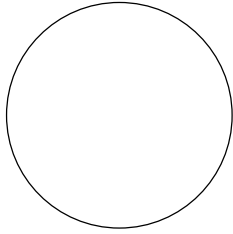




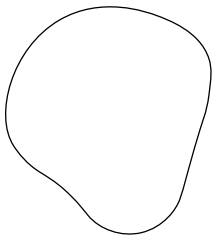
2023

Годовой отчёт



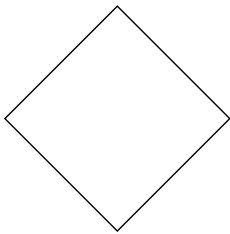
Природа

Постоянное движение и развитие, закономерность и системность, законы и ритмы, хаос и жизнь



Человек

Живой организм, часть природы, способный мыслить, чувствовать и говорить, несовершенный



Технологии

Творение человека и его продолжение, логичное и предсказуемое, утилитарное, создано помогать



AIRI

Единство природы, человека и технологий

Содержание

Вступительное слово	4
Миссия AIRI	5
Ценности AIRI	6
Направления деятельности	7
Ключевые результаты	8
Руководители научных групп	10
Руководство Института	12
Научные результаты	13
Научные публикации	40
Мероприятия и выступления	61
О нас пишут и говорят	75
Партнерства и коллаборации	76
Ученые AIRI в социальных сетях	78
Контакты	79

Третий год жизни Института



Иван Оселедец
Генеральный директор

В 2023 году Институт закрепил статус одного из лидеров в России по искусственному интеллекту и уверенно идёт к краткосрочной цели — стать абсолютным лидером по числу публикаций на конференциях A/A*.

Наша основная цель на ближайшие годы — разработка мультимодальных, мультиагентных систем сильного AI. В этом году наши команды внесли огромный вклад в создание целого ряда больших моделей, таких как Kandinsky, первая в России text-to-video модель Kandinsky Video и первая в России мультимодальная модель OmniFusion.

В стране много коллективов, активно работающих в области искусственного интеллекта: Сколтех, ВШЭ, МФТИ, ИТМО, Иннополис, ИСП РАН, но только AIRI ставит своей основной целью создание и развитие сильного искусственного интеллекта. Я считаю, что мы можем это сделать, так как технологии уже достигли нужного уровня, у нас собраны сильнейшие команды в области AI в стране.

В 2023 году к нам присоединились новые коллеги — группу NLP возглавил Александр Панченко, Дмитрий Дылов организовал лабораторию «Сильный AI в медицине». Мы планируем и дальше создавать новые коллективы.

Человекоцентричность, ответственность и вклад в будущее по-прежнему остаются одними из ключевых ценностей Института. Мы продолжаем активно вести просветительскую деятельность, объединять сообщество энтузиастов вокруг обсуждения насущных вопросов предметной области.

За 2023 год в нашем портфолио сформировался целый ряд крупных научных результатов по фундаментальным и прикладным аспектам AI, с которыми вы можете ознакомиться в этом отчёте.

Миссия AIRI

Создание универсальных систем
искусственного интеллекта, решающих
задачи реального мира

Основная цель Института — найти возможности применения искусственного интеллекта для решения сложных социальных, экономических и научных задач. Научные сотрудники Института занимаются исследованиями в фундаментальных и прикладных областях. В работе над своими проектами они стремятся получить прорывные результаты в области искусственного интеллекта и его приложений, участвуя в формировании глобальной исследовательской повестки.

Ценности AIRI



Человекоцентричность



Свобода научного
выражения



Ответственность
и вклад в будущее



Открытость
и прозрачность



Партнерство
и коллаборация



Направления деятельности



Партнерства

Развитие партнерств с промышленностью, поддержка AI стартапов, создание возможностей для коммерциализации исследований



Вклад в развитие AI

Участие в глобальном технологическом развитии искусственного интеллекта через создание, развитие и поддержку open-source AI проектов



Исследования

Проведение прорывных исследований в области AI и его приложений, участие в формировании глобальной исследовательской повестки, становление центром экспертизы в области AI



Выступления

Проведение международных научных конференций и встреч, создание и поддержка AI соревнований



Популяризация AI

Проведение открытых семинаров, обучающих сессий, дискуссионных групп и популяризация AI в целом

Ключевые результаты



146

публикаций

30

статей
на конференциях A*

43

статьи в журналах
с рейтингом Q1

8

статей
на конференциях A



Руководители научных групп



Артем Шелманов

Обучение на слабо размеченных данных



Артур Кадури

Глубокое обучение в науках о жизни



Евгений Фролов

Группа технологий персонализации



Олег Рогов

Доверенные и безопасные интеллектуальные системы



Александр Панов

Нейросимвольная интеграция



Дмитрий Ветров

Вероятностные методы машинного обучения



Ольга Кардымон

Биоинформатика



Андрей Кузнецов

FusionBrain

Руководители научных групп



**Александр
Панченко**

Вычислительная
семантика



**Елена
Тутубалина**

Domain-specific NLP



**Алексей
Осадчий**

Нейроинтерфейсы



**Семен
Буденный**

Дизайн новых
материалов



**Илья
Макаров**

AI в промышленности



**Евгений
Бурнаев**

Обучаемый интеллект



**Манвел
Аветисян**

Прикладные
проекты



**Дмитрий
Дылов**

Лаборатория «Сильный AI
в медицине»

Руководство Института



**Иван
Оселедец**

Генеральный
директор



**Манвел
Аветисян**

Директор по развитию
прикладных проектов



**Максим
Кузнецов**

Директор проектного
офиса



**Антон
Ризаев**

Финансовый
директор



**Александра
Бройтман**

Директор по маркетингу
и коммуникациям



**Мария
Мараховская**

HR Директор



**Юлия
Никитина**

Руководитель отдела
юридического
сопровождения



**Степан
Мамонтов**

Главный
разработчик




**Константин
Катанов**

Руководитель
направления IT




**Ольга
Суровегина**

Директор по научно-
техническим партнерствам



Научные
результаты



Главные
результаты

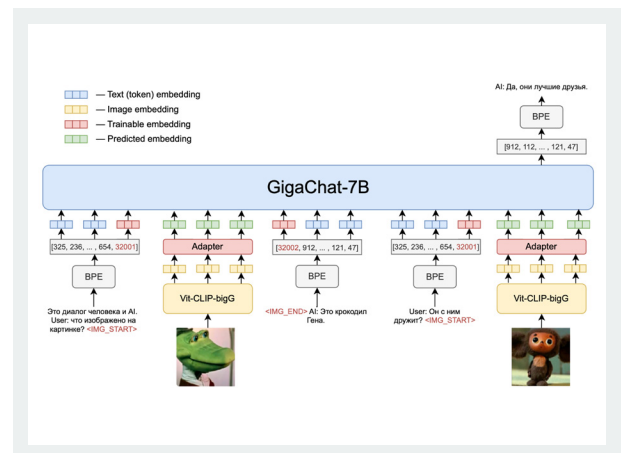


Андрей Кузнецов
FusionBrain,
ведущий научный сотрудник

OmniFusion

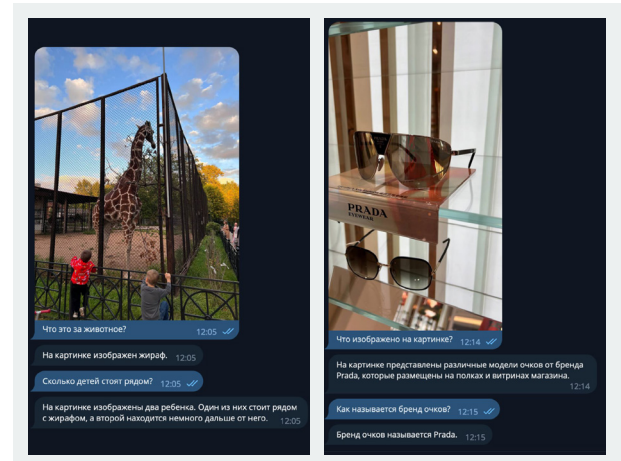
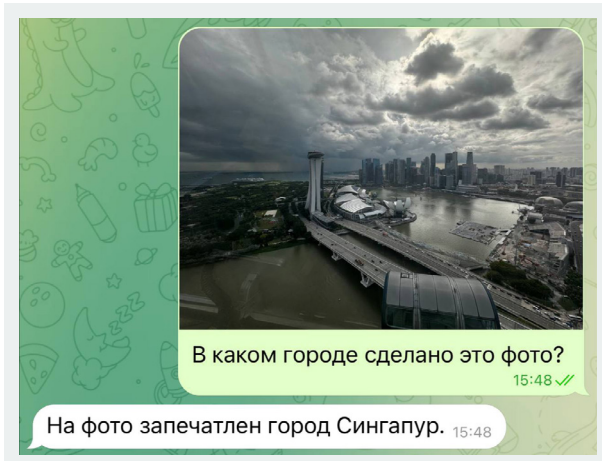
Разработка первой российской мультимодальной диалоговой модели, которая способна поддерживать диалог с пользователем не только на базе текста, но и с помощью изображений

Модель умеет понимать картинки и строить диалог вокруг них, отвечая на разные вопросы и генерируя подробные описания сцен и объектов. В основе модели лежит языковая модель GigaChat-7B и Mistral-7B (для публикации в open source). Обучение проводилось в два этапа. Сначала — крупномасштабное многопрофильное предварительное обучение для выравнивания признаков, затем — тонкая настройка на последующих задачах.



Сравнение OmniFusion с открытой моделью LLaVA на основе 10 различных бенчмарков показало, что качество OmniFusion не уступает, а в ряде случаев даже превосходит конкурента, несмотря на то, что в основе OmniFusion лежит намного более «легкая» языковая модель. В основе LLaVA лежит языковая модель с 13 млрд. параметров, в то время как языковая модель в основе OmniFusion содержит всего 7 млрд. Это значит, что модель более экономичная и быстрая.

Модель уже стабильно работает на английском языке и обучается грамотному владению русским, чтобы стать доступной пользователям, а команда готовит научную публикацию о процессе создания OmniFusion. В настоящее время проводятся эксперименты по внедрению модели в GigaChat, а также обсуждаются возможности применения модели в самостоятельном формате.





Андрей Кузнецов

FusionBrain,
ведущий научный сотрудник



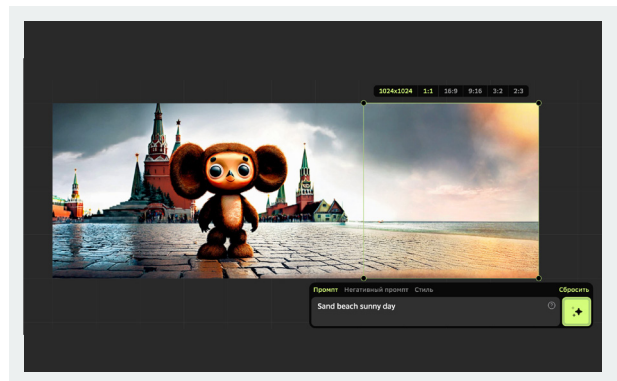
Денис Димитров

FusionBrain,
научный консультант

Kandinsky 2.X, 3.0

Разработка плана исследований и экспериментов для создания новой диффузионной модели синтеза изображений по текстовым описаниям на различных языках.

