

# Семинар Лаборатории КЛПВЭ

27.06.2025

## “Трансформеры в Науке и Жизни”

### ИИ врывается в нашу жизнь...

**История 1.** Я очень хорошо помню, как ровно 10 лет назад мы с одним моим подопечным, студентом Физфака МГУ, занимаясь сбором и анализом данных о тестировании спортсменов, в перерывах болтали о модных тогда словечках БигДата, ДипЛёнинг и т.п. Как сейчас помню, Миша (так звали студента) с увлечением рассказывал о том, что у Гугла есть возможность теперь в строке написать слово “кошечка” и в ответ получить кучу фотографий, рисунков и т.п. с кошечками. Тогда нам это казалось удивительным достижением – шёл 2015 год... Сейчас Михаил – успешный бизнесмен, поселившийся в Дубае, и карьеру сделал, в том числе, и с привлечением ИИ к своим инициативным проектам.

**История 2.** Есть у меня приятель-одноклассник – в школе за одной партой сидели. В 21-м веке он стал руководителем одного из крупнейшего в Беларуси производителя программного обеспечения – правда, в основном для Европейского рынка. Навещая пару раз в году своих родителей, я всегда по пути обратно заезжал к другу на дачу, где за бокалом само-сваренного пива и в тепле сауны мы вели долгие беседы о том, как меняется мир вокруг нас. Он хвастался, что его младший сын настолько быстро схватывает все, что связано с современными технологиями, что ему самому, делающему карьеру в этом направлении, никогда за ним не угнаться. И очень радовался этому, предвкушая, в том числе, и какой уникальный мир откроется перед новым поколением уже в ближайшие года. В принципе, он оказался прав. Я, в общих чертах соглашаясь с таким видением грядущих перемен, чаще выступал в качестве оппонента – отстаивал права на жизнь “старых поколений” и подвергал сомнениям “эффективность ИИ”. Кстати, сам друг и его семья, также как и фирма, – все переехали в Германию и больше мне не с кем вести эти чрезвычайно увлекательные беседы.

**История 3.** Есть у меня ещё один приятель, уже здесь, в Москве. Точнее, время от времени он срывается куда-нибудь типа Бали, Кении или ещё в какие-то отдалённые места, но обилие инициативных проектов заставляет его возвращаться снова. Сейчас он в той или иной степени курирует развитие научно-технологической и образовательной структуры московского “Спартака”. В принципе, у него все проекты такого масштаба – с размахом. Причём чаще всего он занимается тем, что презентует топ-менеджменту и всяким высокопоставленным чиновникам какие-то свои идеи – инновационные, коммерческие, архитектурные, технологические и т.п. Так вот, год назад он мне говорит: «Ты не представляешь, насколько мне стало легче работать после того, как я в качестве помощника начал использовать ИИ. Раньше, чтобы подготовить качественную презентацию мне нужен был целый (мини)штат сотрудников (а следовательно и офис), с каждым из которых надо было провести кучу совещаний, планёрок, просто бесед и т.п. А теперь я практически все делаю в одиночку – только с ИИ. И время создание презентации сократилось в разы! Не говоря уж о финансовых затратах. А качество и полнота этих презентаций только выросли...».

**История 4.** Сели мы как-то в лаборатории на диванчике пить чай-кофе и разговорились об ИИ. Я в этот момент как раз для проекта «ЭРА» завершил предварительный расчёт оптической системы – телескопа Шмидта. И для своих коллег в проекте решил от руки нарисовать эскиз этой системы – прямо на бумажке. И говорю ребятам – Данилу и Вите: А пусть ваш ИИ по этой бумажке мне код напишет, чтобы мы могли получить этот детектор уже сразу в САД-формате или аналоге. И вот... И вот мы сегодня все собрались здесь...

