

## ЗАКЛЮЧЕНИЕ

на основе критериев значимости и доказательности eLife для рукописи

### «Использование больших языковых моделей для решения тестовых заданий по анестезиологии и реаниматологии: сравнительное исследование»

Значимость работы – **полезная**, так как как исследование подчеркивает важность изучения и верификации возможностей больших языковых моделей (LLM) в решении стандартизированных медицинских тестов на русском языке — области, которая ранее была малоизучена. Его выводы носят скорее уточняющий и иллюстративный характер, подтверждая и конкретизируя для российского контекста уже известный из международных публикаций тренд о высокой формальной точности LLM в подобных задачах, что полезно для осмысления их роли в медицинском образовании и предостерегает от некритичного использования.

Доказательность методов исследования **твердая**, поскольку методология исследования является корректной и надежной для поставленной узкой задачи: использован прямой сравнительный дизайн, несколько статистических подходов для анализа и многократные запросы к моделям для повышения стабильности оценки. Однако выводы работы о «высокой точности» LLM поддерживаются данными лишь с существенными оговорками, так как ограничены малым объемом теста (30 вопросов), форматом с одним правильным ответом (не отражающим клиническое мышление) и потенциальным влиянием «утечки» заданий в обучающие данные моделей, что признается авторами как ключевое ограничение.

<https://doi.org/10.21320/1818-474X-2026-2-176-185>

**Значимость\***: выводы, имеющие теоретическое или практическое значение, выходящие за рамки одной области знаний

Ориентир    Фундаментальная    Важная    Ценная    **Полезная**

**Доказательность\*\***: подходящая и надежная методология, соответствующая современному уровню технического обеспечения поставленной задачи

Исключительная    Убедительная    Веская    **Твердая**    Неполная    Неадекватная

#### \* Критерии eLife: оценка научной значимости

- Ориентир: результаты с глубокими последствиями, которые, как ожидается, будут иметь широкое влияние на развитие науки
- Фундаментальные: результаты, которые существенно улучшают наше понимание исследуемой проблемы
- Важные: выводы, имеющие теоретическое или практическое значение, выходящие за рамки одной области знаний
- Ценные: выводы, имеющие теоретическое или практическое значение для конкретной области знаний
- Полезные: результаты, которые подчеркивают важность изучения конкретной проблемы

#### \*\*Критерии eLife: оценка доказательности

- Исключительная: образцовое использование существующих подходов, устанавливающих новые стандарты для области
- Убедительная: доказательства, в которых представлены методы, результаты и их анализ, более убедительные, чем применяемые в настоящее время стандартные методы
- Веская: подходящая и надежная методология, соответствующая современному уровню технического обеспечения поставленной задачи
- Твердая: методы, результаты и их анализ в целом подтверждают выводы с незначительными недостатками
- Неполная: основные выводы подтверждаются лишь частично
- Неадекватная: методы, результаты и их анализ не подтверждают основные выводы