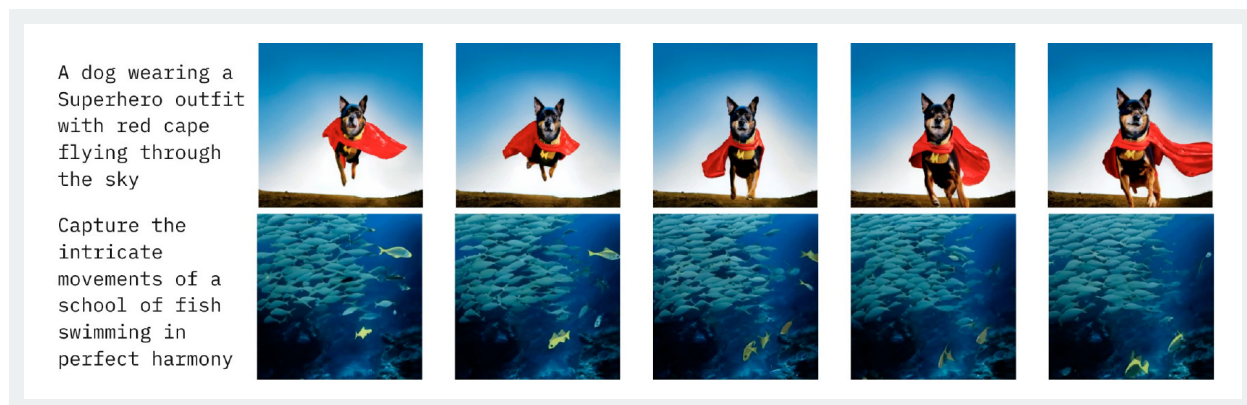
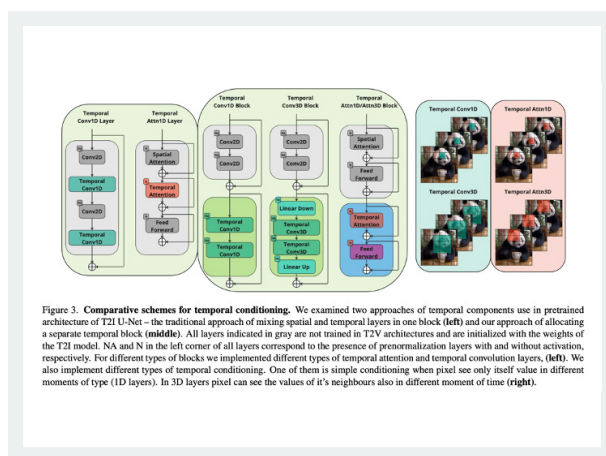
В ходе исследований проводились различные эксперименты с архитектурой решения в части использования image prior составляющей, оценке её вклада в качество генерации. В результате исследований подготовлена и принята статья на конференцию EMNLP (Core A*). Статья также заняла первое место среди статей дня на портале Hugging Face, обогнав статьи Google DeepMind и Carnegie Mellon University.

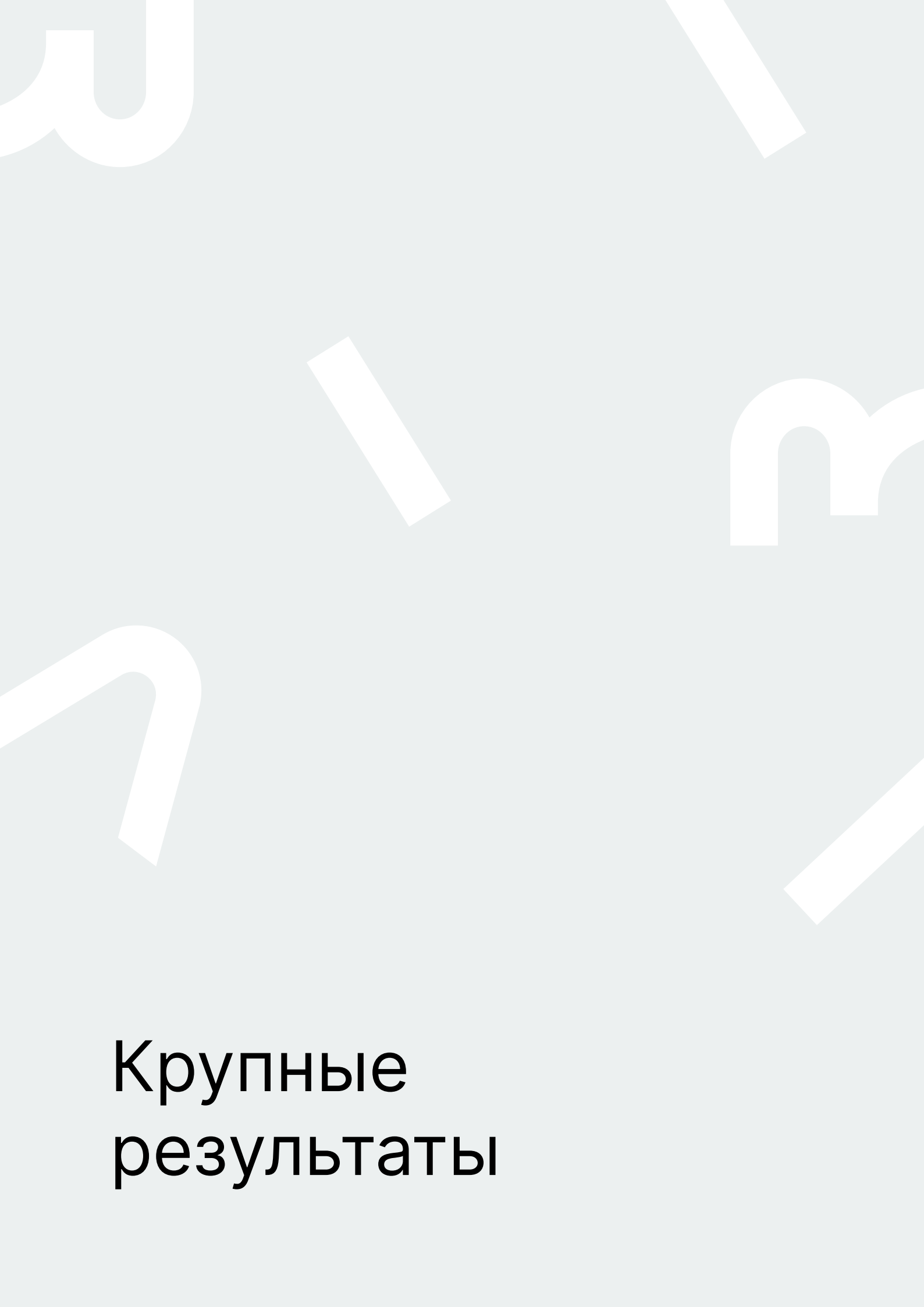


Kandinsky Video

Исследования включали в себя создание первой в РФ end-to-end модели синтеза видео по текстовым описаниям, основанной на диффузионной модели генерации кадров.

Решение состоит из двух этапов: синтез ключевых кадров и синтез интерполяционных кадров, что позволяет контролировать как создание сюжета видео, так и заполнение кадрами до 30 FPS с целью достижения эффекта плавности движений объектов, что повышает визуальное качество синтеза., сейчас проходит рецензирование. Статья также заняла второе место среди статей дня на портале Hugging Face, уступив только статье LeCun'a о новом мультимодальном бенчмарке.





Крупные
результаты

Успехи в области развития технологий обработки естественного языка (NLP)

Группами под руководством Александра Панченко, Артема Шелманова и Елены Тутубалиной был получен ряд новых результатов



Александр Панченко
Ведущий научный сотрудник



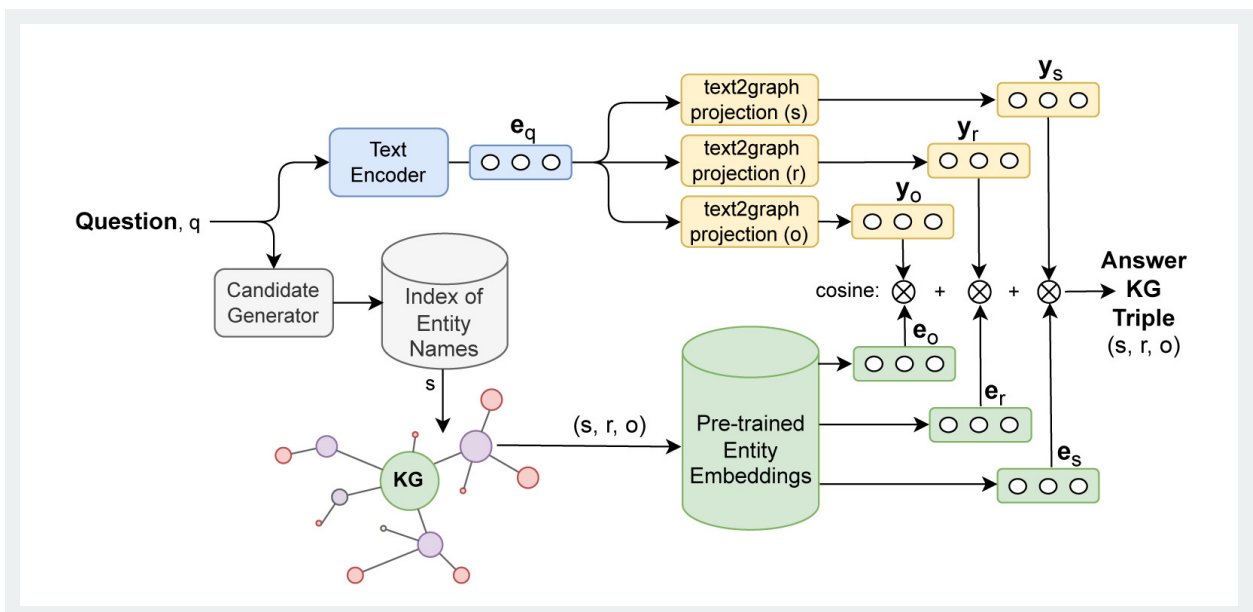
Артем Шелманов
Научный консультант



Елена Тутубалина
Научный консультант

Разработаны новые методы для вопросно-ответных систем с использованием графов знаний, которые позволяют отвечать на простые вопросы лучше, чем это делает ChatGPT.

Разработанные методы позволяют отвечать на вопросы с использованием графа знаний и больших языковых моделей. На нескольких датасетах методы показывают улучшения относительно сильных базовых подходов, включая ChatGPT.



Представлены новые вычислительно эффективные методы для оценивания неопределенности предсказаний языковых моделей.

Разработанный подход комбинирует два state-of-the-art метода для получения значительного прироста качества в задаче выборочной классификации. Такой гибридный подход показал свою эффективность для задач с высокой степенью неоднозначности, например, для определения степени токсичности или тональности текстов. Кроме того, не смотря на то, что данный метод не требует больших вычислительных затрат, он способен превосходить такие методы, как глубокий ансамбль, который требует многократного увеличения вычислительных ресурсов.

Разработана библиотека LM-Polygraph для оценки неопределенности LLM, в том числе моделей, развернутых в виде сервисов.

Оценки неопределенности позволяют детектировать низкокачественные ответы моделей (например, галлюцинации) и отказываться от демонстрации их пользователю.

Разработан новый бенчмарк на основе датасета PAUQ.

Было предложено четыре разбиения text-to-SQL нашего датасета PAUQ, получившие название Shifted PAUQ, для оценки композиционного и мультилингвального обобщения SoTA text-to-SQL моделей. С помощью контекстно-свободных грамматик разработаны метрики оценки композиционности моделей.

Проведено исследование методов оценки неопределенности совместно с методами дебайсинга.

Выявлены наиболее перспективные подходы к дебайсингу, которые оказывают минимальный негативный эффект на оценки неопределенности.

Предложен новый фреймворк Vote'n'Rank, который включает в себя восемь методов агрегации метрик моделей на основе теории многокритериального принятия решений.

Помимо этого, представлены практические рекомендации по использованию фреймворка, которые опираются на теоретические свойства методов агрегации результатов и сценарии потенциального применения.

Rank	σ^{am}	σ^{gm}	σ^{og}	Copeland	Minimax	Plurality	Dowdall	Borda
1	91.18 ↓0	90.89 ↓0	0.074 ↓0	29.00 ↓0	0 ↓0	2.00 ↓0	4.95 ↓0	260.50 ↓0
2	91.07 ↓0	90.78 ↓0	0.075 ↑4	25.00 ↑1	-5.50 ↑1	2.00 ↑13	4.08 ↑13	256.00 ↓0
3	90.88 ↓0	90.56 ↓1	0.076 ↓1	24.00 ↓1	-6.00 ↓1	1.50 ↓0	3.82 ↓0	247.50 ↓0
4	90.86 ↓0	90.48 ↓0	0.076 ↓0	22.00 ↑3	-6.50 ↓2	1.00 ↑1	3.41 ↓0	241.50 ↓0
5	90.74 ↓0	90.44 ↓0	0.077 ↓0	22.00 ↑10	-7.00 ↑2	1.00 ↓3	3.27 ↓3	233.50 ↑1
6	90.66 ↓0	90.34 ↓0	0.078 ↑1	22.00 ↓2	-7.00 ↑9	0.50 ↓0	2.57 ↓1	229.50 ↑1
7	90.48 ↓0	90.11 ↓0	0.082 ↑3	16.00 ↓1	-7.00 ↓1	0.00 ↓3	2.55 ↓0	220.50 ↓2

Table 3: Results of re-ranking the GLUE benchmark. Changes in the system ranks are depicted with arrows, whilst the superscripts denote scores assigned by the aggregation procedure. Notations: =HUMAN; =ERNIE; =STRUCTBERT+CLEVER; =DEBERTA+CLEVER; =DEBERTA/TURINGNLRV4; =MACALBERT+DKM; =T5; =ALBERT+DAAF+NAS; =FUNNEL. The superscript values stand for the voting rules' scores, whilst the subscript values indicate changes in the ranking positions. ↑ x means up x positions, ↓ x means down x positions, ↓↑ means no changes.

