

Большие языковые модели в неврологии

Е.Д. Намиот

Аннотация - Развитие систем искусственного интеллекта (ИИ), в частности, так называемых базовых моделей и больших языковых моделей, открыло новую эру на стыке ИИ и неврологии (нейронауки). Эти модели позволяют работать с разнообразными наборами данных в разных модальностях.

По сравнению с классическими вычислительными подходами, которые, в основном, опирались на традиционные методы машинного обучения, такие модели представляют собой значительный шаг вперед. Они демонстрируют сильную обобщающую способность, могут улавливать сложные пространственно-временные зависимости, обнаруженные в данных. Это достигается, в частности, за счет сквозного обучения непосредственно на необработанных данных. Базовые модели, потенциально, могут применяться во всех основных нейробиологических областях, охватывая нейровизуализацию и обработку данных, интерфейсы мозг-компьютер и нейронное декодирование, молекулярную нейробиологию и геномное моделирование, клиническую помощь, а также приложения, специфичные для конкретных заболеваний, включая неврологические и психиатрические расстройства. Эти модели демонстрируют способность решать основные вычислительные нейробиологические задачи, включая многомодальную интеграцию нейронных данных, пространственно-временную интерпретацию паттернов и разработку трансляционных структур для клинического применения.

Ключевые слова—искусственный интеллект, неврология, базовые модели, языковые модели.

I. ВВЕДЕНИЕ

Генеративные модели Искусственного Интеллекта быстро привлекли к себе внимание во многих областях, как возможные решения для автоматизации различных процессов. Большие языковые модели (LLM) здесь являются, пожалуй, наиболее ярким примером. Простота их использования, похожая на знакомые всем поисковые системы, генерация результатов (ответов) в виде связного текста и обеспечили быстрый рост их применения (попыток применения) в самых разных областях. Решения на основе LLM в медицине [1], безусловно, обладают потенциалом для трансформации различных процессов, от улучшения клинических решений (дифференциальная диагностика, ответы на вопросы) [2] и автоматизации документирования [3], до ускорения исследований (разработка лекарств, анализ данных) и улучшения процесса образования (цифровые двойники - виртуальные пациенты) [4].

При этом разные области применения могут включать использование мультимодальных LLM, которые обрабатывают, например, текст и изображения [5,6].

Многие медицинские применения LLM включают расширенный поиск (RAG), который нужен для обеспечения фактической точности [7]. Модели, при таком подходе, используются вместе со специализированными базами знаний, что уменьшает ошибки и повышает актуальность и точность ответов. Есть уже и специализированные LLM модели для медицины. Например, Med-PaLM от Google [8], которые представляют собой настройку базовых моделей для медицинского домена. Один из критериев для таких LLM – достижение экспертного уровня при прохождении специализированных медицинских экзаменов, типа USMLE [9].

Конечно, на этом пути есть множество не решенных проблем. Например:

- Модели могут генерировать убедительную, но ложную информацию, создавая риски для здравоохранения.
- Может отсутствовать повторяемость выводов
- Модели могут отражать и усиливать предвзятость, присущую в обучающих данных.
- Важным моментом является конфиденциальность и этика, обеспечение безопасности данных.

В настоящей статье мы хотим остановиться на использовании LLM в неврологии.

Основные области применения в неврологии [10,11]:

Клиническая документация: искусственный интеллект, работающий в фоновом режиме, анализирует разговоры пациентов и врачей, автоматически генерируя структурированные клинические заметки (история болезни, физикальное обследование, анамнез), что значительно экономит время.

Диагностика и рассуждения:

- Помощь в сложных дифференциальных диагнозах и локализации неврологических заболеваний.
- Анализ сигналов ЭЭГ и интеграция различных данных для более быстрой диагностики на экспертном уровне.

Когнитивная оценка: автоматизация оценки результатов нейропсихологических тестов (например, заданий на беглость речи) и создание динамических программ когнитивного тренинга.

Исследования и открытия:

- анализ литературы для прогнозирования новых нейробиологических открытий, иногда превосходящих возможности экспертов.
- выявление пациентов с высоким риском неврологических заболеваний в острых состояниях (например, в отделении неотложной помощи).

Поддержка пациентов/опекунов: чат-боты на базе LLM для ответов на вопросы, управления лечением и оказания эмоциональной поддержки.

Нам неизвестны примеры такого использования (языковые модели в неврологии) на русском языке. Представленные примеры относятся к LLM, обученных, в основном, на англоязычных материалах. Проблема национальных языков, на сегодняшний день, является общей для всех доменов, а не только для медицины. Вместе с тем, применительно к медицине, необходимо отметить, что обучение, во всех странах, ведется именно на национальном языке, на нем же всегда сдаются и различные экзамены. Соответственно, не вызывает сомнения тот факт, что возможные применения LLM в российской неврологии должны базироваться на русском языке. Поэтому, одна из основных целей данной статьи – это выявить источники данных (датасеты), на которых обучались подобные модели, чтобы определить потом их локальные аналоги или возможности локализации.

II. ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ

Для повышения эффективности в клинических и исследовательских условиях было разработано несколько специализированных решений:

Neura [12] - масштабируемое, объяснимое решение, специализирующееся на моделях с линейным моделированием для сложных клинических случаев. В ходе слепых оценок оно показало значительно более высокие результаты, чем неврологи, в дифференциальной и окончательной диагностике. В ходе слепой оценки, проведенной на пяти сложных клинических случаях в сравнении с группой из 13 неврологов, система Neura достигла нормализованных показателей 86,17% в целом, 85% для дифференциальной диагностики и 88,24% для окончательного диагноза (55,11%, 46,15% и 70,93% для неврологов) с быстрым временем ответа 28,8 и 19 секунд (9 минут и 37,2 секунды и 8 минут и 51 секунда для неврологов), при этом последовательно предоставляя актуальную и точно цитируемую информацию.

Neura - это решение, использующее донастроенную (дообученную) LLM GPT-4 Turbo от OpenAI. Корпус знаний по клинической неврологии формировался на основе всеобъемлющих учебников по неврологии [13-16], раздела о неврологических расстройствах руководства компании Merck [17] и Википедии [18].

Neuro-Copilot [19] - усовершенствованная система,

используемая в отделениях неотложной помощи для выявления пациентов высокого риска. Она интегрирует неструктурированные данные электронных медицинских карт для принятия важных решений о госпитализации и выписке.

Авторы отмечают, что в неврологии (или вообще в медицине) приложения ИИ, как правило, концентрируются на конкретных неврологических заболеваниях [20], часто из-за опоры на глубокое обучение, которое требует больших объемов данных для достижения адекватной производительности в конкретной проблемной области. Следовательно, исследования решений ИИ, которые затрагивают множество критических точек принятия решений в отделении неотложной помощи или предлагают всестороннюю поддержку для широкого спектра неврологических состояний, остаются ограниченными. В этой связи предполагается, что ключевые точки принятия решений при ведении неврологических случаев в отделении неотложной помощи могут быть эффективно оптимизированы за счет сочетания больших языковых моделей (LLM) и традиционных моделей машинного обучения (ML). С этой целью и был предложен Neuro-Copilot - ансамблевая структура, разработанная для предоставления точной, контекстно-ориентированной поддержки принятия решений неврологам и группам сортировки в отделении неотложной помощи, помогая в таких критически важных задачах, как принятие решений о госпитализации или выписке пациента.

Авторы системы разработали эту ансамблевую структуру, используя возможности LLM Gemini 1.5-pro-002. Модель была усовершенствована с помощью проектирования подсказок и генерации с расширением поиска (RAG). Прогностическое моделирование было достигнуто путем объединения алгоритма eXtreme Gradient Boosting (XGBoost) и логистической регрессии для оптимальной точности принятия клинических решений.

Модель выполняла две основные функции: (1) предоставление числовых оценок (1–7 для общей вероятности положительного ответа и 1–100 для прогнозирования смертности, где 100 означает 100% вероятность смерти в течение заданного периода времени) и (2) генерация переведенных на английский язык сводных данных с рекомендуемыми дальнейшими действиями. Входные данные включали как структурированные, так и неструктурированные данные, извлеченные из электронных медицинских карт, в том числе, результаты неврологического обследования, историю болезни пациента, рентгенологические данные и результаты лабораторных исследований.

Интересная идея - для повышения надежности LLM запускался пять раз для каждого запроса, при этом итоговый числовой показатель рассчитывался как среднее значение этих запусков, а первый сгенерированный текстовый ответ сохранялся для последующего анализа.

Архитектура системы представлена на рисунке 1.

Вывод LLM использовался далее для логистической

регрессии вместе с табличными данными.

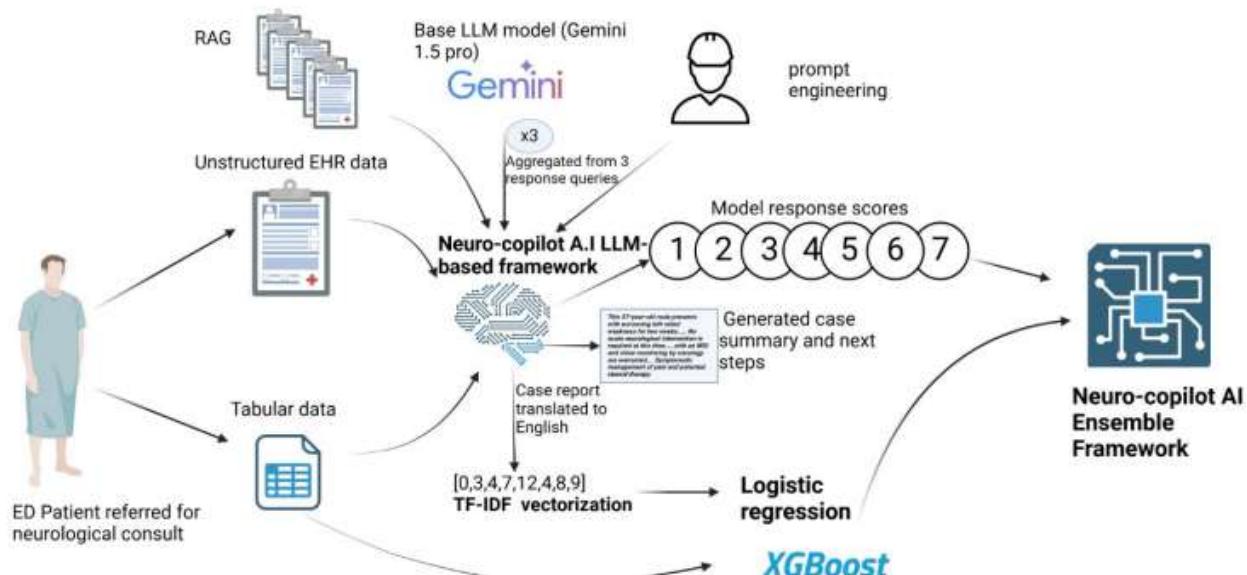


Рис. 1. Архитектура Neuro-Copilot [19].

