

Чем проще, тем лучше: методология упрощения текстов для пациентов с афазией с помощью генеративных моделей



Публикация подготовлена в результате проведения исследования по проекту №24-00-033 «Экспериментальное изучение и моделирование когнитивных механизмов речевой деятельности» в рамках Программы «Научный фонд Национального исследовательского университета “Высшая школа экономики” (НИУ ВШЭ)» в 2025г.

А.В. Колмогорова
akolmorova@hse.ru

А.В. Марголина
avmargolina@edu.hse.ru

А.А. Тельнова
aatelnova@edu.hse.ru

ВВЕДЕНИЕ

Афазия, приобретенное речевое расстройство, часто возникающее в результате инсульта или черепно-мозговой травмы (Pedersen et al., 2004), может нарушить способность человека читать и понимать текст.

Восстановление этой способности является одной из задач логопедии (Yang et al., 2024). Однако существующие тексты в логопедической практике изначально написаны для детей и часто не соответствуют уровню когнитивных и языковых возможностей пациентов, а также не учитывают их интересы и предшествующий опыт.

Для решения этой проблемы была поставлена **цель** – изучить потенциал и ограничения **LLM** в автоматизации поэтапного упрощения текстов для конкретной **целевой группы** – **пациентов с афазией различной степени тяжести.**

Были разработаны **специализированные подходы к упрощению текста по уровням согласно степени выраженности патологии**, позволяющие сохранить сложность и информативность, подходящие для взрослых читателей, и в то же время сделать язык более доступным (Kasdan et al., 2025). Также была оценена способность моделей сохранять смысл в текстовых материалах при одновременном снижении когнитивной нагрузки.

МАТЕРИАЛЫ И МЕТОДЫ

- Создание параллельного корпуса из 134 оригинальных текстов и трёх их адаптированных версий (всего 1011 предложений × 3 уровня).
- Разработка чётких критериев разметки для упрощения текстов:
 - Уровень 1: простые предложения без сложных конструкций.
 - Уровень 2: исключение числительных, иностранных слов, длинных предложений.
 - Уровень 3: минимальная длина (3–5 слов), использование только распространённой лексики.
- Тестирование и сравнение нескольких LLM:
 - Gemma (gemma-2-9b-it)
 - Llama Storm (Llama-3.1-Storm-8B)
 - Qwen (Qwen2.5-7B-Instruct)
 - Vikhr (Vikhr-7B-instruct)
- Клиническая апробация упрощённых текстов среди пациентов с КМА.

КРИТЕРИИ РАЗМЕТКИ

Уровень 1	Уровень 2	Уровень 3
	Простые предложения без скопления деталей	
Отсутствие деепричастных и причастных оборотов		
Отсутствие пассивного залога		
Отсутствие низкочастотных слов (особенно глаголов)		
	Отсутствие числительных	
	Отсутствие иностранных слов (имен, топонимов и пр.)	
	Одна пропозиция в предложении	
	Длина предложения: 5-7 слов	
		Длина предложения: 3-5 слов

СПИСОК ЛИТЕРАТУРЫ

1. Pedersen, P. M., Vinter, K., & Olsen, T. S. (2004). Aphasia after stroke: Type, severity and prognosis. The Copenhagen aphasia study. Cerebrovascular Diseases, 17(1), 35–43. <https://doi.org/10.1159/000073896>

2. Yang, F. A., Huang, C. C., Lu, C. H., Wu, P. J., Escorpizo, R., & Chen, H. C. (2024). Speech therapy for poststroke aphasia: A systematic review and network meta-analysis of randomized controlled trials. Research Square. <https://doi.org/10.21203/rs.3.rs-4312160/v1>

3. Kasdan, A.V., Levy, D.F., Pedisch, I., Wilson, S.M., Herrington, D. A practical guide to translating scientific publications into aphasia-friendly summaries. *Perspect ASHA Spec Interest Groups* 2025; in press

4. Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. arXiv preprint. <https://arxiv.org/abs/1904.09675>