Вероятностные методы машинного обучения

Группой под руководством Айбека Аланова и Дмитрия Ветрова были получены важные результаты



Дмитрий Ветров
Ведущий научный сотрудник

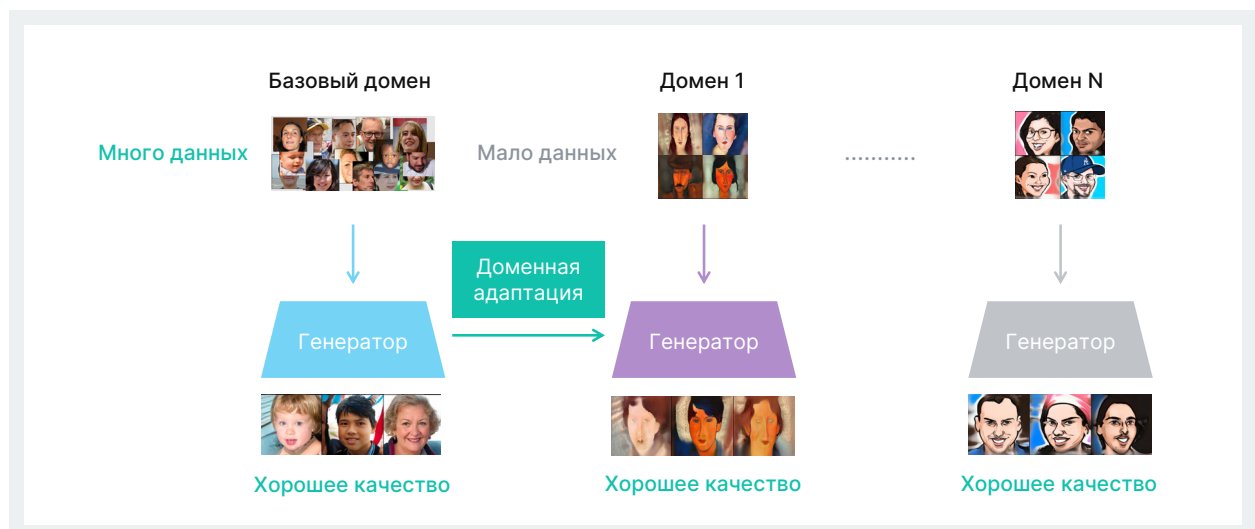


Айбек Аланов
Научный сотрудник

В 5000 раз ускорена доменная адаптация GAN для генерации лиц.

В результате опубликована статья на конференции ICCV. В данной работе рассматривалась область генерации синтетических (неотличимых от реальных) изображений из заданного домена (фотографии лиц, домов, котов). Точнее, решается задача доменной адаптации для sota модели StyleGAN2 — суть задачи в том, чтобы

получить в соответствие каждому синтезированному изображению стилизованную копию из другого домена (рисунок, акварель, мультяшный стиль). Все предшествующие (в том числе sota) методы использовали дообучение полноценной копии StyleGAN2 — предложенный нами метод решает задачу, используя в 5000 раз меньше памяти при обучении и инференсе, и при этом достигает сравнимого качества с существующими подходами.



Улучшение диффузионных моделей.

Представлена новая модель, которая может применяться к различным распределениям в экспоненциальном семействе, что делает ее полезной для моделирования данных на многообразиях с ограничениями, таких как единичная сфера или положительные полуопределенные матрицы (NeurIPS 2023).

Улучшение генеративных моделей для генерации звука.

Совместно с коллегами из Samsung AI Center было написано две статьи, которые были приняты на ICASSP 2023 и InterSpeech 2023.

Разработан метод для манипуляции изображениями с помощью генеративной модели StyleGAN и была подана статья на конференцию CVPR 2024.

В данной работе решается задача обращения (inversion) реальных изображений для модели StyleGAN. Предложенное нами решение полностью превосходит все существующие (в том числе и sota) методы по объективным метрикам реконструкции. Отличительной особенностью решения является высокая способность редактирования реальных изображений и возможность применения на экстремальных out-of-domain примерах.

Разработан метод для редактирования реальных изображений с помощью text-to-image диффузионных моделей и была подана статья на конференцию CVPR 2024.

Предложенный метод обходит все основные (в том числе sota) методы для редактирования по откликам реальных пользователей и в среднем по метрикам. Отличительной особенностью является робастность к выбору гиперпараметров – то есть метод отлично себя показывает при выборе из достаточно большого диапазона.

Разработан метод для эффективного решения задачи HairSwap с помощью генеративной модели StyleGAN и была подана статья на конференцию CVPR 2024.

В данной работе рассматривалась сложная задача переноса прически по входной фотографии. Основная сложность данной задачи заключается в том, чтобы добиться естественного и реалистичного результата при переносе волос. Предложенный метод решает эту задачу без оптимизации, что существенно ускоряет его работу, при этом он достигает лучшего качества по сравнению с другими подходами.



Обучаемый интеллект

Группой под руководством Евгения Бурнаева получены новые результаты в области оптимального транспорта и приложений



Евгений Бурнаев
Ведущий научный сотрудник

Новые методы построения генеративных моделей на основе оптимального транспорта и диффузионных процессов (результаты опубликованы в 3 статьях ICLR, 2023 и 4 статьях NeurIPS, 2023).

Особенно стоит упомянуть статьи на ICLR, 2023 (Notable-top-25% paper) и NeurIPS, 2023 (Oral talk, 3% of accepted papers). В первой работе получен метод построения генеративной модели на основе нового эффективного алгоритма решения задачи слабого оптимального транспорта. Во второй работе исследована эквивалентность задачи непрерывного оптимального транспорта в энтропийной постановке и задачи нахождения диффузионного моста Шредингера; предложен масштабируемый нейросетевой алгоритм для решения обеих задач. Разработанный метод позволяет эффективно решать задачи непарного обучения, например, повышать разрешение изображений.

Новые методы анализа структуры многомерных данных на основе топологического анализа данных (статья на ICLR, 2023).

На основе топологического подхода к оценке размерности предложен подход к детекции фейковых текстов, сгенерированных большими языковыми моделями, например, GPT4 (статья Neurips, 2023).

На основе разработанных подходов к подсчету оптимального транспорта был разработан новый метод, который может работать совместно со множеством существующих оффлайн методов обучения с подкреплением.

В частности, в комбинации с недавно предложенным SOTA методом — ReBRAC, разработанным командой банка Tinkoff, который обеспечивает ускорение и улучшение точности решения задач примерно на 40%, предлагаемый подход позволяет достичь дополнительного улучшения в среднем на 20% на задачах в среде AntMaze.

Удалось показать, что разработанный ранее авторами статьи набор топологических признаков применим для совершенно нового типа данных — речи — при решении стандартных задач классификации, и в некоторых случаях позволяет добиться SOTA результатов (статья Interspeech, 2023).

Новые методы 3D компьютерного зрения и восстановления 3D форм.

Построена выборка разноточных мультимодальных 3D данных, которая позволяет обучать модели для восстановления 3D форм. Данный результат позволит обеспечить научное сообщество качественными данными для обучения различных моделей 3D компьютерного зрения (2 статьи на CVPR, 2023).

Ученые AIRI разработали ряд методов оптимального переноса данных из одного домена в другой с помощью нейронных сетей.

Такие методы сейчас активно исследуются в научном сообществе и часто применяются в генеративном моделировании, обработке изображений и анализе биологических данных.

В основе новых методов лежит теория оптимального транспорта, значительный вклад в развитие которой был сделан советским математиком Леонидом Канторовичем еще в XX веке. Именно поэтому предложенные методы, основанные на нейронных сетях, получили общее название «нейронный оптимальный транспорт». Эти новые подходы обладают большей интерпретируемостью и теоретической обоснованностью, чем многие существующие альтернативные подходы для перевода домена данных, которые обычно строятся на эвристических принципах и не имеют строгого теоретического обоснования.



Рис. (а)



Рис. (б)

Примеры работы разработанного нейросетевого алгоритма на трех задачах переноса домена с непарными обучающими выборками. Рис. (а) — стилизация фотореалистичного портрета под аниме (первые две строки) и генерация изображения здания по заданному изображению пейзажа (вторые две строки). Рис. (б) — генерация изображения обуви по стилю, задаваемому входным изображением сумки (в строках 2-4 показан результат разнообразной генерации).

Вычислительный интеллект

Научным коллективом под руководством Ивана Оселедца был получен ряд результатов



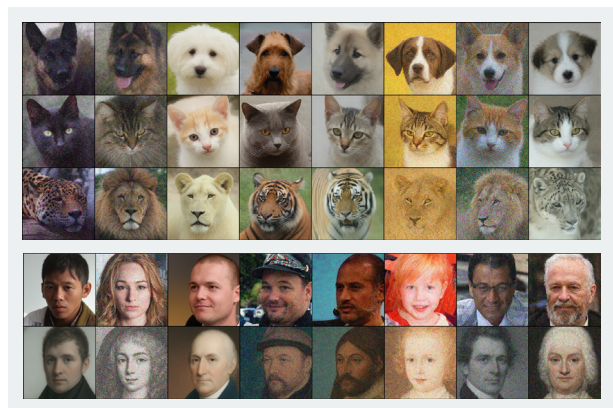
Иван Оселедец
Генеральный директор

Показано, что в ряде случаев диффузионные модели реализуют оптимальный транспорт между исходным распределением и нормальным распределением, сформулирована и протестирована гипотеза в общем случае.

Показано, что две диффузионные модели, обученные на двух отдельных датасетах, дают «близкие» латентные коды для «близких» картинок. Статья уже процитирована в 2023 году 18 раз, согласно Google Scholar. Статья опубликована на ICLR 2023.

Выполнен ряд работ по развитию методов машинного обучения для математического моделирования.

Предложен метод быстрого увеличения датасета за счет использования свойств уравнения (аугментация данных) для обучения нейронных операторов. Опубликовано статья на ICML 2023.



Equation	Model	simple datasets			complex datasets		
		\times	\checkmark	g	\times	\checkmark	g
Convection-diffusion	FNO	0.067	0.048	28%	0.510	0.418	18%
	DeepONet	0.675	0.567	16%	—	—	—
	DiResNet	0.023	0.010	56%	0.312	0.225	28%
	MLP	0.094	0.050	49%	0.566	0.496	12%
	U-Net	0.069	0.031	55%	0.419	0.364	13%
	SNO	0.086	0.066	23%	0.416	0.373	10%
Elliptic alpha	FNO	0.066	0.036	46%	0.306	0.207	32%
	DeepONet	—	0.826	—	—	—	—
	DiResNet	0.105	0.021	80%	0.160	0.133	17%
	MLP	0.088	0.053	40%	0.322	0.253	21%
	U-Net	0.093	0.070	25%	0.386	0.194	50%
	SNO	0.082	0.050	39%	0.251	0.209	17%
Elliptic beta	FNO	0.034	0.021	38%	0.181	0.126	30%
	DeepONet	—	0.832	—	—	0.946	—
	DiResNet	0.099	0.022	78%	0.089	0.062	30%
	MLP	0.069	0.035	50%	0.238	0.138	42%
	U-Net	0.070	0.067	4%	0.170	0.143	16%
	SNO	0.068	0.038	44%	0.187	0.144	23%
Wave	FNO	0.200	0.159	21%	0.650	0.628	3%
	DeepONet	—	—	—	—	—	—
	DiResNet	0.053	0.048	9%	0.43	0.38	12%
	MLP	0.313	0.295	6%	—	0.99	—
	U-Net	—	—	—	0.57	0.52	9%
	SNO	0.37	0.37	0%	—	—	—
Navier-Stokes			n^1			n^2	
	FNO	0.005	0.003	40%	0.022	0.010	55%
	UNet	0.019	0.09	53%	0.069	0.037	46%
	DiResNet	0.021	0.015	29%	0.073	0.045	38%
	MLP	0.082	0.066	38%	0.082	0.066	20%
	SNO	0.004	0.003	25%	0.013	0.008	38%

Выполнен ряд работ по развитию тензорных методов: конструктивный подход к построению тензорных аппроксимаций для функций специального вида.