## ИИ в образовании

### Обучение и искусственный интеллект

#### *Университет штата Огайо (США)*

Достижения в области создания инструментов искусственного интеллекта, таких как Copilot и ChatGPT, быстро трансформируют преподавание и обучение. Штат Огайо лидирует со смелой инициативой по внедрению ИИ в учебную программу. Инициатива AI Fluency <https://oaa.osu.edu/ai-fluency>, которая будет запущена осенью 2025 года для первокурсников, гарантирует, что каждый выпускник университета штата Огайо в 2029 году будет свободно владеть ИИ.

### **Особенности преподавания и обучения для Gen AI: Командный подход**

Как эффективно и с соблюдением этических норм использовать инструменты ИИ – один из многих ключевых вопросов, стоящих перед преподавателями. Учитывая быстро развивающийся характер генеративного ИИ, ни один офис не может решить все связанные с этим образовательные задачи. К счастью, различные подразделения университета используют совместный подход, что в значительной степени способствует получению рекомендаций и предоставлению ресурсов. Перечисленные ниже ресурсы предоставлены разнообразной группой сотрудников университета, объединившихся, чтобы предложить как централизованную, так и распределенную поддержку тем преподавателям, которые ориентируются на влияние ИИ на преподавание и обучение.

- 1) *Институт изучения и преподавания им. Майкла В. Дрейка* расширит возможности программирования и руководства по применению ИИ в обучении. Преподаватели могут посетить семинары по разработке курсов на основе ИИ, интеграции ИИ в задания и использованию GenAI или запросить консультацию, чтобы обсудить возникающие проблемы и получить рекомендации по подходам к обучению, основанным на фактических данных (evidence-informed approaches to teaching).
- 2) *Центр изучения и преподавания письменной речи* работает с преподавателями над решением сложных интеллектуальных задач чтения, устной речи и письма в контексте генеративного ИИ, которые необходимы для участия в академической, профессиональной и общественной сферах.
- 3) *Отдел дистанционного обучения Колледжа искусств и наук* предоставляет рекомендации по написанию правил поведения, которыми вы можете поделиться со своими студентами, чтобы проинформировать их о том, что им разрешено, а что нет, при использовании инструментов ИИ для выполнения работы по вашему курсу.
- 4) *Учебно-методический отдел университетских библиотек* может оказать поддержку преподавателям в изучении связей между генеративным ИИ, информационной грамотностью и исследованиями, включая рассмотрение того, как студенты могут использовать инструменты ИИ в рамках процесса исследований и поиска информации. Университетские библиотеки могут предоставить преподавателям рекомендации по использованию общих инструментов и ресурсов в процессе преподавания, разработки курсов или заданий.
- 5) *Центр учебно-методических ресурсов (TLRC)* – это виртуальный центр ресурсов для преподавателей штата Огайо. Ниже приведены основные ресурсы TLRC, предлагающие рекомендации по обучению с использованием ИИ:...

## ...и в науке

*История 5.* Лет этак 10 назад я натолкнулся на книгу с названием что-то вроде «Профессии ближайшего будущего». Книга была сделана в формате классно иллюстрированного альбома, и появление многих профессий действительно напрашивалось, какие-то были слегка притянуты за уши и скорее заполняли конъюнктурные ниши автора. Но, уверен, что автор и представить себе не мог, как могут поменяться набор выбранных им профессий с учётом последних достижений в ИИ, в особенности больших языковых моделей (LLM) – теме, которой будет посвящена практическая часть нашего сегодняшнего семинара. В контексте этого мне почему-то вспомнилась совсем другая профессия – из мира прошлого. Профессию, которую во всей красе застал Михаил Зотов и я, и с которой вряд ли кто-то ещё из присутствующих здесь встречался. В 80-е был такой человек Телемастер. Его вызывали, если что-то переставало работать в одном из главных достояний советской семьи – Телевизоре. Телевизоры тогда считались чуть ли не живым существом и фактически приравнивались члену семьи. С «телевизионным доктором» частенько встречались дети, так как его вызывали в рабочее время, когда родители были на работе. Приходил этакий дядька с усами и с ящичком инструментов, откручивал заднюю крышку пузатого телевизора и что-то там крутил, винтил, менял. После этого обычно телевизор начинал показывать что-то удобоваримое. А если нет, то вечером приходил глава семейства и показывал как «донастроить говорящую коробку», ударив его аккуратнo сбоку... Сейчас телевизор – «чёрный ящик», залезть внутрь которого не возникнет мысль ни у одного здравомыслящего человека. И я не сомневаюсь, что такая же участь в скором времени ждёт и LLM-специалистов. Это сейчас они ещё стараются «заглянуть под капот нейросетки», выяснить как она работает, усовершенствовать её. Лет через пять вам будут впаривать продвинутые версии, которые справляются с той или иной задачей лучше других.