Были оценены ключевые клинические исходы, такие как госпитализация и смертность. Случайная выборка из 100 случаев была проанализирована тремя опытными неврологами для оценки соответствия прогнозов ИИ экспертной клинической оценке.

Данные для обучения: 1368 последовательных случаев из отделения неотложной помощи. Из них 1118 случаев составили обучающую выборку, а фиксированная контрольная тестовая выборка состояла из 250 случаев «подробной когорты», содержащих более подробные аннотации, которые использовались только для оценки. Клиническая информация извлекалась единообразно с помощью электронной системы поиска записей, способной получить доступ ко всем клиническим и лабораторным результатам. При сборе данных выявили всех пациентов, которые прошли неврологическую консультацию в отделении неотложной помощи с января по февраль 2024 года. Критерии включения включали пациентов старше 18 лет с анамнезом заболевания. Критерии исключения включали пациентов младше 18 лет. Отметим, что все пациенты были из одного и того же медицинского учреждения (это к вопросу о возможном смещении).

Результаты: были ретроспективно проанализированы 1368 последовательных пациентов отделения неотложной помощи, которые прошли неврологическую консультацию, оценивая их клинические особенности, диагностические тесты и исходы госпитализации. Пациенты, поступившие в отделение неотложной помощи, как правило, были старше и имели более высокие показатели смертности, более короткие интервалы до неврологического обследования и более высокую частоту острого инсульта по сравнению с выписанными пациентами. В ходе основного анализа ($n = 250$) нейро-модель искусственного интеллекта (ИИ) продемонстрировала значимые показатели эффективности, достигнув площади под кривой (AUC) 0,88 для прогнозирования госпитализации в целом по

сравнению с фактическими результатами, AUC 0,86 для госпитализации в неврологическое отделение, 0,93 для оценки риска долгосрочной смертности и 1 для оценки риска смертности в течение 48 часов¹. Прогнозы нейромодели ИИ показали сильную корреляцию с консенсусом экспертов (коэффициент корреляции Пирсона 0,79, $p < 0,001$), что указывает на ее способность обеспечивать последовательную поддержку в условиях расходящихся клинических мнений.

Разработанные авторами подсказки (промпты) доступны для изучения².

Объём научной литературы экспоненциально растёт. Одна из проблем для учёных — успевать за развитием событий. Одним из решений является подход, основанный на взаимодействии человека и машины, при котором учёные взаимодействуют с обширной базой знаний нейробиологической литературы, называемой BrainGPT. BrainGPT обучена распознавать закономерности данных в нейробиологической литературе, используя последние достижения в области машинного обучения для моделей больших языков.

BrainGPT функционирует как генеративная модель научной литературы, позволяя исследователям предлагать планы исследований в качестве подсказок, на основе которых BrainGPT будет генерировать вероятные закономерности данных, отражающие текущий анализ научной литературы. Моделисты могут использовать BrainGPT для оценки своих моделей на основе общего понимания предметной области (например, мгновенный метаанализ). BrainGPT может

¹AUC показывает, насколько хорошо модель отличает положительные примеры от отрицательных: значение от 0 до 1, где 1 - идеальная модель, 0.5 - случайное угадывание, а 0 - наихудшая модель. AUC - это вероятность, что модель присвоит более высокий балл случайно выбранному положительному примеру, чем отрицательному. Чем ближе к 1, тем лучше модель разделяет классы; значения 0.7-0.8 считаются приемлемыми, а >0.9 - выдающимися.

² <https://www.mdpi.com/article/10.3390/jcm1417633/s1>

помочь выявить аномальные результаты, будь то указывающие на прорыв или содержащие ошибку.

Важно отметить, что BrainGPT не резюмирует статьи и не осуществляет поиск статей. В таких случаях модели, использующие обширные языковые средства, часто прибегают к конфабуляции, что потенциально вредно. Вместо этого BrainGPT объединяет существующие знания, слишком обширные для человеческого понимания, чтобы помочь людям расширять научные горизонты.

В своей статье в Nature, коллектив авторов, возглавляемый профессором Xiaoliang Luo³, отмечает, что вопрос, могут ли большие языковые модели, обученные на общем тексте и научных статьях, предсказывать результаты экспериментов, остается открытым. Если бы предсказания LLM превосходили предсказания экспертов-людей, научная практика и темпы открытий радикально изменились бы. Авторы рассматривают этот вопрос применительно к нейробиологии, которая является обширной и междисциплинарной областью. Отмечается, что предсказание в нейробиологии должно представлять собой сложную задачу для экспертов-людей по нескольким причинам:

- (1) часто существуют тысячи релевантных научных статей,
- (2) отдельное исследование может быть шумным или ненадежным и может не воспроизводиться,
- (3) нейробиология — это многоуровневая область исследований, охватывающая поведение и молекулярные механизмы,
- (4) методы анализа разнообразны и могут быть сложными,
- (5) как и используемые методы, которые включают различные методы визуализации мозга, исследования поражений, генную модификацию, фармакологические вмешательства и так далее.

