

Дата	ФИО докладчика	Аффилиации, регалии	Тема доклада	Аннотация
17-сен-2025	Сергей Игоревич Николенко	д.ф.-м.н., ведущий научный сотрудник ПОМИ РАН, профессор СПбГУ, Head of AI в компании Synthesis AI	Как оценивать данные для LLM: от простых фильтров к influence functions	<p>Не все данные одинаково полезны для обучения. Иногда модель становится "умнее" после ещё 50 ГБ текста, а иногда — нет. Какие именно документы оказались наиболее полезны, как узнать это заранее? Об этом мы и поговорим на первом семинаре Лаборатории Маркова.</p> <p>Начнём с простых методов: как собрать чистый корпус данных, можно ли обучить вспомогательные модели, которые будут оценивать качество и полезность датасетов, какие есть полезные сигналы в этом отношении (спойлер: энтропия неплохо работает).</p> <p>Но главное содержание нашего разговора — функции влияния (influence functions, IF). Они отвечают на вопрос о том, как изменится метрика качества, если убрать тот или иной документ из датасета. И здесь внезапно появляется много интересной математики — ну ладно, достаточно много по меркам глубокого обучения. Влияние точки данных легко оценить теоретически, но простые методы совсем не масштабируются, и чтобы вывести это на уровень LLM, нужно что-то придумывать. Мы разберём несколько базовых методов (LiSSA, Arnoldi-IF), важные приближения (LoGra, EK-FAC) и методы ранней оценки данных (LinFiK/ALinFiK).</p>
24-сен-2025	Константин Сергеевич Яковлев	ЛМ	Ищи. Планируй. Учись. (Search. Plan. Learn)	<p>Способность искать и находить решение нетривиальных задач, планировать на несколько шагов вперёд и обучаться (на чужом или собственном опыте) — эти свойства традиционно приписывают интеллектуальным системам (или, как это принято сейчас говорить, искусственному интеллекту). Неудивительно, что эти (под)области ИИ активно развивались с самого зарождения этой науки и развиваются до сих пор.</p> <p>В рамках доклада мы рассмотрим ряд интересных и практически мотивированных задач, которые решаются исследователями с помощью методов эвристического поиска, автоматического планирования и машинного обучения и которые входят в сферу моих интересов. Например: как спланировать множество неконфликтных путей для <math>n</math> мобильных агентов, действующих в общей среде? Или: как решить сложную логическую головоломку вроде кубика Рубика? Или: как по данным датчиков от робота построить карту местности, в которой он перемещается? Особое внимание будет уделено открытым задачам, не имеющим в настоящий момент решения, но которые (на мой взгляд) вполне под силу решить в рамках курсовой/дипломной работы и внести свой вклад в науку.</p> <p>Я надеюсь, что этот доклад придаст мотивации тем из вас, кто хочет заняться научной деятельностью, и поможет сориентироваться тем, кто сейчас выбирает тему КР/ВКР.</p>
1-окт-2025	Кирилл Тыщук	ex-Yandex, ex-Perplexity AI, Incoming Research Engineer, DeepMind	Пост-обучение LLM с подкреплением: последние новости в алгоритмах и их наградах	<p>Мы проведём обзорную экскурсию по современным алгоритмам обучения с подкреплением для больших языковых моделей. Кратко напомним стандартный пайплайн обучения LLM, разберём, почему на финальном этапе нужна именно RL-оптимизация и как конструируются награды.</p> <p>А затем подробнее остановимся на технических деталях самых популярных алгоритмов:</p> <ul style="list-style-type: none"> <li>— базовый PPO, которым обучали первый ChatGPT;</li> <li>— простой DPO, который делает "RL без RL";</li> <li>— новый GRPO + verifiable rewards, которым обучали DeepSeek R1</li> </ul>

8-окт-2025	Александр Панов	Д.ф.-м.н., доцент, заведующий лабораторией Когнитивных когнитивных систем в Институте искусственно го интеллекта AIRI, директор Центра когнитивного моделировани я Московского физико- технического института (МФТИ)	AI Scientist — большой вызов для современных универсальных систем искусственного интеллекта	В докладе обсудим задачу автоматизации научных исследований, применение в ней больших языковых и мультимодальных моделей. Поговорим о современных ИИ-инструментах, используемых в научных исследованиях. Отметим ключевые научные проблемы на пути создания ИИ-исследователя. Обсудим проект создания робототехнического агента-исследователя в задачах автоматизации процессов в химических лабораториях.