В частности, получено существенное ускорение расчета кооперативных игр и получен оптимальный алгоритм вычисления перманента. Этот алгоритм в конце 2023 года получил дальнейшее развитие для ускорения моделирования бозонных сэмплеров, которые играют важную роль в фотонных квантовых компьютерах. Разработан новый тензорный метод оптимизации типа «черный ящик» PROTES, который

существенно превзошел аналоги, в том числе, пакет Nevergrad (Meta) на целом ряде бенчмарков. Опубликована статья на NeurIPS 2023. Разработан новый метод оптимизации функций, представленных в малоранговом формате. Статья опубликована в SISC (Q1).

Разработан метод уменьшения памяти для обучения нейросетевых моделей за счет квантизации активаций в обратном проходе (метод FewBit).

Сокращение памяти без потери точности составило 5-10%.

Опубликована статья на ICML 2023.

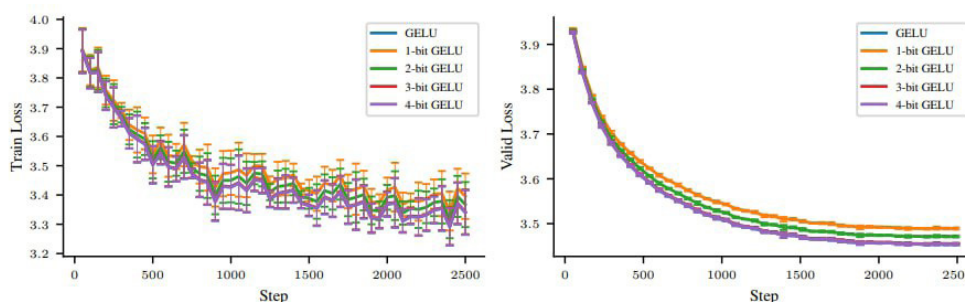


Figure 4. Dynamic of loss values in finetuning of ruDALL-E Malevich with Few-bit GELU activations.

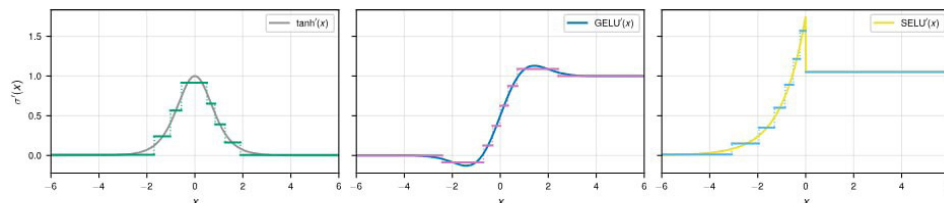
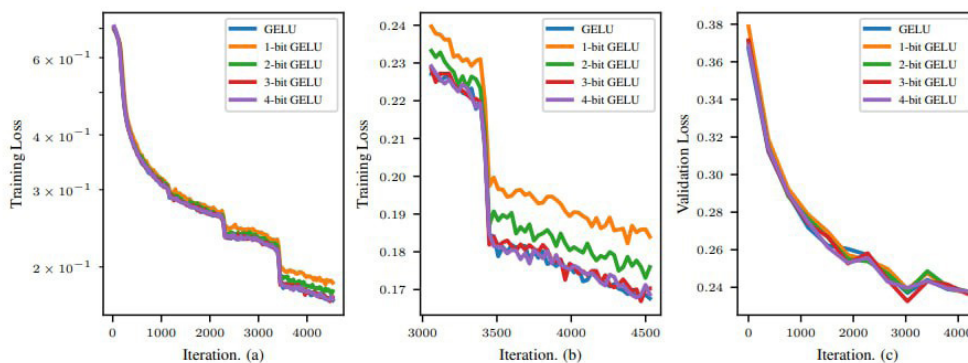


Figure 1. Optimized 3-bit piecewise-constant approximations of the derivatives of activation functions.

Нейросимвольная интеграция

Группой под руководством Александра Панова получены следующие результаты



Александр Панов
Ведущий научный сотрудник

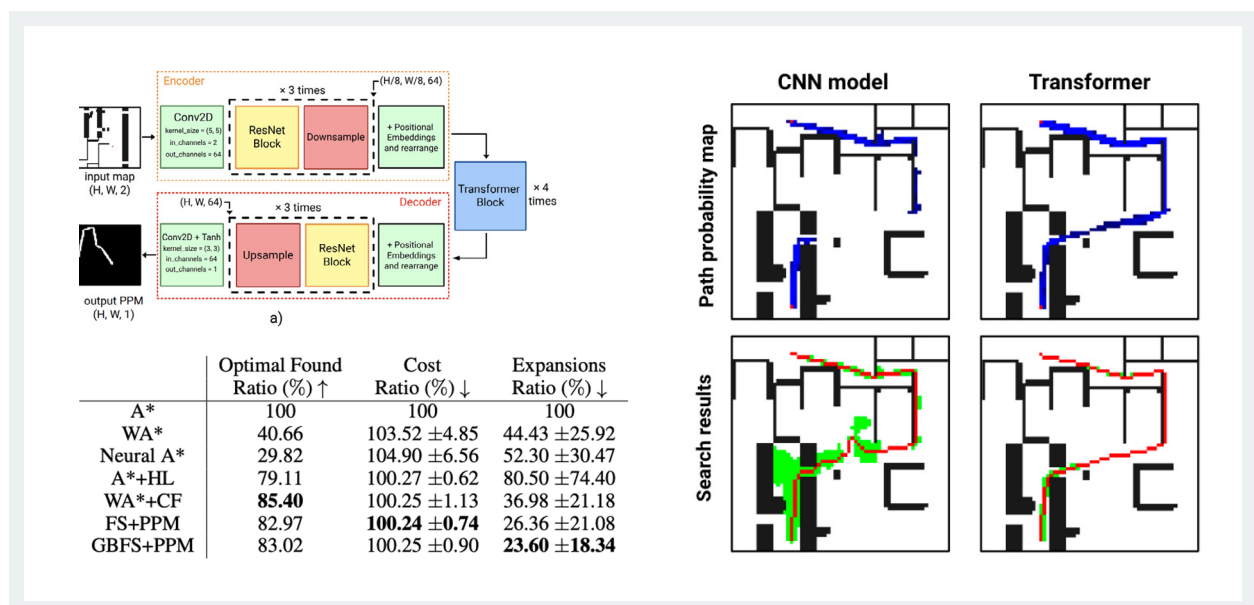
Предложен метод интеграции трансформерных нейросетевых моделей и алгоритмов эвристического поиска для задачи поиска пути на графе регулярной декомпозиции (планировании траектории по графу-сетке).


Работа опубликована на конференции AAAI 2023 (Core A*).

Предложены новые методы планирования траектории мобильного агента в среде с движущимися

препятствиями, учитывающие кинодинамические ограничения объекта управления, в том числе методы, комбинирующие обучаемые стратегии локального избегания столкновений и классические алгоритмы планирования (эвристический поиск, планирование на основе случайного семплирования).

Работы опубликованы на конференции AAAI 2023 (Core A*), журнале RA-L (Q1, Top-1 in Robotics).

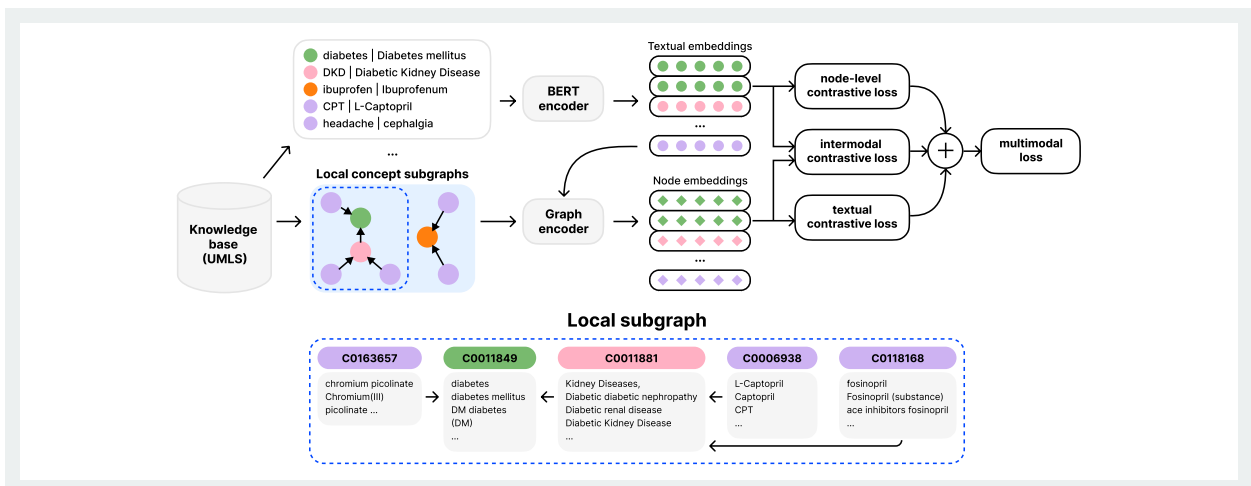




Прикладные
результаты

NLP

- Новый метод оценки неопределенности предсказаний внедрен в систему медицинской диагностики, которая разрабатывается Sber AI Lab. Система развернута в клиниках Москвы.
- Разработка методов для кросс-язычной и мультязычной детоксификации текстов. В рамках данной работы были представлены методы «перевода» токсичного текстового сообщения, например, содержащего нецензурные выражения в эквивалентное по смыслу сообщение, переписанное в нейтральном стиле. В нашем исследовании впервые были протестированы методы детоксификации в мультязычной и кросс-язычной постановке. Последняя подразумевает, что одновременно решаются две задачи — перевод и детоксификация текста. Результаты метода были представлены на международной конференции AACL-IJCNLP-2023 в Индонезии. Кроме этого, была принята заявка на проведение международного соревнования на тему кросс-язычной детоксификации текстов в рамках форума CLEF-2024.
- Предложена модель Graph-Enriched Biomedical Entity Representation Transformer (GEBERT), основанная на использовании графовых нейронных сетей архитектуры GraphSAGE и GAT для обогащения языковой модели дополнительной информацией из графа знаний. По результатам экспериментов на пяти англоязычных корпусах предложенная модель на основе архитектуры GraphSAGE превзошла существующие передовые модели нормализации именованных сущностей, не использующие граф знаний, на всех пяти корпусах по точности (accuracy). Опубликована статья “Graph-Enriched Biomedical Entity Representation Transformer” в сборнике международной конференции CLEF 2023 и статья “Graph-Enriched Biomedical Language Models: A Research Proposal” в сборнике AACL-IJCNLP SRW 2023 (CORE B).



Нейросимвольная интеграция

Группой под руководством Александра Панова получены следующие результаты

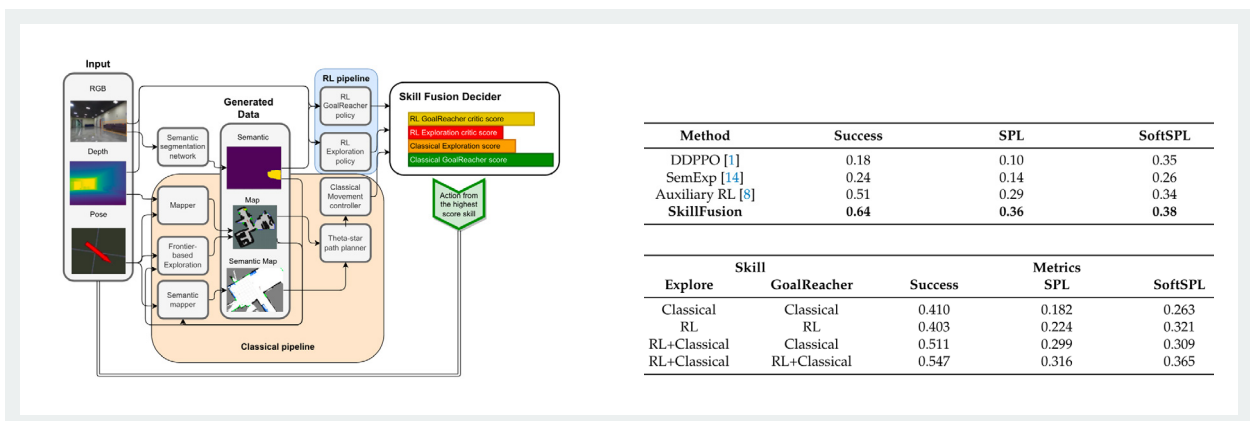


Александр Панов
Ведущий научный сотрудник

- Предложен ряд методов решения задачи многоагентного планирования траекторий (в децентрализованной постановке), опирающихся на интеграцию методов эвристического поиска и обучения с подкреплением (как на основе модели, так и без). Одна из публикаций по этой теме в журнале TNNLS (Q1, top-1 по нейросетевым методам).
- Разработан модуль планирования поведения интеллектуального агента по инструкциям на естественном языке на основе LLM. Апробация модуля проходила как на специально

сгенерированных данных, так и в составе системы управления робототехнической платформы на полигоне (совместно с Центром робототехники Сбера).

