского образования, и LLM открывают перспективу масштабируемой генерации клинических кейсов.



IF = 9.2

Safranek и соавторы показали, что GPT-4, инструктированный промптом с указанием нозологии, возраста пациента, стадии заболевания и дидактической цели, генерирует сценарии, по содержательности релевантности не уступающие кейсам, составленным преподавателями [19]. Ключевой методический приём состоит в поэтапном раскрытии информации (staged disclosure): промпт должен предписывать модели разделить презентацию случая на три-четыре этапа, каждый из которых содержит новый массив данных – от жалоб и анамнеза через результаты физикального осмотра к лабораторно-инструментальным данным. Такая структура воспроизводит логику клинического мышления и стимулирует активное формулирование гипотез студентами.

Объяснение топографической анатомии и техники хирургических вмешательств представляет отдельную педагогическую задачу, где LLM могут служить «интерактивным тьютором». Промпт

вида «опиши послойный хирургический доступ при лапароскопической холецистэктомии, указывая анатомические ориентиры, используемые инструменты и потенциальные ятрогенные осложнения на каждом этапе» генерирует детализированное описание, пригодное для подготовки к операции [20]. Генерация вопросов множественного выбора (MCQ) также существенно выигрывает от структурированного промптинга: Strong и коллеги установили, что качество MCQ, созданных GPT-4, сопоставимо с экзаменационными вопросами, составленными опытными преподавателями, при условии спецификации уровня сложности, количества дистракторов и требования обоснования правильного ответа [21]. Gilson и соавторы подтвердили, что генерируемые объяснения к каждому варианту ответа повышают обучающую ценность MCQ и могут использоваться для самостоятельной подготовки студентов [22].

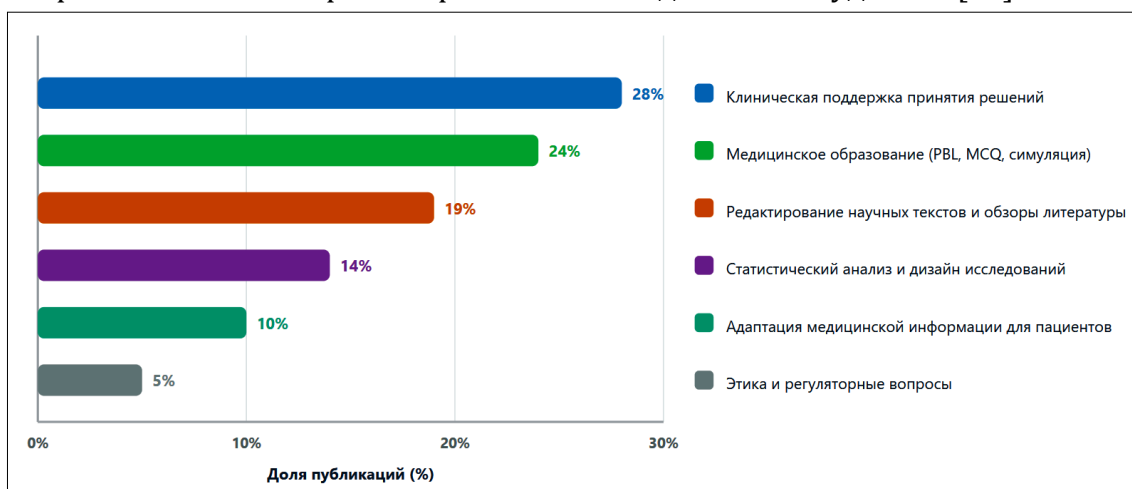


Рисунок 1. Распределение областей применения prompt

engineering в медицине по данным публикаций 2022-2025 гг.



Prompt engineering в клинической практике.