Большинство медицинских бенчмарков оценивают способность LLM к извлечению базовых знаний и рассуждениям, которые, как правило, ориентированы на прошлое. Эти бенчмарки структурированы в формате вопросов и ответов, где модели должны продемонстрировать обширные знания о мире, извлечь соответствующую информацию на основе контекста вопроса и дать правильный ответ. Однако ни один из этих бенчмарков не подходит для оценки способности моделей прогнозировать новые результаты, что по своей сути ориентировано на будущее.

Ориентация на прошлое — это извлечение (вспоминание) фактической информации. Например, студент вспоминает о признаках какого-либо заболевания. Существующие критерии оценки в научных областях по сути являются оглядывающимися назад, поскольку они делают упор на извлечение общепринятых фактов для решения задач, требующих

ответа на вопросы и рассуждений. Критерии оценки, ориентированные на будущее, предполагают прогнозирование новых результатов на основе прошлых данных. Могут присутствовать две формы неопределенности: алеаторическая (из-за присущей случайности) и эпистемическая (из-за недостатка знаний). Например, болельщик предсказывает, какая команда выиграет, основываясь на своих знаниях об играх, их игре до настоящего момента и так далее. Также будут присутствовать присущие случайные факторы, такие как погода, ошибки судей и т.д.

В работе [22] и представлен разработанный подход для проверки способности LLM прогнозировать результаты нейробиологических исследований, получивший название BrainBench (рис. 2). LLM прошли обширное обучение на научной литературе, включая нейробиологию. BrainBench оценивает, насколько LLM уловили фундаментальные закономерности методов и результатов, лежащие в основе структуры нейробиологии. Могут ли LLM превзойти экспертов-людей в этом перспективном teste? В частности, BrainBench оценивает, насколько хорошо испытуемый может прогнозировать результаты нейробиологических исследований на основе методов, представляя две версии аннотации из недавней журнальной статьи. Задача испытуемого — предсказать результат исследования, выбирая между оригинальной и измененной версией. Измененная аннотация существенно меняет результат исследования, сохраняя при этом общую согласованность.

Тестовые примеры BrainBench были взяты из недавних аннотаций к статьям в Journal of Neuroscience по пяти областям нейронауки: поведенческая/когнитивная, системная/схемная, нейробиология заболеваний, клеточная/молекулярная и эмбриональная/пластиность/восстановление.

Участники тестирования выбирали между оригинальной аннотацией и измененной версией, существенно изменившей результат при сохранении согласованности. Экспертам и моделям LLM было поручено выбрать правильный (то есть оригинальный) вариант из двух предложенных. Эксперты делали выбор и оценивали свою уверенность и компетентность в онлайн-опросе. Модели LLM оценивались как выбравшие аннотацию с меньшей степенью сложности (то есть, текстовый фрагмент, который был менее неожиданным для модели), а их уверенность была пропорциональна разнице в степени сложности между двумя вариантами.

То, что является галлюцинацией в задачах, ориентированных на прошлое [23], является обобщением или предсказанием в задачах, ориентированных на будущее. BrainBench позволяет количественно оценить эту способность к прогнозированию и сравнить результаты с результатами экспертов.

³ Department of Experimental Psychology, University College London, London, UK

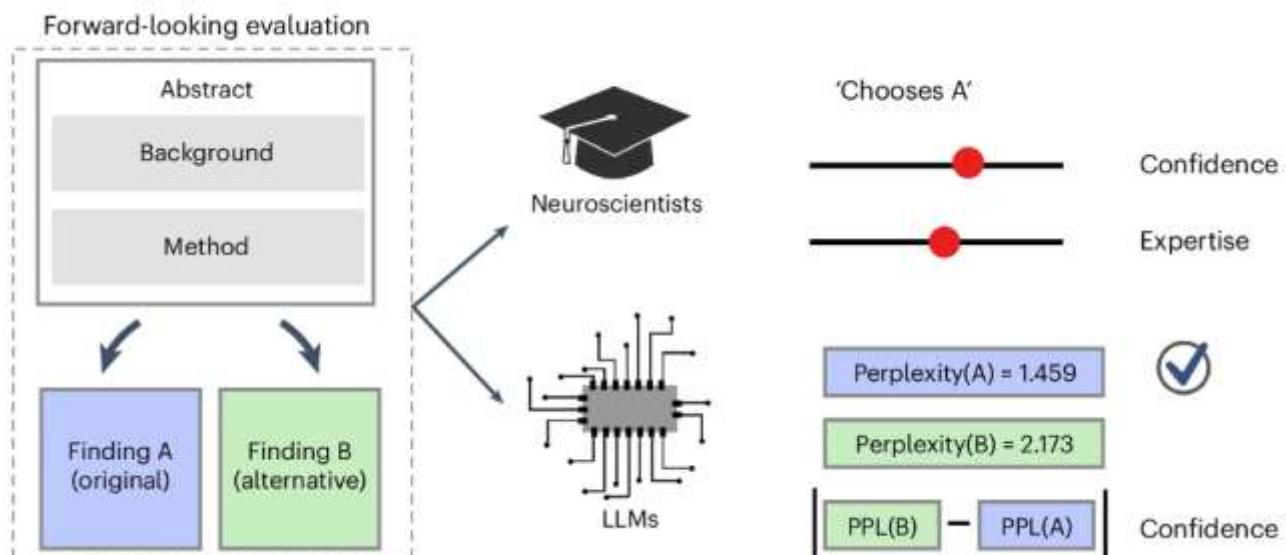


Рис. 2. BrainBench [22]

В такой постановке, LLM значительно превзошли экспертов в BrainBench, и это преимущество увеличилось, когда было проведено дополнительное обучение базовой LLM в области нейронаук, которое авторы назвали «BrainGPT».