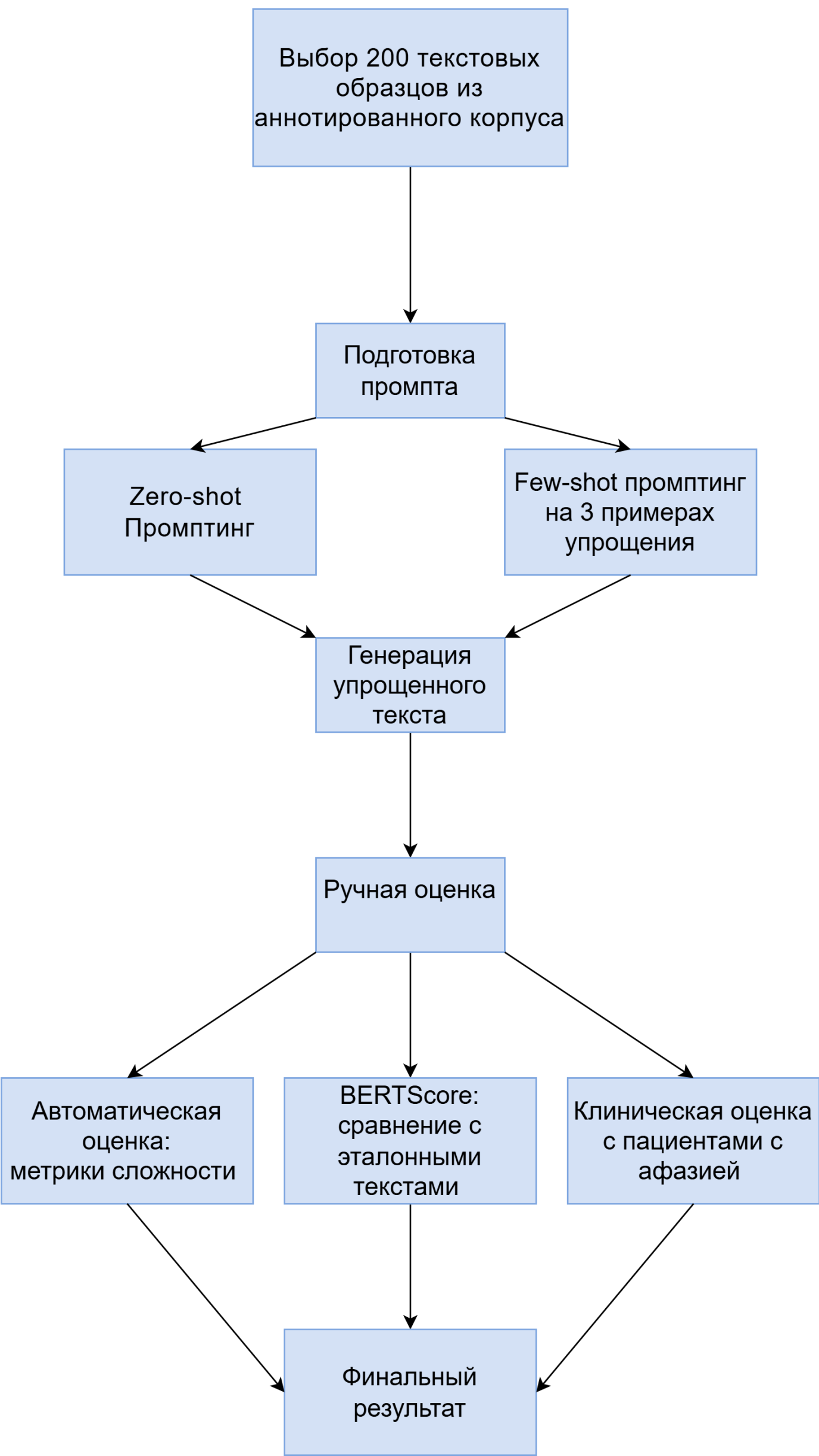
5. McLaughlin, G. (1969). SMOG grading: A new readability formula. Journal of Reading, 12(8), 639–646.