- Разработан подход SkillFusion, в котором классические и обучаемые алгоритмы были реализованы в качестве навыков агента, и был предложен модуль принятия решений о выборе навыка, который основан на оценке внутренней полезности каждого навыка. На основе данного результата было предложено лучшее решение на соревновании Habitat Challenge CVPR 2023 и опубликована статья в журнале Robotics (Q1).



AI в промышленности

Группой под руководством Ильи Макарова получены следующие результаты



Илья Макаров
Старший научный сотрудник

В области Industrial ML:

- Метод оценки состояния оборудования на основе комбинации моделей глубокого обучения и существующих моделей износа, интерпретируемый со стороны существующих диагностических правил.
- Метод организации и интерпретации низкоразмерного латентного пространства при решении задачи деградации индустриального оборудования.
- Генеративная диффузионная модель отказов оборудования с построением condition на основе близких в латентном пространстве траекторий деградации.
- Фреймворк и проведены масштабные эксперименты по защите систем диагностики неисправностей от состязательных атак.
- Unsupervised Fault Diagnosis по сенсорным данным химических предприятий:

- Получено SOTA решение, основанное на self-supervised предобучении и deep clustering, улучшающие метрики качества относительно существующих подходов на 20–30%;
- Показано, что разработанный способ предобучения позволяет моделям, обученным в semi-supervised сеттинге с экстремально малым количеством размеченных данных, достигать результатов близких SOTA моделей, обученных в supervised сеттинге.

В области Graph ML:

- Предложена SOTA модель TI-DC-GNN для работы с временными графами, на основе предложенных графов каузальности и следствия, что позволило улучшить качество и кратко- и долгосрочных предсказаний.
- Разработан новый SOTA SSL подход для предобучения моделей для предсказания свойств молекул, который сохраняет важные доменные знания.

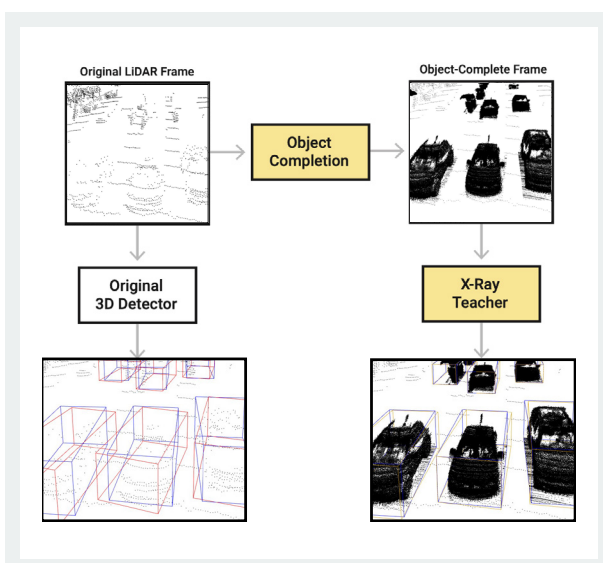
- Адаптирован метод мягкой кластеризации графа к задаче исследования в рексистемах, что позволило улучшить качество адаптации к изменениям в пользовательском поведении.

В области Computer Vision:

- Новые подходы к 3D детекции на основе дистилляции учителя на дополненных данных с тяжелым студентом в supervised и semi-supervised сеттингах в задаче автоматической генерации разметки. В результате:

a. Разработан подход, основанный на объединении последовательности кадров LIDAR с помощью алгоритмов регистрации облаков точек, который достигает SOTA результатов в semi-supervised сеттинге и стабильно улучшает качество моделей в supervised сеттинге;

b. Найдена ошибка в способах сравнения semi-supervised моделей 3D детекции на популярном датасете ONCE и предложен модифицированный способ обучения моделей для её решения.



В области Natural Language Processing:

- Организация работы над прототипом ассистента менеджера с использованием LLM на базе мультиагентного подхода (BabyAGI). Проект выполняется в сотрудничестве с Sber AI Lab.
- Установлена связь между теоретическими достижениями в области доступа к данным на основе онтологий и практическими приложениями машинного обучения, демонстрируя, что методы машинного обучения могут быть эффективным и нетрадиционным подходом для ответа на теоретические онтологические запросы с помощью покрывающей аксиомы.
- Предложен подход, который применяет графовые нейронные сети (GNN) к задаче индукции смысла слова (WSI) с использованием сети совместного появления, демонстрирует удовлетворительные результаты с низким потреблением вычислительных ресурсов.

В области Deep Reinforcement Learning:

- Разработка модели, которая включает в себя расширенные функции агрегации и механизмы внимания в сверточных сетях графов, что значительно повышает эффективность глубокого обучения с подкреплением при навигации по разреженным вознаграждениям, частично наблюдаемым трехмерным средам.
- Новые подходы в задаче интерпретируемой адаптации с реальных на синтетические данные с последующим обучением и обратной доменной адаптацией.

Биоинформатика

Группой под руководством Ольги Кардымон получены следующие результаты



Ольга Кардымон
Научный сотрудник

- Разработано и выложено в открытый доступ семейство языковых моделей для ДНК GENA — LM. Показана эффективность применения GENA для решения широкого спектра геномных задач. Согласно полученным данным, модели GENA-LM показывают характеристики не уступающие, а в некоторых случаях превосходящие модели мировых конкурентов (Google, Nvidia, Stanford), представленные в 2023 году. GENA-LM с высокой точностью предсказывает расположение ключевых элементов генома: промоторов генов, сайтов сплайсинга, терминаторов генов, регуляторных элементов. Аннотация различных геномных элементов на основе последовательности ДНК открывает возможность для создания инструментов для интерпретации мутаций в геноме человека.
- Совместно с Дмитрием Ветровым и коллегами из ВШЭ обучена модель для безусловной диффузионной генерации белков DIMA по текстовым (аминокислотным) данным. Разработан алгоритм сравнения генерируемых последовательностей по качеству и разнообразию как между кластерами, так и с конкурентами (EvoDiff-seq, nanoGPT, SeqDesign).

- Разработаны и опубликованы ряд моделей:

модель оценки стабильности белков при появлении одиночных мутаций «PROSTATA» (Bioinformatics, Q1, IF=5.8);

совместно разработанная с МГНЦ им. Бочкова модель и датасет для поиска клинически значимых регионов в не кодирующем белок начале последовательности генов (Nucleic Acids Research, Q1, IF = 19.16);

модель для интерпретации влияния на гены мутаций в не кодирующих белки участках ДНК в разных типах клеток (GigaScience, Q1, IF = 7.658).

Семейство языковых моделей GENA_LM



1–3 тыс.

скачиваний
каждый месяц с HF

15 883

скачиваний за 2023 год

2

международных семинара

Дизайн новых материалов

Группой под руководством Семена Буденного получены следующие результаты



Семен Буденный
Научный консультант

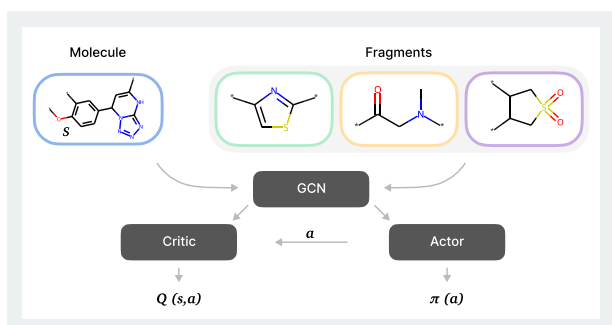
- Комплекс программного обеспечения ASCAD Light для автоматической детекции дефектов на солнечных ячейках и локализации узлов производственной линии, связанных с их образованием. Проведено успешное опытно-промышленное испытание разработанного комплекса в производственном ландшафте компании Hevel. Научные результаты опубликованы в журнале Solar Energy (Q1).
- Eco4cast, open-source библиотека на Python, снижающая углеродный след от обучения моделей AI за счет распределения облачных вычислений по временным интервалам и регионам с наименьшей текущей углеродной стоимостью электроэнергии. Научные результаты с экспериментальной проверкой эффективности библиотеки были опубликованы в журнале «Доклады математики» (Q2).
- Решения команды в области ESG — ASCAD Light, Eco2AI и Eco4cast — представлены на климатическом саммите ООН.
- Репозиторий Cd- and Zn-doped CsPbI₃ Energetics с датасетом и кодом постобработки предсказаний термодинамических свойств легированных перовскитов. Результаты, полученные с применением разработанного гибридного DFT/GNN подхода, опубликованы в журнале Computational Materials Science (Q1).
- С целью привлечения физико-химического научного сообщества к созданию AI-решений в области материаловедения созданы сайт и репозиторий 2DMD at a Glance, содержащие алгоритмы постобработки и упрощенного доступа к данным открытого датасета свойств точечных дефектов в двумерных материалах (2DMD).
- Прототип библиотеки для автоматизированного проектирования изделий из метаматериалов для 3D печати, в частности, имплантов костей.
- Разработана суррогатная модель симуляции литографического процесса в микроэлектронике, а также специализированные метрика и функция потерь, учитывающие как форму, так и топологию паттернов фотошаблонов. В рамках партнерства со Sber Robotics Lab модель успешно прошла первичный этап экспериментального тестирования.

Глубокое обучение в науках О ЖИЗНИ

Группой под руководством Артура Кадурина
получены следующие результаты



Артур Кадурина
Инженер-исследователь



- Совместно с командой Елены Тутубалиной разработаны новые модели для обработки медицинских текстов с учетом доменных знаний, представленных в виде графов (принята на конференцию CLEF), а так же создан бенчмарк языковых моделей для проверки их на химические знания (подана на ICLR).
- Разработан метод генерации новых молекулярных структур с высокой афинностью к белку-мишени. Статья принята в журнал TMLR и зарегистрировано соответствующее ПО.
- Разработан метод оптимизации молекулярных конформаций, позволяющий сократить вычислительную сложность на несколько порядков. Статья подана на конференцию A* ICLR.
- Совместно с командой Ольги Кардымон ведется разработка метода дизайна белков на основе обучения с подкреплением.
- Готовится к публикации вторая версия бенчмарка и датасета для решения квантово-химических задач nabraDFT.
- Совместно с командой Семена Буденного ведется разработка методов дизайна плоских материалов.
- Вместе с командой физиков из МГУ было проведено исследование по оценке магнитокалорического эффекта в бинарных материалах с легированием.

Нейроинтерфейсы

Группой под руководством Алексея Осадчего получены следующие результаты

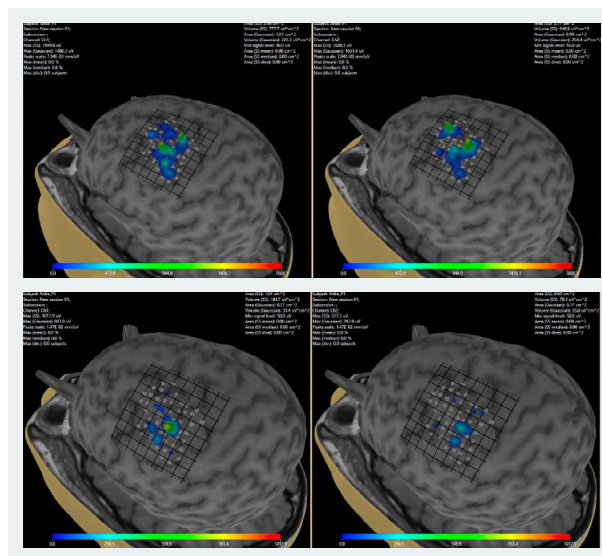


Алексей Осадчий
Ведущий научный сотрудник

- Внедрение результатов проекта «Мгновенная нейрообратная связь» в практику.

Реализация алгоритма мгновенной (задержка 3 мс, <3% от периода ритма) оценки фазы ритма мозга на борту отечественного электроэнцефалографа NVX52 и апробация в экспериментах по состояние-зависимой транскраниальной магнитной стимуляции.