Предварительно обученные LLM могут служить основой для дальнейшего обучения в области нейронауки с целью повышения производительности, оцениваемой с помощью BrainBench. В данном примере авторы использовали дообучение по методу LoRA [24] для базовой LLM, Mistral-7B-v0.1. LoRA эффективно расширяет возможности универсальных линейных моделей обучения за счет введения в существующую модель обучаемых параметров низкого ранга (называемых «адаптерами»). Этот процесс позволяет эффективно дорабатывать модель для решения последующих задач без необходимости чрезмерно ресурсоемкого обучения всей модели. В данном случае авторы выполнили тонкую настройку Mistral-7B-v0.1, используя более 1,3 миллиарда токенов из публикаций по нейронауке, охватывающих 100 журналов в период с 2002 по 2022 год (332,807 аннотаций и 123,085 полных статей), что улучшило производительность BrainBench на 3%.

Поскольку тестовый датасет, в данном случае, это оригинальные и модифицированные статьи, то для создания тестовых примеров, модифицировались (с помощью экспертов и/или LLM ChatGPT) опубликованные работы. Модифицированная версия существенно изменила результаты, не затрагивая методы и контекст. Были внесены минимальные изменения, которые повлияли на основной результат. Например, модифицированный вариант по сравнению с оригиналом мог поменять местами роли двух областей мозга в результатах, изменить направление результата (например, заменить «уменьшается» на «увеличивается») и так далее. Любые изменения сохраняли согласованность реферата, что иногда требовало нескольких изменений (например, замена

нескольких «уменьшается» на «увеличивается»). Другими словами, модифицированные рефераты должны были быть эмпирически разными, но не логически противоречивыми. И добровольцам, и GPT-4 были даны инструкции, соответствующие указанным выше основным критериям. В статье приводится точная подсказка (промпт) для модификации.

Весь код, связанный с данной работой, доступен на GitHub⁴. Веса и тренировочные данные доступны на HuggingFace⁵.

На рисунке 3 [25] представлены 5 основных областей применения генеративных моделей в неврологии. Эти области могут быть систематически сгруппированы в

- нейровизуализацию и обработку данных,
- интерфейсы мозг-компьютер
- геномное моделирование
- клинические приложения
- другие расстройства головного мозга

Эти приложения используют способность крупномасштабных моделей ИИ расшифровывать сложные нейронные паттерны и выявлять лежащие в основе законы организации и функционирования мозга.

Каждая группа использует различные архитектуры и методы, начиная от предварительного обучения с использованием базовых моделей на основе мультимодальных нейронных данных до тонкой настройки с помощью интерпретируемых фреймворков и решения конкретных задач нейробиологических исследований и клинической практики. Эти методы создают вычислительную основу для понимания нейронных механизмов и позволяют применять их в диагностике, терапии и нейротехнологиях в лабораторных и клинических приложениях.

Этот свежий (октябрь 2025) обзор содержит 55+ примеров использования LLM в неврологии.