15-окт-2025	Алексей Андреевич Власов Андрей Алексеевич Лазтин	Лаборатория Маркова, МКН СПбГУ	Состояния важнее внимания? От классических SSM к S4/S5 и Mamba	<p>Трансформеры стали стандартом для работы с последовательностями, но современное моделирование живо «не трансформером единым». Их ключевое ограничение — квадратичное время работы от длины входа. Мы разберём State Space Models (SSM) — в том числе нашумевшую Mamba — как альтернативу: эти модели масштабируются линейно, естественно описывают динамику во времени и лучше работают с дальними зависимостями.</p> <p>Мы начнём с классических SSM (с помощью которых людей отправляли на Луну), и их практических версий в ML. В качестве мотивации посмотрим на Hippo: как можно «сжать прошлое и восстановить его обратно». Дальше обсудим связь с оптимальным управлением и сигналами, и покажем, как линейная масштабируемость достигается структурой SSM, а стабильность обучения обеспечивается не только корректной инициализацией, но и выбором метода дискретизации.</p> <p>Разберём три удобных взгляда на такие модели — как непрерывные динамические системы, как свёрточные операторы (через структурированные матрицы) и как рекуррентное обновление состояния. Покажем эволюцию от S4 (SISO) к S5 (в том числе MIMO), и чем эти апдейты важны на практике.</p> <p>Завершим кратким обзором Mamba: что в ней означает «селективность» и как устроено обновление скрытого состояния. Заодно обсудим, почему эта модель вызвала такой фурор; но подробное изложение архитектур Mamba и Mamba 2 будет в следующем докладе.</p>
22-окт-2025	Владислав Дмитриевич Ушаков Максим Юрьевич Романов	Яндекс Лаборатория Маркова, МКН СПбГУ	Погружение в детали реализации Mamba	<p>Мы продолжаем детальный разбор SSM-based моделей. В начале семинара разберём архитектуру Mamba — что такое «селективность», как устроено обновление состояния, затем пройдемся по минимальным имплементациям S4 и Mamba. А потом погрузимся в детали эффективной реализации по следующему плану.</p> <ol style="list-style-type: none"> <li>1. Краткое введение в GPU-вычисления: архитектура GPU, модель массового параллелизма.</li> <li>2. Постановка задачи префиксной суммы (Scan) и эволюция алгоритмов её решения на GPU: от Sequential Scan до Work-Efficient Parallel Scan.</li> <li>3. Разберём, где и как Parallel Scan применяется в Mamba, а также какие ещё GPU-оптимизации используются при обучении.</li> </ol>

29-окт-2025	Андрей Алексеевич Лазтин	Лаборатория Маркова, МКН СПбГУ	Mamba-2: SSM и механизм внимания	<p>Мы продолжаем разбор моделей на основе State Space Models и переходим к Mamba-2 — архитектуре, выросшей из идеи Structured State-Space Duality (SSD), которая показывает, что SSM и механизм внимания можно рассматривать как два взгляда на одну и ту же структуру. Этот подход объединяет сильные стороны обоих направлений: эффективность и линейную масштабируемость SSM с выразительностью и гибкостью механизмов внимания.</p> <p>Мы разберём, как устроен новый слой Mamba-2, чем он отличается от первой версии и как через SSD удаётся свести вычисления к матричным операциям, хорошо оптимизированным под GPU. Именно благодаря этому Mamba-2 работает в несколько раз быстрее Mamba-1 и остаётся конкурентоспособной с трансформерами по качеству на задачах языкового моделирования.</p>
5-ноя-2025	Никита Денисов, Александр Комнацкий	Лаборатория Маркова, МКН СПбГУ	Последние новости: Gated Delta Networks и Mamba-3	<p>Мы заканчиваем разбор моделей на основе SSM. В этот раз расскажем сразу про два свежих продолжения этой линии работ: Gated Delta Networks (ICLR 2025) и Mamba-3 (подана на ICLR 2026). Обе модели развивают идеи линейных рекуррентных архитектур и показывают, как SSM может конкурировать с трансформерами на всё более сложных задачах.</p> <p>Gated DeltaNet строится на базе Mamba-2 и объединяет два подхода к управлению памятью — «gating» и «delta rule». Это позволяет модели выбирать, что забывать полностью, а что обновлять выборочно, при этом сохраняя эффективность и адаптированность к длинным контекстам.</p> <p>Mamba-3 делает следующий шаг — переосмысливает саму структуру SSM с точки зрения эффективного инференса. В ней появляются более выразительная рекуррентная динамика, комплексные обновления состояния и MIMO-формулировка (multi-input, multi-output), что делает её быстрее и мощнее без увеличения вычислительных затрат на порождение.</p>