ФРАГМЕНТ ПАРАЛЛЕЛЬНОГО ДАТАСЕТА

Оригинал	Уровень 1	Уровень 2	Уровень 3
Любитель железа даже придумал удобный способ поедания вилок и ножей: чтобы было легче глотать, он ломал их пополам.	Любитель железа даже придумал легкий способ есть вилки и ножи: он ломал их пополам и глотал.	Мужчина ломал вилки и ножи на две части. Так было проще глотать их.	Мужчина ломал вилки. Так было проще есть их.
Абсолютное большинство детей не понимает и не может еще понимать, что собака – это прежде всего огромная ответственность.	Большинство детей еще не понимает ответственности за собаку.	Многие дети не понимают ответственности за собаку.	Собака - большая ответственность. Многие дети этого не понимают.
Маленький командир алы со взмохшим лбом и в темной от пота на спине белой рубашке, находившийся внизу холма у открытого подъема, то и дело подходил к кожаному ведру в первом взводе, черпал из него пригоршнями воду, пил и мочил свой тюрбан.	Маленький командир алы то и дело подходил к кожаному ведру. Он черпал из него воду, пил ее и мочил свой тюрбан.	Маленький командир то и дело черпал воду. Он пил ее и мочил свою шапку.	Командир постоянно черпал воду из ведра. Он пил ее.

ХОД ИССЛЕДОВАНИЯ

Чтобы оценить эффективность каждой выбранной языковой модели, мы разработали **пайплайн** эксперимента. Использовался набор из 200 идентичных семплов, отобранных из датасета.



Эксперименты проводились с использованием **zero-shot** и **few-shot промптинга**. При **zero-shot** модели были даны разработанные инструкции. Для **few-shot** были включены три примера упрощенных текстов.

После **генерации** проводилась **двухэтапная оценка** каждой модели. Сначала была проведена **предварительная ручная оценка**, чтобы исключить модели с неудовлетворительными результатами. Затем тексты моделей сравнивались с «золотым стандартом» и оценивались с помощью следующих **метрик**: SMOG, ASL, TTR, FRE GL, BERTScore. Далее тексты прошли **клиническую апробацию** среди пациентов с КМА.

ANNOTATION

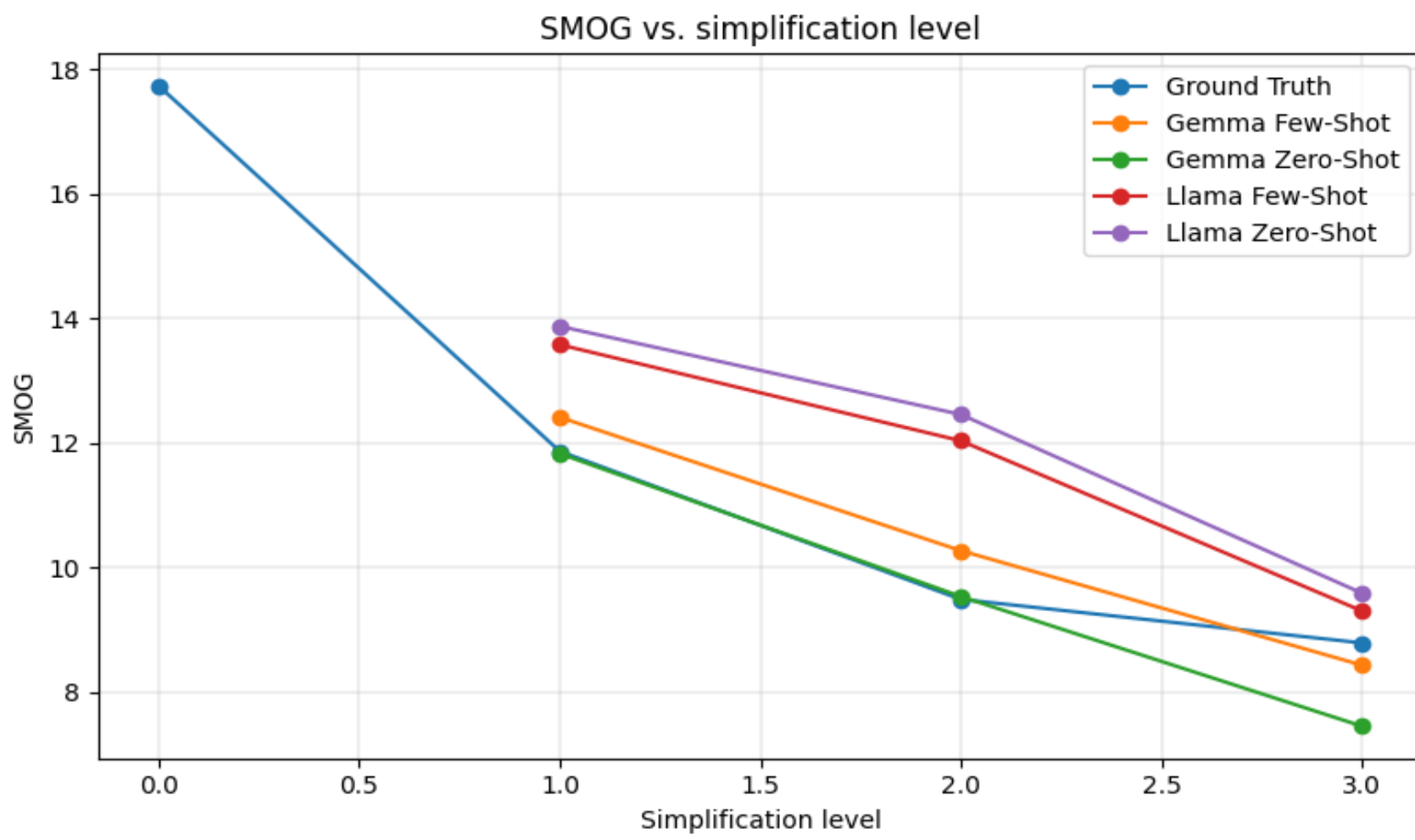
- Purpose:** To investigate the effectiveness of using LLMs for multilevel text simplification for patients with aphasia.
- Method:** A parallel corpus consisting of 134 in Russian was prepared. Russian-speaking annotators generated three simplified versions for each original sentence, corresponding to three levels of disease severity. Our designed dataset were used in a few-shot prompting strategy applied to four LLMs.
- Results:** Models were evaluated using linguistic complexity metrics and BERTScore for semantic similarity. The top LLM's output was clinically validated with 12 complex motor aphasia patients.
- Conclusion:** Few-shot prompted automated simplification effectively adapts texts for complex motor aphasia patients.