⁴ <https://github.com/braingpt-lovelab/BrainBench>

⁵ <https://huggingface.co/BrainGPT>

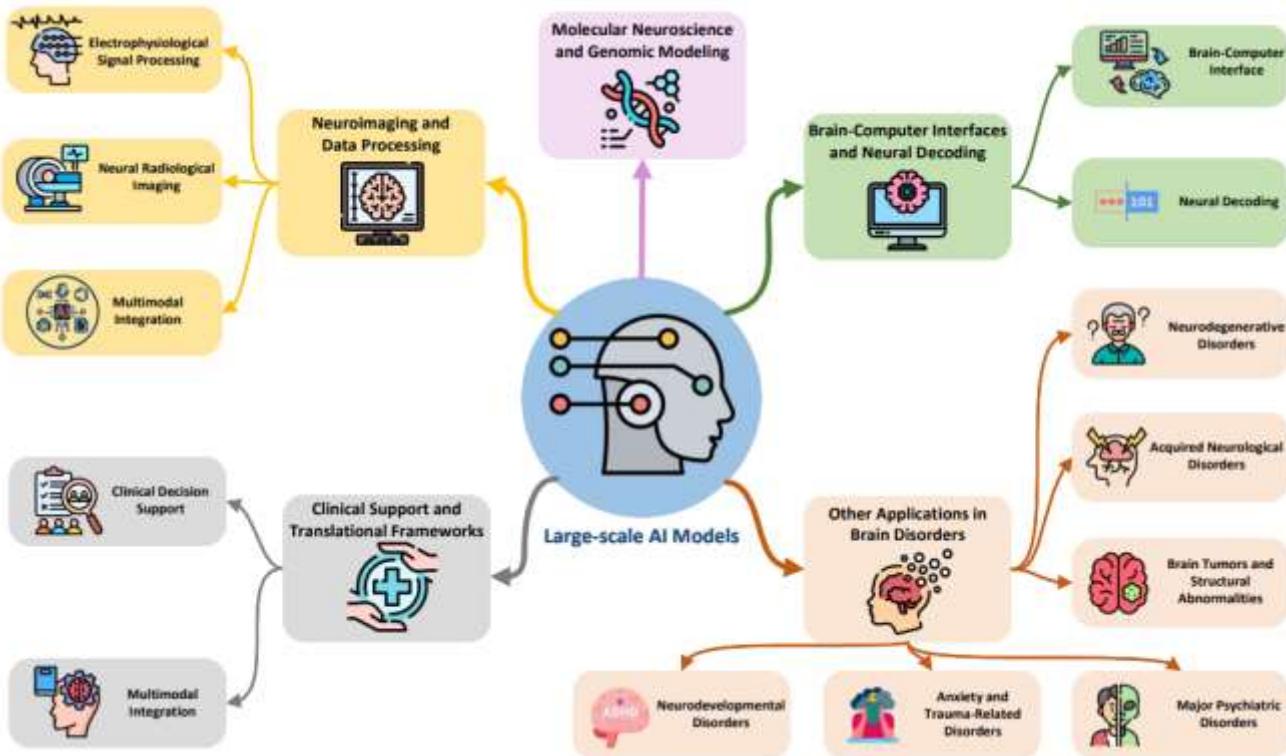


Рис.3 Базовые модели в неврологии [25].

Если остановиться на используемых архитектурах, то, практически везде, это тонкая настройка базовых моделей. В трех случаях использовался подход с RAG и дважды – интеграция графа знаний.

III ДАТАСЕТЫ

Если говорить об использованных датасетах, то по обзору [25] можно сделать следующие заключения. Используются:

- Публичные датасеты. Например, коллекция открытых датасетов по нейронаукам⁶. Их довольно много, например, по электроэнцефалограммам⁷ или MRI⁸.
- Медицинские карты (истории болезней)
- Учебники по неврологии
- Образовательные ресурсы по неврологии. Ранее упоминался, например, справочник Merck
- Литература по неврологии. PubMed
- Нейрохирургический атлас [26]
- Онтологии нейронаук⁹
- Описания клинических испытаний [28]

Здесь нужно отметить, что исследователи в Российской Федерации имеют хорошую экспертизу в области анализа медицинских изображений. Сервис «КТ головного мозга»¹⁰ на базе искусственного интеллекта ИИ, разработанный совместно Лабораторией искусственного интеллекта Сбербанка и СберМедИИ,

стал первым алгоритмом по направлению «Ишемический инсульт» в московском эксперименте по применению компьютерного зрения в лучевой диагностике. В рамках этого же эксперимента был создан и открытый датасет по компьютерной томографии головного мозга¹¹.

V ЗАКЛЮЧЕНИЕ

Специализированные модели, такие как Neura, показали высокие результаты в анализе сложных случаев, даже превзойдя группы неврологов по скорости и точности выполнения определенных задач.

По-прежнему, существуют опасения по поводу надежности, присущей моделям предвзятости, потенциального усугубления неравенства в здравоохранении и ограниченного истинного клинического мышления (галлюцинации).

Гибридные подходы, сочетающие мощные общие языковые модели машинного обучения с тщательно отобранными медицинскими базами данных (генерация с расширенным поиском - RAG), оказываются более эффективными, чем автономные модели, с точки зрения фактической точности в клинических условиях. В части наличия исходных данных для дообучения (настройки) базовых моделей используются размеченные медицинские карты, а также научные публикации. Такие источники доступны на национальных языках, что, в принципе, делает возможным построение LLM для неврологии, работающей с русским языком.

Отмечается, что для успешной интеграции крайне важно создание культуры, благоприятствующей ИИ, четкие цели и взаимодействие с заинтересованными сторонами. При этом LLM рассматриваются как дополнение, а не замена человеческого опыта, что требует контроля и интеграции в рабочие процессы.

⁶ <https://github.com/openlists/ElectrophysiologyData>

⁷ <https://github.com/meagmohit/EEG-Datasets>

⁸ <https://openneuro.org/>

⁹ <https://scicrunch.org/scicrunch/interlex/dashboard>

¹⁰ <https://sbermed.ai/our-algorithms/ct-stroke>

¹¹ <https://www.m24.ru/news/medicina/16022024/666686>

БЛАГОДАРНОСТИ

Хотелось бы поблагодарить сотрудников кафедры Информационной безопасности факультета БМК за плодотворные дискуссии и обсуждения.