Привязка момента стимуляции к фазе сенсомоторного ритма позволила в несколько раз повысить точность локализации кортикальных зон представления мышц кисти. В верхнем ряду представлены карты амплитуды вызванного моторного ответа, полученные стандартным методом. Нижний ряд — карты, полученные в режиме фазозависимой стимуляции в реальном времени с использованием модифицированного NVX-52.



Обучаемый интеллект

Группой под руководством Евгения Бурнаева получены следующие результаты



Евгений Бурнаев
Ведущий научный сотрудник

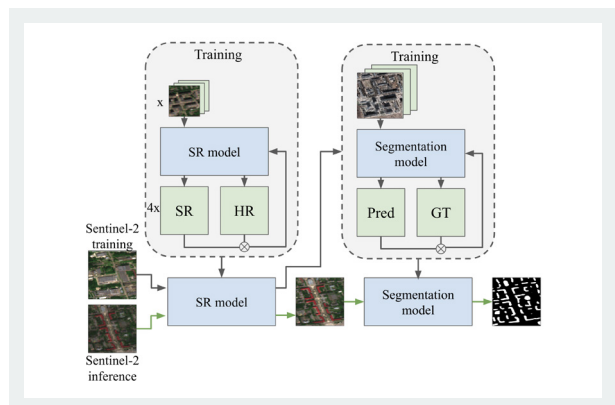
Новые общие алгоритмы ML (2 статьи Q1)

- Разноточный алгоритм автоматического поиска нейросетевых архитектур
- Кластеризация временных точечных процессов для обработки пользовательских данных
- Алгоритм обучения одноклассовой машины опорных векторов с привилегированной информацией

Приложения алгоритмов ML (9 статей Q1)

- Оптимальный выбор портфеля активов
- Обнаружения нейродегенеративных заболеваний на основе структурного МРТ и морфометрических признаков
- Обнаружения депрессивных расстройств
- Применение методов моделирования многообразий для анализа микробиома кишечника человека
- Сегментация зданий по данным ДЗЗ

- Моделирование наводнений по данным ДЗЗ
- Классификация типов землепользования по данным ДЗЗ
- Обзор приложений алгоритмов ML в задачах устойчивого развития



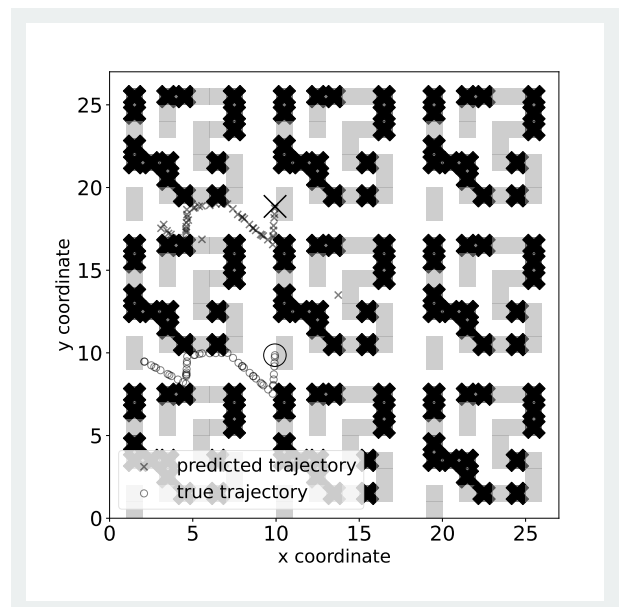
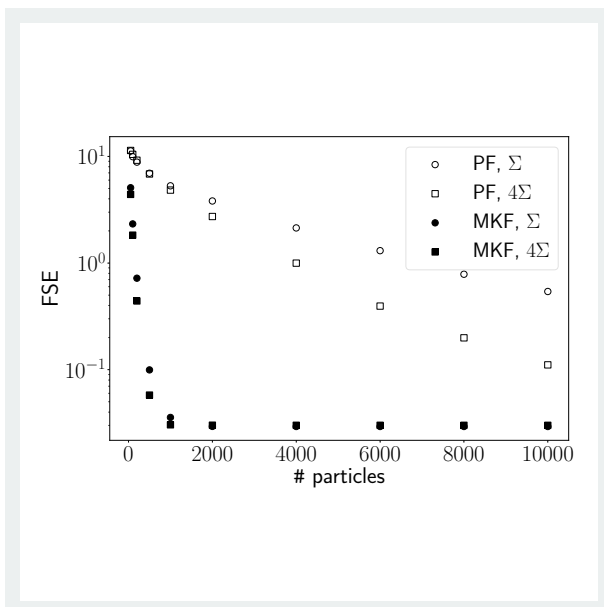
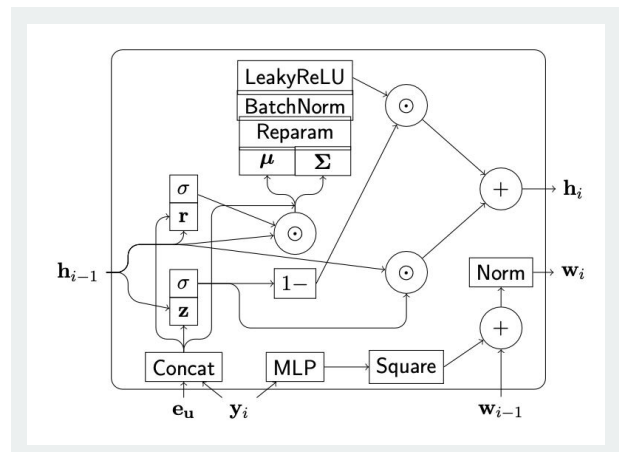
Вычислительный интеллект


Группой под руководством Ивана Оселедца получены следующие результаты



Иван Оселедец
Генеральный директор

- Предложена модификация архитектуры PFRNN, сфокусированная на решении задачи определения положения объектов и существенно более эффективная по памяти и качеству фильтрации.
- Разработан метод решения задачи локализации объектов по зашумлённым измерениям с помощью комбинации фильтра Калмана и частичного фильтра.



The image features a dark teal background with a repeating pattern of Hebrew characters in a lighter teal color. In the center, there is a stylized sun-like symbol composed of a central semi-circle and several radiating rectangular bars. The text 'Научные публикации' is positioned in the lower-left area in a white, sans-serif font.

Научные
публикации

Статьи на конференциях A*

8

NeurIPS

3

EMNLP

5

ICLR

2

ISMAR

3

ICML

4

ACL

1

ICCV

2

CVPR

2

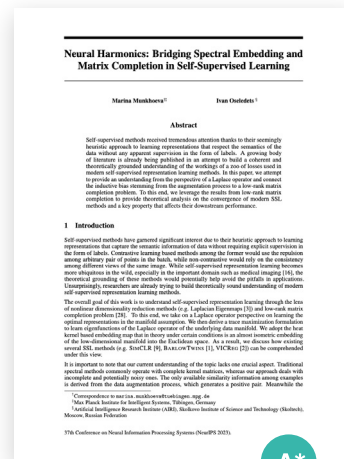
AAAI

NeurIPS

Neural Harmonics: Bridging Spectral Embedding and Matrix Completion in Self-Supervised Learning

Marina Munkhoeva, Ivan Oseledets

Self-supervised methods received tremendous attention thanks to their seemingly heuristic approach to learning representations that respect the semantics of the data without any apparent supervision in the form of labels. A growing body of literature is already being published in an attempt to build a coherent and theoretically grounded understanding of the workings of a zoo of losses used in modern self-supervised representation learning methods. In this paper, we attempt to provide an understanding from the perspective of a Laplace operator and connect the inductive bias stemming from the augmentation process to a low-rank matrix completion problem. To this end, we leverage the results from low-rank matrix completion to provide theoretical analysis on the convergence of modern SSL methods and a key property that affects their downstream performance. other methods on different downstream tasks.

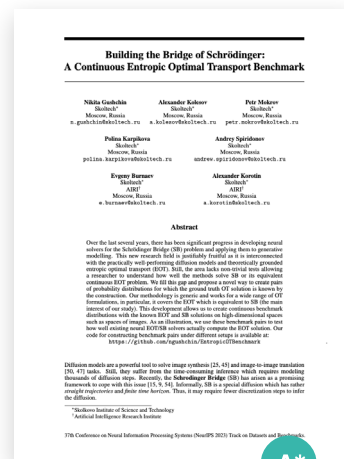


Ссылка на источник

Building the Bridge of Schrödinger: A Continuous Entropic Optimal Transport Benchmark

Nikita Gushchin, Alexander Kolesov, Petr Mokrov, Polina Karpikova, Andrei Spiridonov, Evgeny Burnaev, Alexander Korotin

Over the last several years, there has been significant progress in developing neural solvers for the Schrödinger Bridge (SB) problem and applying them to generative modelling. This new research field is justifiably fruitful as it is interconnected with the practically well-performing diffusion models and theoretically grounded entropic optimal transport (EOT). Still, the area lacks non-trivial tests allowing a researcher to understand how well the methods solve SB or its equivalent continuous EOT problem. We fill this gap and propose a novel way to create pairs of probability distributions for which the ground truth OT solution is known by the construction. Our methodology is generic and works for a wide range of OT formulations, in particular, it covers the EOT which is equivalent to SB (the main interest of our study). This development allows us to create continuous benchmark distributions with the known EOT and SB solutions on high-dimensional spaces such as spaces of images. As an illustration, we use these benchmark pairs to test how well existing neural EOT/SB solvers actually compute the EOT solution.



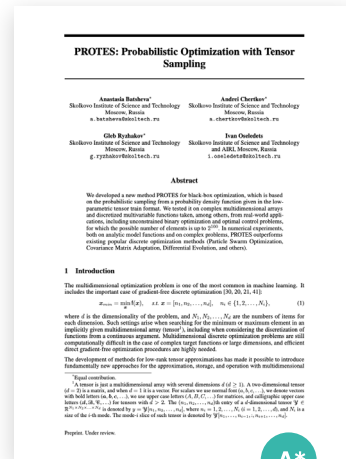
Ссылка на источник

NeurIPS

PROTES: Probabilistic Optimization with Tensor Sampling

Anastasia Batsheva, Andrei Chertkov, Gleb Ryzhakov, Ivan Oseledets

We developed a new method PROTES for black-box optimization, which is based on the probabilistic sampling from a probability density function given in the low-parametric tensor train format. We tested it on complex multidimensional arrays and discretized multivariable functions taken, among others, from real-world applications, including unconstrained binary optimization and optimal control problems, for which the possible number of elements is up to 2^{100} . In numerical experiments, both on analytic model functions and on complex problems, PROTES outperforms existing popular discrete optimization methods (Particle Swarm Optimization, Covariance Matrix Adaptation, Differential Evolution, and others).



[Ссылка на источник](#)

Star-Shaped Denoising Diffusion Probabilistic Models

Andrey Okhotin, Dmitry Molchanov, Vladimir Arkhipkin, Grigory Bartosh, Viktor Ohanesian, Aibek Alanov, Dmitry Vetrov

Denoising Diffusion Probabilistic Models (DDPMs) provide the foundation for the recent breakthroughs in generative modeling. Their Markovian structure makes it difficult to define DDPMs with distributions other than Gaussian or discrete. In this paper, we introduce Star-Shaped DDPM (SS-DDPM). Its star-shaped diffusion process allows us to bypass the need to define the transition probabilities or compute posteriors. We establish duality between star-shaped and specific Markovian diffusions for the exponential family of distributions and derive efficient algorithms for training and sampling from SS-DDPMs. In the case of Gaussian distributions, SS-DDPM is equivalent to DDPM. However, SS-DDPMs provide a simple recipe for designing diffusion models with distributions such as Beta, von Mises-Fisher, Dirichlet, Wishart and others, which can be especially useful when data lies on a constrained manifold. We evaluate the model in different settings and find it competitive even on image data, where Beta SS-DDPM achieves results comparable to a Gaussian DDPM.



[Ссылка на источник](#)

NeurIPS

Intrinsic Dimension Estimation for Robust Detection of AI-Generated Texts

Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Serguei Barannikov, Irina Piontkovskaya, Sergey Nikolenko, Evgeny Burnaev

Rapidly increasing quality of AI-generated content makes it difficult to distinguish between human and AI-generated texts, which may lead to undesirable consequences for society. Therefore, it becomes increasingly important to study the properties of human texts that are invariant over different text domains and varying proficiency of human writers, can be easily calculated for any language, and can robustly separate natural and AI-generated texts regardless of the generation model and sampling method. In this work, we propose such an invariant for human-written texts, namely the intrinsic dimensionality of the manifold underlying the set of embeddings for a given text sample. We show that the average intrinsic dimensionality of fluent texts in a natural language is hovering around the value 9 for several alphabet-based languages and around 7 for Chinese, while the average intrinsic dimensionality of AI-generated texts for each language is ≈ 1.5 lower, with a clear statistical separation between human-generated and AI-generated distributions. This property allows us to build a score-based artificial text detector. The proposed detector's accuracy is stable over text domains, generator models, and human...