12-ноя-2025	Захар Варфоломеев	Songsterr, ex. Yandex Music	VQ-VAE: когда нужны дискретные представления и как их готовить	<p>VQ-VAE (<a href="https://t.me/vf_science/119">https://t.me/vf_science/119</a>) — популярная нейросетевая архитектура, которая активно применяется в моделях, работающих с аудио (<a href="https://arxiv.org/pdf/2411.01156#page=2.73">https://arxiv.org/pdf/2411.01156#page=2.73</a>), изображениями (<a href="https://arxiv.org/pdf/2406.06525">https://arxiv.org/pdf/2406.06525</a>), видео (<a href="https://arxiv.org/pdf/2405.15223#page=3.81">https://arxiv.org/pdf/2405.15223#page=3.81</a>), в мультимодальных моделях (<a href="https://arxiv.org/abs/2405.09818">https://arxiv.org/abs/2405.09818</a>), а также в 3D-моделировании, трекинге движений и естественных науках (биологии, химии). На основе VQ-VAE создаются модели для сжатия и восстановления данных, которые сейчас используются во многих стриминговых платформах. Также VQ-VAE применяется для задач порождения (<a href="https://arxiv.org/pdf/2306.05284#page=3.14">https://arxiv.org/pdf/2306.05284#page=3.14</a>) или извлечения информации; например, на нём работает моё любимое audio2midi (<a href="https://t.me/vf_science/126">https://t.me/vf_science/126</a>), которое я делаю, или rapoptic segmentation (<a href="https://arxiv.org/pdf/2309.15505#page=1.53">https://arxiv.org/pdf/2309.15505#page=1.53</a>).</p> <p>В общем, идея полезная, но у VQ-VAE есть конкретные сценарии применения и проблемы, о которых речь пойдёт далее:</p> <ul style="list-style-type: none"> <li>— он вычисляет N центроидов кластеров по картам признаков, полученным из кодировщика; кластеризация приводит к потере информации;</li> <li>— инициализация центроидов кластеров значительно влияет на метрики для любой задачи; всем известный k-means даёт здесь полезную последовательность идей;</li> <li>— через операцию VQ не протекают градиенты, поэтому мы вынуждены использовать аппроксимации, чаще всего STE (<a href="https://proceedings.mlr.press/v202/huh23a/huh23a.pdf">https://proceedings.mlr.press/v202/huh23a/huh23a.pdf</a>);</li> <li>— поскольку градиент останавливается после VQ, он не доходит до словаря и кодировщика; нужны вспомогательные функции потерь. Одна из них будет приближать выходы кодировщика к элементам словаря, вторая — приближать элементы словаря к выходам кодировщика; а это приводит к сложностям в оптимизации.</li> </ul> <p>Поэтому для некоторых моделей в CV используют (<a href="https://arxiv.org/pdf/2410.07073v2">https://arxiv.org/pdf/2410.07073v2</a>) просто патчи, как в ViT (<a href="https://arxiv.org/pdf/2010.11929">https://arxiv.org/pdf/2010.11929</a>). Это тоже интересный момент, и мы ещё вернемся к нему, а пока идём дальше.</p> <p>— Часто вспомогательные лоссы считают L2-расстояние между конкретными семплами, и возникает коллапс словаря, когда модель использует только небольшую часть токенов, игнорируя остальные. Это снижает эффективность модели и происходит потому, что мы приближаем друг к другу лишь конкретные семплы, и ближе всех к выходам кодировщика будут последние, а большинство предыдущих окажутся дальше последних и никогда не будут выбраны. Поэтому создавать большие словари (10K+) не имеет смысла.</p> <p>— При обучении с нуля подбор числа элементов словаря и их размерности — это трудоёмкий процесс: каждый раз нужно инициализировать словарь заново и проверять разные центроиды кластеров.</p>
19-ноя-2025	Максим Николаев	Лаборатория Маркова, МКН СПбГУ	BART	<p>BART, Bayesian Additive Regression Trees — это непараметрическая модель, которая наследует выразительность от ансамблей деревьев решений и описание неопределённости от байесовского подхода. Как это часто бывает, апостериорное распределение этой модели краткой аналитической формы не имеет, поэтому для работы с ней используются методы Монте-Карло на марковских цепях (MCMC). Мы обсудим устройство модели, а также используемые методы MCMC. В оставшееся время обсудим направления для дальнейших исследований. Материал не требует специфической подготовки, но будет полезным понимание основных понятий теории цепей Маркова вплоть до стационарного распределения.</p>