### **AI Agents for Ground-Based Gamma Astronomy**

*Dmitriy Kostunin, Vladimir Sotnikov, Sergo Golovachev, and Alexandre Strube (Germany)*

<https://arxiv.org/abs/2503.00821>

Demo: <https://majestix-vm8.zeuthen.desy.de/>

Детекторы гамма-астрономии нового поколения значительно усложняются, поскольку в их состав входят десятки телескопов. Такой масштабный скачок создаёт значительные трудности в управлении работой систем и оффлайн-анализе данных. Методы, зависящие от продвинутой подготовки персонала и сложного программного обеспечения, становятся все более трудными по мере роста сложности системы, и бросают новые вызовы их пользователям. Для решения этих задач мы предлагаем разработку агентов искусственного интеллекта на основе больших языковых моделей (LLM) с точной настройкой инструкций. Эти агенты соответствуют конкретной документации и базам кода, понимают контекст окружающей среды, работают с внешними API и общаются с людьми на естественном языке. Используя передовые возможности современных LLM, которые могут обрабатывать и сохранять огромные объёмы информации, эти ИИ-агенты предлагают преобразующий подход к управлению системой и анализу данных, автоматизируя сложные задачи и предоставляя интеллектуальную помощь. Мы представляем два прототипа, которые интегрируются с конвейерами обсерватории СТА для работы и анализа данных. Первый прототип автоматизирует внедрение модели данных и обслуживание базы данных конфигурации системы управления массивом и сбора данных (ACADA). Второй прототип представляет собой приложение для генерации кода с открытым доступом, предназначенное для анализа данных на основе платформы Gammaru framework.

## LLM vs. МАТЕМАТИКА

### Конкурс “Кто обыграет o4-mini?!”

Сайт конкурса: <https://frontiermath-symposium.epoch.ai/>

Статья про бенчмарк frontiermath: <https://arxiv.org/abs/2411.04872>

В середине мая был созван математический конклав. Тридцать известных математиков с разных частей планеты приехали в Беркли. Каждая задача, которую не мог решить **o4-mini**, приносила математику, который её придумал, вознаграждение в размере 7500 долларов.

В течение двух дней математики задавали боту сложные математические задачи. Оказалось, что бот способен ответить на некоторые из самых сложных решаемых задач в мире.

«У меня есть коллеги, которые буквально сказали, что эти модели приближаются к математическому гению», – рассказал Кен Оно, математик из Университета Вирджинии, руководитель турнира.

Рассматриваемый чат-бот работает на основе o4-mini, так называемой большой языковой модели (LLM). Он был обучен OpenAI делать очень сложные выводы. Аналог Google, Gemini 2.5 Flash, обладает схожими способностями. Как и LLM, которые работали в более ранних версиях ChatGPT, o4-mini учится предсказывать следующее слово в последовательности. Однако по сравнению с этими более ранними LLM, o4-mini и его эквиваленты являются более легкими, более гибкими моделями, которые обучаются на специализированных наборах данных с более сильным подкреплением от людей. Такой подход приводит к чат-боту, способному гораздо глубже погрузиться в сложные проблемы математики, чем традиционные LLM.