БИБЛИОГРАФИЯ

- [1] Thirunavukarasu, Arun James, et al. "Large language models in medicine." *Nature medicine* 29.8 (2023): 1930-1940.
- [2] Shool, Sina, et al. "A systematic review of large language model (LLM) evaluations in clinical medicine." *BMC Medical Informatics and Decision Making* 25.1 (2025): 117.
- [3] Goyal, Sagar, et al. "Healai: A healthcare llm for effective medical documentation." *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 2024.
- [4] Lucas, Harrison C., Jeffrey S. Upperman, and Jamie R. Robinson. "A systematic review of large language models and their implications in medical education." *Medical education* 58.11 (2024): 1276-1285.
- [5] AlSaad, Rawan, et al. "Multimodal large language models in health care: applications, challenges, and future outlook." *Journal of medical Internet research* 26 (2024): e59505.
- [6] Nam, Yoojin, et al. "Multimodal large language models in medical imaging: current state and future directions." *Korean Journal of Radiology* 26.10 (2025): 900.
- [7] Ng, Karen Ka Yan, Izuki Matsuba, and Peter Chengming Zhang. "RAG in health care: a novel framework for improving communication and decision-making by addressing LLM limitations." *Nejm Ai* 2.1 (2025): Alra2400380.
- [8] Singhal, Karan, et al. "Toward expert-level medical question answering with large language models." *Nature Medicine* 31.3 (2025): 943-950.
- [9] Shaikh, Yahya, et al. "Collaborative intelligence in AI: Evaluating the performance of a council of AIs on the USMLE." *PLOS Digital Health* 4.10 (2025): e0000787.
- [10] Moura, Lidia, et al. "Implications of large language models for quality and efficiency of neurologic care: emerging issues in neurology." *Neurology* 102.11 (2024): e209497.
- [11] Shah, Rushabh, and Fabrice Jotterand. "Large Language Models in Neurology Treatment Decision-Making: a Scoping Review." *Journal of Medical Systems* 49.1 (2025): 1-8.
- [12] Barrit, Sami, et al. "Neura: a specialized large language model solution in neurology." *medRxiv* (2024): 2024-02.
- [13] Ropper, A. H., M. A. Samuels, and J. P. Klein. "Adams and Victor's Principles of Neurology." 2014.
- [14] Jankovic, Joseph, John C. Mazzotta, and Scott L. Pomeroy. Bradley and Daroff's Neurology in Clinical Practice-E-Book. Elsevier Health Sciences, 2021.
- [15] Campbell, William Wesley, and Russell N. DeJong. DeJong's the neurologic examination. No. 2005. Lippincott Williams & Wilkins, 2005.
- [16] Merritt, Hiram Houston. Merritt's neurology. Lippincott Williams & Wilkins, 2010.
- [17] MSD [Neurology](https://www.msdmanuals.com/professional/searchresults?query=neurology) https://www.msdmanuals.com/professional/searchresults?query=neurology Dec 2025
- [18] Wikipedia https://en.wikipedia.org/wiki/Category:Neurological_disorders Retrieved: Dec, 2025
- [19] Gorenstein, Alon, et al. "Neuro-Copilot AI: Advanced LLM Framework for Neurological Patients in the Emergency Room (S27.004)." *Neurology*. Vol. 104. No. 7_Supplement_1. Hagerstown, MD: Lippincott Williams & Wilkins, 2025.
- [20] Voigtlaender, Sebastian, et al. "Artificial intelligence in neurology: opportunities, challenges, and policy implications." *Journal of neurology* 271.5 (2024): 2258-2273.
- [21] BrainGPT <https://braingpt.org/> Retrieved: Dec, 2025
- [22] Luo, Xiaoliang, et al. "Large language models surpass human experts in predicting neuroscience results." *Nature human behaviour* 9.2 (2025): 305-315.
- [23] Kim, Yubin, et al. "Medical hallucinations in foundation models and their impact on healthcare." *arXiv preprint arXiv:2503.05777* (2025).
- [24] Le, Thanh-Dung, et al. "The Impact of LoRA Adapters on LLMs for Clinical Text Classification Under Computational and Data Constraints." *IEEE Access* (2025).
- [25] Yang, Shihao, et al. "Foundation and Large-Scale AI Models in Neuroscience: A Comprehensive Review." *arXiv preprint arXiv:2510.16658* (2025).
- [26] Z. E. Teton, R. S. Freedman, S. B. Tomlinson, J. R. Linzey, A. Onyewuenyi, A. S. Khahera, "The neurosurgical atlas: advancing neurosurgical education in the digital age," *Neurosurgical focus*, vol. 48, no. 3, p. E17, 2020
- [27] B. K. Hendricks, and A. A. Cohen-Gadol, "The neurosurgical atlas: advancing neurosurgical education in the digital age," *Neurosurgical focus*, vol. 48, no. 3, p. E17, 2020
- [28] M. Jullien, M. Valentino, H. Frost, P. O'Regan, D. Landers, and A. Freitas, "Nli4ct: Multievidence natural language inference for clinical trial reports," *arXiv preprint arXiv:2305.03598*, 2023.

Статья получена 29 октября 2025.

Е.Д. Намиот – ординатор ФФМ МГУ имени М.В. Ломоносова (e-mail: enamiot@gmail.com).

Large Language Models in Neuroscience

Eugenia Namiot

Abstract - The development of artificial intelligence (AI) systems, in particular, so-called foundation models and large language models, has opened a new era at the intersection of AI and neuroscience. These models allow working with diverse datasets across different modalities.