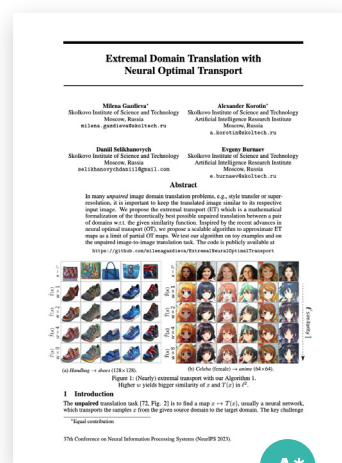


[Ссылка на источник](#)

Extremal Domain Translation with Neural Optimal Transport

Milena Gazdieva, Alexander Korotin, Daniil Selikhanovych, Evgeny Burnaev

In many unpaired image domain translation problems, e.g., style transfer or super-resolution, it is important to keep the translated image similar to its respective input image. We propose the extremal transport (ET) which is a mathematical formalization of the theoretically best possible unpaired translation between a pair of domains w.r.t. the given similarity function. Inspired by the recent advances in neural optimal transport (OT), we propose a scalable algorithm to approximate ET maps as a limit of partial OT maps. We test our algorithm on toy examples and on the unpaired image-to-image translation task.



[Ссылка на источник](#)

EMNLP

Kandinsky: an Improved Text-to-Image Synthesis with Image Prior and Latent Diffusion

Anton Razzhigaev, Arseniy Shakhmatov, Anastasia Maltseva, Vladimir Arkhipkin, Igor Pavlov, Ilya Ryabov, Angelina Kuts, Alexander Panchenko, Andrey Kuznetsov, Denis Dimitrov

Text-to-image generation is a significant domain in modern computer vision and has achieved substantial improvements through the evolution of generative architectures. Among these, there are diffusion-based models that have demonstrated essential quality enhancements. These models are generally split into two categories: pixel-level and latent-level approaches. We present Kandinsky, a novel exploration of latent diffusion architecture, combining the principles of the image prior models with latent diffusion techniques. The image prior model is trained separately to map text embeddings to image embeddings of CLIP. Another distinct feature of the proposed model is the modified MoVQ implementation, which serves as the image autoencoder component. Overall, the designed model contains 3.3B parameters. We also deployed a user-friendly demo system that supports diverse generative modes such as text-to-image generation, image fusion, text and image fusion, image variations generation, and text-guided inpainting/outpainting. Additionally, we released the source code and...

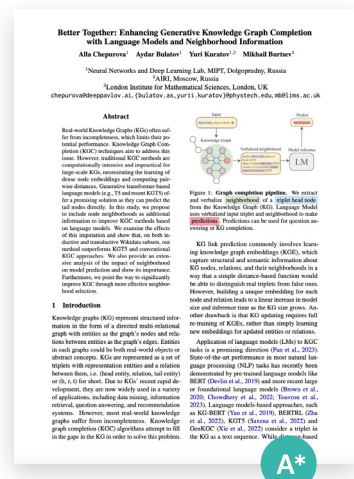


Ссылка на источник

Better Together: Enhancing Generative Knowledge Graph Completion with Language Models and Neighborhood Information

Alla Chepurova, Aydar Bulatov, Yuri Kuratov, Mikhail S. Burtsev

Real-world Knowledge Graphs (KGs) often suffer from incompleteness, which limits their potential performance. Knowledge Graph Completion (KGC) techniques aim to address this issue. However, traditional KGC methods are computationally intensive and impractical for large-scale KGs, necessitating the learning of dense node embeddings and computing pairwise distances. Generative transformer-based language models (e.g., T5 and recent GPT5) offer a promising solution as they can predict the tail nodes directly. In this study, we propose to include node neighborhoods as additional information to improve KGC methods based on language models. We examine the effects of this imputation and show that, on both inductive and transductive Wikidata subsets, our method outperforms GPT5 and conventional KGC approaches...



Ссылка на источник

EMNLP

LM-Polygraph: Uncertainty Estimation for Language Models

Ekaterina Fadeeva, Roman Vashurin, Akim Tsvigun, Artem Vazhentsev, Sergey Petrakov, Kirill Fedyanin, Daniil Vasilev, Elizaveta Goncharova, Alexander Panchenko, Maxim Panov, Timothy Baldwin, Artem Shelmanov

Recent advancements in the capabilities of large language models (LLMs) have paved the way for a myriad of groundbreaking applications in various fields. However, a significant challenge arises as these models often «hallucinate», i.e., fabricate facts without providing users an apparent means to discern the veracity of their statements. Uncertainty estimation (UE) methods are one path to safer, more responsible, and more effective use of LLMs. However, to date, research on UE methods for LLMs has been focused primarily on theoretical rather than engineering contributions. In this work, we tackle this issue by introducing LM-Polygraph, a framework with implementations of a battery of state-of-the-art UE methods for LLMs in text generation tasks, with unified program interfaces in Python. Additionally, it introduces an extendable benchmark for consistent evaluation of UE techniques by researchers, and a demo web application that enriches the standard chat dialog with confidence scores, empowering end-users to discern unreliable responses. LM-Polygraph is compatible with the most recent LLMs, including BLOOMz, LLaMA-2, ChatGPT, and GPT-4, and is designed to support future releases of similarly-styled LLMs.



[Ссылка на источник](#)

Learning topology-preserving data representations

Ilya Trofimov, Daniil Cherniavskii, Eduard Tulchinskii, Nikita Balabin, Evgeny Burnaev, Serguei Barannikov

We propose a method for learning topology-preserving data representations (dimensionality reduction). The method aims to provide topological similarity between the data manifold and its latent representation via enforcing the similarity in topological features (clusters, loops, 2D voids, etc.) and their localization. The core of the method is the minimization of the Representation Topology Divergence (RTD) between original high-dimensional data and low-dimensional representation in latent space. RTD minimization provides closeness in topological features with strong theoretical guarantees. We develop a scheme for RTD differentiation and apply it as a loss term for the autoencoder. The proposed method «RTD-AE» better preserves the global structure and topology of the data manifold than state-of-the-art competitors as measured by linear correlation, triplet distance ranking accuracy, and Wasserstein distance between persistence barcodes.

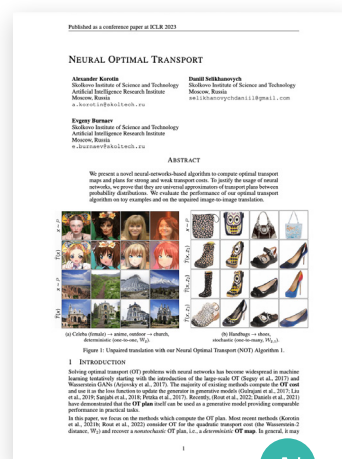


Ссылка на источник

Neural Optimal Transport

Alexander Korotin, Daniil Selikhanovych, Evgeny Burnaev

We present a novel neural-networks-based algorithm to compute optimal transport maps and plans for strong and weak transport costs. To justify the usage of neural networks, we prove that they are universal approximators of transport plans between probability distributions. We evaluate the performance of our optimal transport algorithm on toy examples and on the unpaired image-to-image translation.

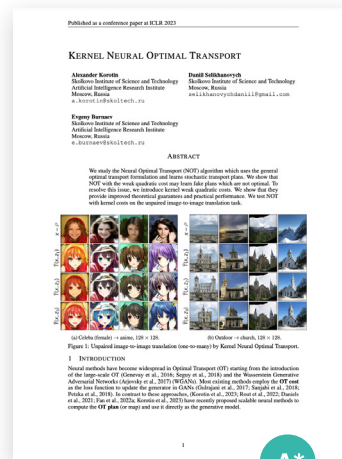


Ссылка на источник

Kernel Neural Optimal Transport

Alexander Korotin, Daniil Selikhanovych, Evgeny Burnaev

We study the Neural Optimal Transport (NOT) algorithm which uses the general optimal transport formulation and learns stochastic transport plans. We show that NOT with the weak quadratic cost may learn fake plans which are not optimal. To resolve this issue, we introduce kernel weak quadratic costs. We show that they provide improved theoretical guarantees and practical performance. We test NOT with kernel costs on the unpaired image-to-image translation task.

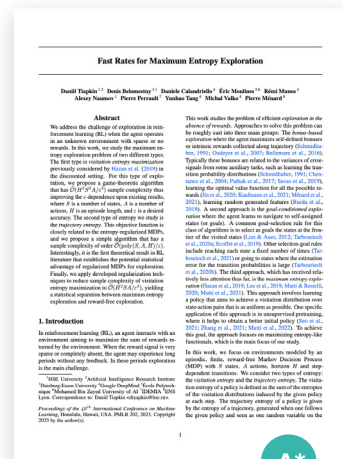


[Ссылка на источник](#)

Fast Rates for Maximum Entropy Exploration

Daniil Tiapkin, Denis Belomestny, Daniele Calandriello, Eric Moulines, Remi Munos, Alexey Naumov, Pierre Perrault, Yunhao Tang, Michal Valko, Pierre Menard

We address the challenge of exploration in reinforcement learning (RL) when the agent operates in an unknown environment with sparse or no rewards. In this work, we study the maximum entropy exploration problem of two different types. The first type is visitation entropy maximization previously considered by Hazan et al. (2019) in the discounted setting. For this type of exploration, we propose a game-theoretic algorithm that has $\tilde{O}(H^3 S^2 A / \epsilon^2)$ sample complexity thus improving the ϵ -dependence upon existing results, where S is a number of states, A is a number of actions, H is an episode length, and ϵ is a desired accuracy. The second type of entropy we study is the trajectory entropy. This objective function is closely related to the entropy-regularized MDPs, and we propose a simple algorithm that has a sample complexity of order $\tilde{O}(\text{poly}(S, A, H) / \epsilon)$. Interestingly, it is the first theoretical result in RL literature that establishes the potential statistical advantage of regularized MDPs for exploration. Finally, we apply developed regularization techniques to reduce sample complexity of visitation entropy maximization to $\tilde{O}(H^2 S A / \epsilon^2)$, yielding a statistical separation between maximum entropy exploration and reward-free exploration.



[Ссылка на источник](#)

General Covariance Data Augmentation for Neural PDE Solvers

Vladimir Fanaskov, Tianchi Yu, Alexander Rudikov, Ivan V. Oseledets

The growing body of research shows how to replace classical partial differential equation (PDE) integrators with neural networks. The popular strategy is to generate the input-output pairs with a PDE solver, train the neural network in the regression setting, and use the trained model as a cheap surrogate for the solver. The bottleneck in this scheme is the number of expensive queries of a PDE solver needed to generate the dataset. To alleviate the problem, we propose a computationally cheap augmentation strategy based on general covariance and simple random coordinate transformations. Our approach relies on the fact that physical laws are independent of the coordinate choice, so the change in the coordinate system preserves the type of a parametric PDE and only changes PDE's data (e.g., initial conditions, diffusion coefficient). For tried neural networks and partial differential equations, proposed augmentation improves test error by 23% on average. The worst observed result is a 17% increase in test error for multilayer perceptron, and the best case is a 80% decrease for dilated residual network.

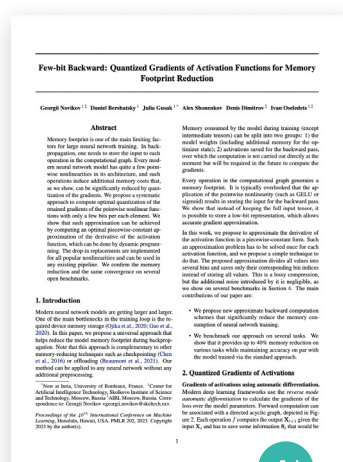


Ссылка на источник

Few-bit Backward: Quantized Gradients of Activation Functions for Memory Footprint Reduction

Georgii Novikov, Daniel Bershtatsky, Julia Gusak, Alex Shonenkov, Denis Dimitrov, Ivan Oseledets

Memory footprint is one of the main limiting factors for large neural network training. In backpropagation, one needs to store the input to each operation in the computational graph. Every modern neural network model has quite a few pointwise nonlinearities in its architecture, and such operations induce additional memory costs that, as we show, can be significantly reduced by quantization of the gradients. We propose a systematic approach to compute optimal quantization of the retained gradients of the pointwise nonlinear functions with only a few bits per each element. We show that such approximation can be achieved by computing an optimal piecewise-constant approximation of the derivative of the activation function, which can be done by dynamic programming. The drop-in replacements are implemented for all popular nonlinearities and can be used in any existing...