Чтобы отслеживать прогресс o4-mini, OpenAI ранее поручил Epoch AI, некоммерческой организации, которая проводит сравнительный анализ LLM, придумать 300 математических вопросов, решения которых ещё не были опубликованы. Даже традиционные LLM могут правильно ответить на многие сложные математические вопросы. Тем не менее, когда Epoch AI задал нескольким таким моделям эти вопросы, которые были непохожи на те, на которых они обучались, наиболее успешные смогли решить менее 2%, что показало, что этим LLM не хватает способности рассуждать. Но o4-mini оказался совсем другим.

Epoch AI наняла Эллиота Глейзера, недавно получившего докторскую степень по математике, для разработки нового бенчмарка, получившего название FrontierMath, в сентябре 2024 года. Проект собрал новые вопросы по разным уровням сложности, причем первые три уровня охватывали задачи уровня бакалавриата, магистратуры и исследовательского уровня.

К апрелю 2025 года Глейзер обнаружил, что o4-mini может решить около 20% вопросов. Затем он перешёл к четвёртому уровню: набору вопросов, которые были бы сложными даже для академического математика. Только небольшая группа людей в мире была бы способна разработать такие вопросы, не говоря уже о том, чтобы ответить на них. Участвующие в проекте математики должны были подписать соглашение о неразглашении, требующее от них общаться исключительно через приложение для обмена сообщениями Signal. (Другие формы контакта, такие как традиционная электронная почта, потенциально могли быть просканированы LLM и непреднамеренно обучить его, тем самым загрязнив набор данных.)

Epoch AI провела очную встречу в субботу, 17 мая, и воскресенье, 18 мая. Там участники должны были завершить последнюю партию вопросов-задач. 30 участников были разделены на группы по 6 человек. В течение двух дней ученые соревновались друг с другом, чтобы придумать задачи, которые они могли решить, но которые сбили бы с толку бота.

За 10 минут бот справлялся с задачами уровня PhD, на которую у человека ушли бы недели. Сначала изучал литературу, потом решал "игрушечные" версии задач для тренировки, а затем на основе придуманных им самим игрушечных задач решал основную.

Из сотен попыток математикам удалось "обмануть" бота лишь 10 раз...

Вердикт экспертов: "Это то, что делал бы очень, очень хороший аспирант – даже больше".

"Вспомните, как в 2020 году все поражались, что модель GPT-3 могла (не всегда и с ошибками) выполнять сложение трехзначных чисел. А теперь представьте, где мы будем через несколько лет ..." – Джек Кларк.

## **Language Models Use Trigonometry to Do Addition**

*Subhash Kantamneni, Max Tegmark (MIT)*

Статья: <https://arxiv.org/abs/2502.00873>

Математические рассуждения становятся все более важным показателем возможностей LLM, однако нам не хватает понимания того, как LLM обрабатывают даже простые математические задачи. Чтобы решить эту проблему, мы проанализировали, как три LLM среднего размера (GPT-J, Pythia-6.9B, and Llama3.1-8) вычисляют сложение. Сначала обнаружилось, что число в них представлены в виде спирали, которая тесно связана с задачами сложения и вычитания, а также имеет причинно-следственное отношение к целочисленному делению, умножению и модульной арифметике. Затем мы предлагаем, чтобы LLM вычисляли сложение, манипулируя этой обобщённой спиралью с использованием алгоритма "Clock": для решения  $a+b$  спирали для  $a$  и  $b$  обрабатываются для получения спирали ответа  $a+b$ , которая затем считывается в логиты модели. Демонстрируя, что LLM представляют числа в виде спирали и манипулируют этой спиралью для выполнения сложения, мы представляем первое объяснение математических возможностей LLM на уровне представления.