Compared to classical computational approaches, which mainly relied on traditional machine learning methods, these models represent a significant step forward. They demonstrate strong generalization capabilities and can capture complex spatio-temporal dependencies found in the data. This is achieved, in particular, through end-to-end learning directly on raw data. Foundation models can potentially be applied in all major neurobiological fields, encompassing neuroimaging and data processing, brain-computer interfaces and neural decoding, molecular neurobiology and genomic modeling, clinical care, and disease-specific applications, including neurological and psychiatric disorders. These models demonstrate the ability to solve fundamental computational neuroscience problems, including multimodal integration of neural data, spatio-temporal pattern interpretation, and the development of translational frameworks for clinical applications.

Keywords— artificial intelligence, neuroscience, foundation models, language models.

REFERENCES

- [1] Thirunavukarasu, Arun James, et al. "Large language models in medicine." *Nature medicine* 29.8 (2023): 1930-1940.
- [2] Shool, Sina, et al. "A systematic review of large language model (LLM) evaluations in clinical medicine." *BMC Medical Informatics and Decision Making* 25.1 (2025): 117.
- [3] Goyal, Sagar, et al. "Healai: A healthcare llm for effective medical documentation." *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 2024.
- [4] Lucas, Harrison C., Jeffrey S. Upperman, and Jamie R. Robinson. "A systematic review of large language models and their implications in medical education." *Medical education* 58.11 (2024): 1276-1285.
- [5] AlSaad, Rawan, et al. "Multimodal large language models in health care: applications, challenges, and future outlook." *Journal of medical Internet research* 26 (2024): e59505.
- [6] Nam, Yoojin, et al. "Multimodal large language models in medical imaging: current state and future directions." *Korean Journal of Radiology* 26.10 (2025): 900.
- [7] Ng, Karen Ka Yan, Izuki Matsuba, and Peter Chengming Zhang. "RAG in health care: a novel framework for improving communication and decision-making by addressing LLM limitations." *Nejm Ai* 2.1 (2025): A1ra2400380.
- [8] Singhal, Karan, et al. "Toward expert-level medical question answering with large language models." *Nature Medicine* 31.3 (2025): 943-950.
- [9] Shaikh, Yahya, et al. "Collaborative intelligence in AI: Evaluating the performance of a council of AIs on the USMLE." *PLOS Digital Health* 4.10 (2025): e0000787.
- [10] Moura, Lidia, et al. "Implications of large language models for quality and efficiency of neurologic care: emerging issues in neurology." *Neurology* 102.11 (2024): e209497.
- [11] Shah, Rushabh, and Fabrice Jotterand. "Large Language Models in Neurology Treatment Decision-Making: a Scoping Review." *Journal of Medical Systems* 49.1 (2025): 1-8.
- [12] Barrit, Sami, et al. "Neura: a specialized large language model solution in neurology." *medRxiv* (2024): 2024-02.
- [13] Ropper, A. H., M. A. Samuels, and J. P. Klein. "Adams and Victor's Principles of Neurology. 2014
- [14] Jankovic, Joseph, John C. Mazziotta, and Scott L. Pomeroy. Bradley and Daroff's Neurology in Clinical Practice-E-Book. Elsevier Health Sciences, 2021.
- [15] Campbell, William Wesley, and Russell N. DeJong. *DeJong's the neurologic examination*. No. 2005. Lippincott Williams & Wilkins, 2005.
- [16] Merritt, Hiram Houston. *Merritt's neurology*. Lippincott Williams & Wilkins, 2010.
- [17] MSD <https://www.msdmanuals.com/professional/searchresults?query=neurology> Dec 2025
- [18] Wikipedia https://en.wikipedia.org/wiki/Category:Neurological_disorders Retrieved: Dec, 2025
- [19] Gorenstein, Alon, et al. "Neuro-Copilot AI: Advanced LLM Framework for Neurological Patients in the Emergency Room (S27.004)." *Neurology*. Vol. 104. No. 7_Supplement_1. Hagerstown, MD: Lippincott Williams & Wilkins, 2025.
- [20] Voigtlaender, Sebastian, et al. "Artificial intelligence in neurology: opportunities, challenges, and policy implications." *Journal of neurology* 271.5 (2024): 2258-2273.
- [21] BrainGPT <https://braingpt.org/> Retrieved: Dec, 2025
- [22] Luo, Xiaoliang, et al. "Large language models surpass human experts in predicting neuroscience results." *Nature human behaviour* 9.2 (2025): 305-315.
- [23] Kim, Yubin, et al. "Medical hallucinations in foundation models and their impact on healthcare." *arXiv preprint arXiv:2503.05777* (2025).
- [24] Le, Thanh-Dung, et al. "The Impact of LoRA Adapters on LLMs for Clinical Text Classification Under Computational and Data Constraints." *IEEE Access* (2025).
- [25] Yang, Shihao, et al. "Foundation and Large-Scale AI Models in Neuroscience: A Comprehensive Review." *arXiv preprint arXiv:2510.16658* (2025).
- [26] Z. E. Teton, R. S. Freedman, S. B. Tomlinson, J. R. Linzey, A. Onyewuenyi, A. S. Khahera,
- [27] B. K. Hendricks, and A. A. Cohen-Gadol, "The neurosurgical atlas: advancing neurosurgical education in the digital age," *Neurosurgical focus*, vol. 48, no. 3, p. E17, 2020
- [28] M. Jullien, M. Valentino, H. Frost, P. O'Regan, D. Landers, and A. Freitas, "Nli4ct: Multievidence natural language inference for clinical trial reports," *arXiv preprint arXiv:2305.03598*, 2023.