Клиническая поддержка принятия решений (clinical decision support, CDS) представляет собой, пожалуй, наиболее дискутируемую область применения LLM. Kanjee и соавторы продемонстрировали, что GPT-4, получив структурированное описание клинического случая с полным набором данных анамнеза, физикального осмотра, лабораторных и инструментальных исследований, верно идентифицировал редкий диагноз в 64% случаев из серии диагностических задач, опубликованных в NEJM Case Records [23]. Savage и коллеги подтвердили, что стратегия «diagnostic reasoning prompting», при которой модели предписывается эксплицировать ход дифференциально-диагностического рассуждения, повышает точность итогового диагноза по сравнению с прямым запросом «что это за диагноз?» [24]. Практическая ценность такого подхода состоит не в замене клинического мышления врача, а в расширении когнитивного горизонта: модель может предложить редкие диагнозы, которые клиницист мог бы не рассмотреть в условиях когнитивной перегрузки или anchoring bias.

Предоперационное планирование – ещё одна ниша, где грамотно сконструированный промпт способен дать ощутимый результат. Хирург может сформулировать запрос, содержащий данные визуализации (размер и локализацию опухоли, соотношение с магистральными сосудами,

вариантную анатомию), и получить предложение по выбору хирургического доступа, объёму резекции и прогнозируемым техническим сложностям. Mehandru и коллеги оценивали LLM как «клинических агентов» и пришли к выводу, что при достаточной контекстуализации промпта модели демонстрируют способность к интеграции множественных переменных, аппроксимируя процесс принятия решений опытного консультанта [25]. Адаптация медицинской терминологии для пациентов – часто недооцениваемый аспект клинической коммуникации – также значительно упрощается при помощи промптов с чётко сформулированным целевым уровнем health literacy. Chen и соавторы показали, что сообщения, сгенерированные LLM для пациентов, оценивались врачами как более эмпатичные и понятные по сравнению с сообщениями, написанными самими клиницистами [26].

Этические ограничения и риски. Центральным риском использования LLM остаётся феномен «галлюцинаций» – генерации фактически неверных, но стилистически убедительных утверждений. В медицинском контексте галлюцинация модели может проявляться в виде несуществующей ссылки на клиническое руководство, искажённых показателей чувствительности и специфичности диагностического теста или вымышленного лекарственного



взаимодействия. Omiye и соавторы установили, что LLM могут воспроизводить расовые предрассудки, присутствующие в обучающих данных, генерируя клинические рекомендации, дифференцированные по расовому признаку без научного обоснования [27]. Единственным надёжным механизмом противодействия галлюцинациям является верификация любого выхода LLM через принципы доказательной медицины (EBM): сопоставление рекомендации с актуальными клиническими руководствами, систематическими обзорами Cochrane и оригинальными исследованиями с высоким уровнем доказательности.

Конфиденциальность персональных данных пациентов представляет собой вторую критическую проблему. Передача деперсонализированных, но потенциально реидентифицируемых клинических данных в LLM, функционирующую на внешних серверах, создаёт риск нарушения требований GDPR, HIPAA и локальных законодательных актов о защите персональных данных [28]. Meskó и Toró настаивают на необходимости регуляторного надзора за LLM в здравоохранении, подчёркивая, что ни одна модель не прошла формальной сертификации как медицинское устройство, и ответственность за любое клиническое решение, принятое с участием ИИ, лежит исключительно на враче-специалисте [29]. Данный принцип – «human-in-the-loop» – должен рассматриваться как

нередуцируемое условие интеграции LLM в клинический workflow. Goh и соавторы экспериментально показали, что врачи, некритически принимающие рекомендации LLM, демонстрируют снижение качества диагностического рассуждения по сравнению с контрольными условиями без ИИ, что свидетельствует о феномене «automation bias» [30].

Перспективы и направления развития. Prompt engineering трансформируется из узкоспециализированного навыка в базовую компетенцию медицинского профессионала. Параллель с освоением статистических пакетов (SPSS, R, Stata) в 1990-2000-х годах представляется уместной: подобно тому, как владение методами многомерного анализа стало необходимым элементом исследовательской квалификации, умение конструировать эффективные промпты для LLM становится обязательным компонентом цифровой грамотности врача-исследователя XXI века. Liévin и соавторы продемонстрировали, что стратегия few-shot chain-of-thought prompting повышает точность ответов LLM на клинические вопросы до 80-90% в зависимости от предметной области, что значительно превосходит результаты при использовании zero-shot подходов [4]. Эти данные свидетельствуют о том, что потенциал LLM определяется не столько архитектурой модели, сколько квалификацией пользователя, формулирующего запрос.