НУГ «КОГНИТИВНЫЕ ИССЛЕДОВАНИЯ ЯЗЫКА»
tg: <https://t.me/cognitiveling>

РЕЗУЛЬТАТЫ

Модели **Qwen** и **Vikhr** не прошли ручную оценку, поэтому дальнейшая работа проводилась с **Gemma** и **Llama Storm**.

Few-shot промптинг показал наилучший результат по выбранным метрикам при поуровневой генерации текста, что обеспечивает оптимальный баланс между контролируемым упрощением и сохранением лингвистической целостности.



BERTScore, zero-shot

Модель	Уровень 1	Уровень 2	Уровень 3
Llama	0.788	0.772	0.754
Gemma	0.789	0.777	0.772

BERTScore, few-shot

Модель	Уровень 1	Уровень 2	Уровень 3
Llama	0.812	0.777	0.764
Gemma	0.805	0.785	0.778

Обе модели дают результаты, которые схожи с «золотым стандартом».

КЛИНИЧЕСКАЯ ВАЛИДАЦИЯ

Валидация проводилась с **12** пациентами с разной степенью КМА. Тексты сгруппированы по степени афазии: №1, №2 – средняя степень; №3, №4 – средне-легкая; №5, №6 – легкая.

Тексты №1, №3, №5 – тексты из логопедических материалов.

Тексты №2, №4, №6 – тексты из нашей выборки.

	Текст № 1	Текст № 2	Текст №3	Текст №4	Текст №5	Текст №6
Средний балл из 1000	733	893	919	616	979	786
Средняя оценка интереса к тексту из 100	50	75	66	33	33	66

Формула: Средняя оценка текста = (100 - скорость чтения) + 100*(точность и полнота ответов)+ 100*1(интересный текст) + 100*1(несложный текст).

ВЫВОДЫ

- При использовании LLM с инструкциями, мы можем получать тексты с показателями сложности, которые сопоставимы с эталонными значениями.
- Как показывают значения метрики BERTScore и результаты клинической валидации, содержание текста, соответствующего интересам пациента, сохраняется.
- Перспективы:**
 - Доработка инструкций-промптов (например, для сюжетного упрощения)
 - Валидация инструмента в клинической практике
- Ограничения:**
 - Вычислительные ресурсы, долгая оценка в клинических условиях