3-дек-2025	Артём Важенцев	Сколтех	Оценка неопределённости в NLP: от классификации к большим языковым моделям	<p>Оценка неопределённости (UQ) является важным инструментом повышения надёжности систем обработки естественного языка — от классификации текстов до порождения ответов большими языковыми моделями. Она позволяет снижать риск критических ошибок, выявлять неоднозначные случаи и обнаруживать галлюцинации при порождении текста.</p> <p>Несмотря на быстрый прогресс в этой области, существующие подходы часто оказываются нестабильными на сложных и неоднозначных данных и плохо переносятся с задач классификации на авторегрессионные языковые модели. В докладе обсуждаются ключевые вызовы оценки неопределённости в NLP. Для задач текстовой классификации рассматриваются как базовые, так и state-of-the-art методы, направленные на повышение надёжности выборочной классификации. Для задач порождения текста большими языковыми моделями рассматриваются современные методы оценки неопределённости, а также анализируется роль attention механизма в выявлении галлюцинаций и низкокачественных ответов.</p>

10-дек-2025	Яна Вейцман	Universität Göttingen (work done while at Saarland University)	The Expressive Capacity of State Space Models: A Formal Language Perspective	<p>State Space Models (SSM) за последние годы стремительно набрали популярность, демонстрируя качество языкового моделирования, сопоставимое с трансформерами. Однако их фундаментальные теоретические свойства до сих пор остаются малоизученными. В докладе Яны Вейцман будут представлены результаты исследования 2024 года, в котором впервые проводится систематический теоретический анализ выразительной мощности SSM по сравнению с трансформерами и классическими RNN в рамках парадигмы формальных языков.</p> <p>Мы обсудим пересекающиеся, но различающиеся сильные стороны SSM и трансформеров. В частности, речь пойдёт о способности SSM эффективно решать задачи слежения за состоянием в star-free языках — области, где трансформеры испытывают существенные трудности. Кроме того, будет показано, что SSM способны оптимально моделировать ограниченные иерархические структуры без явной симуляции стека. При этом в общепринятом дизайне современных SSM скрыто конструктивное ограничение, которое существенно снижает их выразительность.</p> <p>В докладе также будет обсуждаться, какие практические выводы следуют из этих теоретических результатов для будущего архитектур SSM, развития языковых моделей и систем вроде Mamba, а также какие новые направления исследований они открывают.</p> <p>Доклад будет интересен специалистам в области формальных языков и архитектур языковых моделей. Язык доклада — английский.</p>
17-дек-2025	Константин Сергеевич Яковлев Обловатный Родион Игоревич Смирнов Станислав Константинович	Лаборатория Маркова, МКН СПбГУ	Разбор статей NeurIPS-2025	<p>Лучшие из пяти тысяч статей NeurIPS, получившие best paper awards, уже опубликованы. Так что на заключительном заседании семинара мы решили провести их краткий разбор. Будет семь (или шесть, как получится) докладов по 10-15 минут:</p> <ul style="list-style-type: none"> <li>— Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond)</li> <li>— Gated Attention for Large Language Models: Non-linearity, Sparsity, and Attention-Sink-Free</li> <li>— 1000 Layer Networks for Self-Supervised RL: Scaling Depth Can Enable New Goal-Reaching Capabilities</li> <li>— Why Diffusion Models Don't Memorize: The Role of Implicit Dynamical Regularization in Training</li> <li>— Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?</li> <li>— Optimal Mistake Bounds for Transductive Online Learning</li> <li>— Superposition Yields Robust Neural Scaling</li> </ul>
23-дек-2025	Александр Куликов	Neapolis University Pafos	Булевы схемы: недавние результаты и открытые задачи	<p>Булевы схемы представляют собой простую, естественную и выразительную модель вычислений, достаточно мощную, чтобы моделировать алгоритмы. Схемы имеют применения в таких областях компьютерных наук, как алгоритмы, теория сложности, криптография, формальная верификация, проектирование аппаратуры, машинное обучение, SAT-солверы.</p> <p>Я расскажу про недавние результаты нашей группы в области синтеза схем и нижних оценок на размер схем и время работы алгоритмов. Отдельное внимание уделю открытым задачам и перспективным, на мой взгляд, направлениям для сотрудничества в этой области.</p>