IF = 9.2

Заклучение. Интеграция курсов по prompt engineering в программы медицинских вузов, резидентур и программ непрерывного медицинского образования представляется стратегически обоснованным шагом. При этом критически необходимо подчёркивать двойственную природу инструмента: LLM способна существенно расширить когнитивные ресурсы врача, однако при некритическом использовании она же может стать источником

систематических ошибок. Формирование культуры ответственного использования ИИ, основанной на принципах ЕВМ, транспарентности промптов и обязательной экспертной верификации, станет определяющим фактором того, сумеет ли медицинское сообщество реализовать трансформативный потенциал генеративных технологий без ущерба для безопасности пациентов.

References:

1. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198.
2. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical competency examinations. *arXiv preprint arXiv:2303.13375*. 2023.
3. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172–180.
4. Liévin V, Hother CE, Motzfeldt AG, Winther O. Can large language models reason about medical questions? *Patterns*. 2024;5(3):100943.
5. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. 2023;6:1169595.
6. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*. 2023;11(6):887.
7. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930–1940.
8. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388(13):1233–1239.
9. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst*. 2023;47(1):33.
10. Wang X, Chen Y, Wang L, Liu Z, Liu J. Prompt engineering in consistency and reliability with the medical AI correspondence. *npj Digit Med*. 2024;7(1):207.
11. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst*. 2022;35:24824–24837.



12. Savage T, Nayak A, Gallo R, Rangan E, Chen JH. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *npj Digit Med.* 2024;7(1):20.
13. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv preprint arXiv:2302.11382.* 2023.
14. Tang L, Sun Z, Idnay B, Nestor JG, Soroush A, Elber Melo PA, et al. Evaluating large language models on medical evidence summarization. *npj Digit Med.* 2023;6(1):158.
15. Clusmann J, Kolbinger FR, Muti HS, Grberó GI, Nikberó JN, Schneider V, et al. The future landscape of large language models in medicine. *Commun Med.* 2023;3(1):141.
16. Agrawal M, Hegselmann S, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. *arXiv preprint arXiv:2205.12689.* 2022.
17. Sandmann S, Riepenhausen S, Plagwitz L, Dugas M. Systematic analysis of ChatGPT, Google search, and Llama 2 for clinical decision support. *Nat Commun.* 2024;15(1):2050.
18. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. *Narra J.* 2023;3(1):e103.
19. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large language models in medical education: applications and implications. *JMIR Med Educ.* 2023;9:e50945.
20. Bhayana R, Krishna S, Bleakney RR. Chatbots and large language models in radiology: a practical primer for clinical and research applications. *Radiology.* 2024;310(2):e232756.
21. Strong E, DiGiammarino A, Weng Y, Basaviah P, Hosamani P, Kumar A, et al. Chatbot vs medical student performance on free-response clinical reasoning examinations. *JAMA Intern Med.* 2023;183(9):1028–1030.
22. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ.* 2023;9:e45312.
23. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. *JAMA.* 2023;330(1):78–80.
24. Savage T, Nayak A, Gallo R, Rangan E, Chen JH. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *npj Digit Med.* 2024;7(1):20.
25. Mehandru N, Miao BY, Almaraz ER, Sushil M, Butte AJ, Alaa A. Evaluating large language models as agents in the clinic. *npj Digit Med.* 2024;7(1):84.
26. Chen S, Guevara M, Moningi S, Hoebbers F, Kann BH. The effect of using a large language model to respond to patient messages. *Lancet Digit Health.* 2024;6(5):e379–e381.
27. Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. *npj Digit Med.* 2023;6(1):195.
28. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digit Med.* 2023;6(1):120.



29. Duong D, Solomon BD. Analysis of large-language model versus human performance for genetics questions. *Eur J Hum Genet.* 2024;32(4):466–468.
30. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman J, et al. Large language model influence on diagnostic reasoning: a randomized clinical vignette study. *JAMA Netw Open.* 2024;7(10):e2440969.