Ссылка на источник

Efficient Out-of-Domain Detection for Sequence to Sequence Models

Artem Vazhentsev, Akim Tsvigun, Roman Vashurin, Sergey Petrakov, Daniil Vasilev, Maxim Panov, Alexander Panchenko, Artem Shelmanov

Sequence-to-sequence (seq2seq) models based on the Transformer architecture have become a ubiquitous tool applicable not only to classical text generation tasks such as machine translation and summarization but also to any other task where an answer can be represented in a form of a finite text fragment (e.g., question answering). However, when deploying a model in practice, we need not only high performance but also an ability to determine cases where the model is not applicable. Uncertainty estimation (UE) techniques provide a tool for identifying out-of-domain (OOD) input where the model is susceptible to errors. State-of-the-art UE methods for seq2seq models rely on computationally heavyweight and impractical deep ensembles. In this work, we perform an empirical investigation of various novel UE methods for large pre-trained seq2seq models T5 and BART on three tasks: machine translation, text summarization, and question answering. We apply computationally lightweight density-based UE methods to seq2seq models and show that they often outperform heavyweight deep ensembles on the task of OOD detection.



Ссылка на источник

Layerwise universal adversarial attack on NLP models

O Tsymboi, D Malaev, A Petrovskii, I Oseledets

In this work, we examine the vulnerability of language models to universal adversarial triggers (UATs). We propose a new white-box approach to the construction of layerwise UATs (LUATs), which searches the triggers by perturbing hidden layers of a network. On the example of three transformer models and three datasets from the GLUE benchmark, we demonstrate that our method provides better transferability in a model-to-model setting with an average gain of 9.3% in the fooling rate over the baseline. Moreover, we investigate triggers transferability in the task-to-task setting. Using small subsets from the datasets similar to the target tasks for choosing a perturbed layer, we show that LUATs are more efficient than vanilla UATs by 7.1% in the fooling rate.



Ссылка на источник

Hybrid Uncertainty Quantification for Selective Text Classification in Ambiguous Tasks

Artem Vazhentsev, Gleb Kuzmin, Akim Tsvigun, Alexander Panchenko, Maxim Panov, Mikhail Burtsev, Artem Shelmanov

Many text classification tasks are inherently ambiguous, which results in automatic systems having a high risk of making mistakes, in spite of using advanced machine learning models. For example, toxicity detection in user-generated content is a subjective task, and notions of toxicity can be annotated according to a variety of definitions that can be in conflict with one another. Instead of relying solely on automatic solutions, moderation of the most difficult and ambiguous cases can be delegated to human workers. Potential mistakes in automated classification can be identified by using uncertainty estimation (UE) techniques. Although UE is a rapidly growing field within natural language processing, we find that state-of-the-art UE methods estimate only epistemic uncertainty and show poor performance, or under-perform trivial methods for ambiguous tasks such as toxicity detection. We argue that in order to create robust uncertainty estimation methods for ambiguous tasks it is necessary to account also for aleatoric uncertainty. In this paper, we propose a new uncertainty...



Ссылка на источник

A System for Answering Simple Questions in Multiple Languages

A Razzhigaev, M Salnikov, V Malykh, P Braslavski, A Panchenko

Our research focuses on the most prevalent type of queries – simple questions — exemplified by questions like “What is the capital of France?”. These questions reference an entity such as “France”, which is directly connected (one hop) to the answer entity “Paris” in the underlying knowledge graph (KG). We propose a multilingual Knowledge Graph Question Answering (KGQA) technique that orders potential responses based on the distance between the question’s text embeddings and the answer’s graph embeddings. A system incorporating this novel method is also described in our work. Through comprehensive experimentation using various English and multilingual datasets and two KGs — Freebase and Wikidata — we illustrate the comparative advantage of the proposed method across diverse KG embeddings and languages. This edge is apparent even against robust baseline systems, including seq2seq QA models, search-based solutions and intricate rule-based pipelines. Interestingly, our research underscores that even advanced...

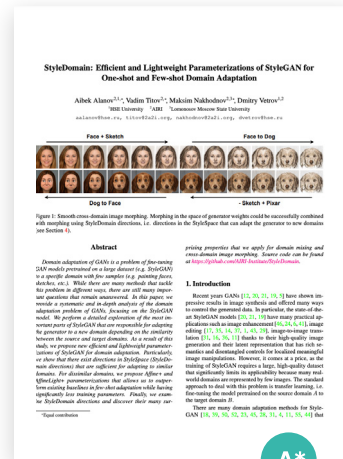


Ссылка на источник

StyleDomain: Efficient and Lightweight Parameterizations of StyleGAN for One-shot and Few-shot Domain Adaptation

Aibek Alanov, Vadim Titov, Maksim Nakhodnov, Dmitry P. Vetrov

Domain adaptation of GANs is a problem of fine-tuning GAN models pretrained on a large dataset (e.g. StyleGAN) to a specific domain with few samples (e.g. painting faces, sketches, etc.). While there are many methods that tackle this problem in different ways, there are still many important questions that remain unanswered. In this paper, we provide a systematic and in-depth analysis of the domain adaptation problem of GANs, focusing on the StyleGAN model. We perform a detailed exploration of the most important parts of StyleGAN that are responsible for adapting the generator to a new domain depending on the similarity between the source and target domains. As a result of this study, we propose new efficient and lightweight parameterizations of StyleGAN for domain adaptation. Particularly, we show that there exist directions in StyleSpace (StyleDomain directions) that are sufficient for adapting to similar domains. For dissimilar domains, we propose Affine+ and AffineLight+ parameterizations that allows us to outperform existing baselines in few-shot adaptation while having significantly less training parameters. Finally, we examine StyleDomain directions and discover their many surprising properties that we apply for domain mixing and cross-domain image morphing.



[Ссылка на источник](#)

Sphere-Guided Training of Neural Implicit Surfaces

Andreea Dogaru, Andrei-Timotei Ardelean, Savva Ignatyev, Egor Zakharov, Evgeny Burnaev

In recent years, neural distance functions trained via volumetric ray marching have been widely adopted for multi-view 3D reconstruction. These methods, however, apply the ray marching procedure for the entire scene volume, leading to reduced sampling efficiency and, as a result, lower reconstruction quality in the areas of high-frequency details. In this work, we address this problem via joint training of the implicit function and our new coarse sphere-based surface reconstruction. We use the coarse representation to efficiently exclude the empty volume of the scene from the volumetric ray marching procedure without additional forward passes of the neural surface network, which leads to an increased fidelity of the reconstructions compared to the base systems. We evaluate our approach by incorporating it into the training procedures of several implicit surface modeling methods and observe uniform improvements across both synthetic and real-world datasets.

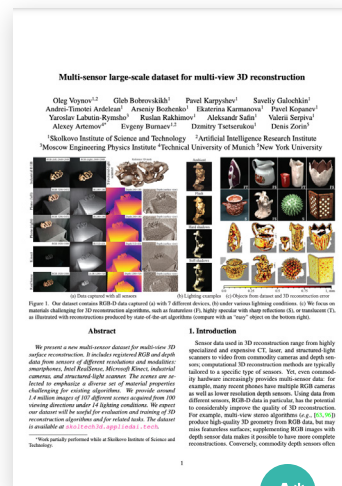


[Ссылка на источник](#)

Multi-sensor large-scale dataset for multi-view 3D reconstruction

Oleg Voynov, Gleb Bobrovskikh, Pavel Karpyshev, Saveliy Galochkin, Andrei-Timotei Ardelean, Arseniy Bozhenko, Ekaterina Karmanova, Pavel Kopanev, Yaroslav Labutin-Rymsho, Ruslan Rakhimov, Aleksandr Safin, Valerii Serpiva, Alexey Artemov, Evgeny Burnaev, Dzmitry Tsetserukou, Denis Zorin

We present a new multi-sensor dataset for multi-view 3D surface reconstruction. It includes registered RGB and depth data from sensors of different resolutions and modalities: smartphones, Intel RealSense, Microsoft Kinect, industrial cameras, and structured-light scanner. The scenes are selected to emphasize a diverse set of material properties challenging for existing algorithms. We provide around 1.4 million images of 107 different scenes acquired from 100 viewing directions under 14 lighting conditions. We expect our dataset will be useful for evaluation and training of 3D reconstruction algorithms and for related tasks.

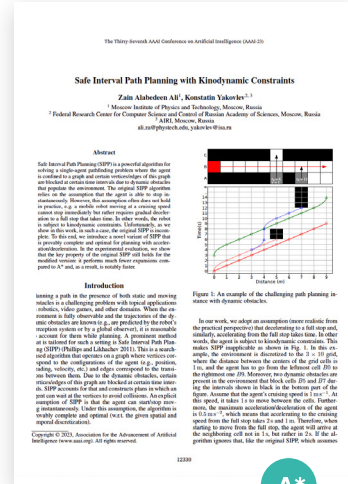


[Ссылка на источник](#)

Safe Interval Path Planning with Kinodynamic Constraints

Zain Alabedeen Ali, Konstatin Yakovlev

Safe Interval Path Planning (SIPP) is a powerful algorithm for solving a single-agent pathfinding problem where the agent is confined to a graph and certain vertices/edges of this graph are blocked at certain time intervals due to dynamic obstacles that populate the environment. The original SIPP algorithm relies on the assumption that the agent is able to stop instantaneously. However, this assumption often does not hold in practice, e.g. a mobile robot moving at a cruising speed cannot stop immediately but rather requires gradual deceleration to a full stop that takes time. In other words, the robot is subject to kinodynamic constraints. Unfortunately, as we show in this work, in such a case, the original SIPP is incomplete. To this end, we introduce a novel variant of SIPP that is provably complete and optimal for planning with acceleration/deceleration. In the experimental evaluation, we show that the key property of the original SIPP still holds for the modified version: it performs much fewer expansions compared to A* and, as a result, is notably faster.

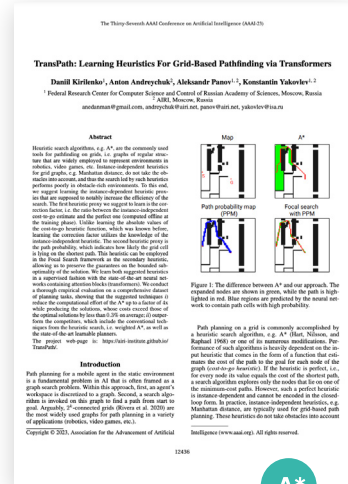


[Ссылка на источник](#)

TransPath: Learning Heuristics for Grid-Based Pathfinding via Transformers

Daniil Kirilenko, Anton Andreychuk, Aleksandr Panov, Konstantin Yakovlev

Heuristic search algorithms, e.g. A*, are the commonly used tools for pathfinding on grids, i.e. graphs of regular structure that are widely employed to represent environments in robotics, video games, etc. Instance-independent heuristics for grid graphs, e.g. Manhattan distance, do not take the obstacles into account, and thus the search led by such heuristics performs poorly in obstacle-rich environments. To this end, we suggest learning the instance-dependent heuristic proxies that are supposed to notably increase the efficiency of the search. The first heuristic proxy we suggest to learn is the correction factor, i.e. the ratio between the instance-independent cost-to-go estimate and the perfect one (computed offline at the training phase). Unlike learning the absolute values of the cost-to-go heuristic function, which was known before, learning the correction factor utilizes the knowledge of the instance-independent heuristic. The second heuristic proxy is the path probability, which indicates how likely the grid cell is lying on the shortest path. This heuristic can be employed in the Focal Search framework as the secondary heuristic, allowing us to preserve the guarantees on the bounded sub-optimality of the solution. We learn both suggested heuristics in a supervised fashion with the state-of-the-art neural networks containing attention blocks (transformers). We conduct a thorough empirical evaluation on a comprehensive dataset of planning tasks, showing that the suggested techniques i) reduce the computational effort of the A* up to a factor of 4x while producing the solutions, whose costs exceed those of the optimal solutions by less than 0.3% on average; ii) outperform the competitors, which include the conventional techniques from the heuristic search, i.e. weighted A*, as well as the state-of-the-art learnable planners. and, as a result, is notably faster.

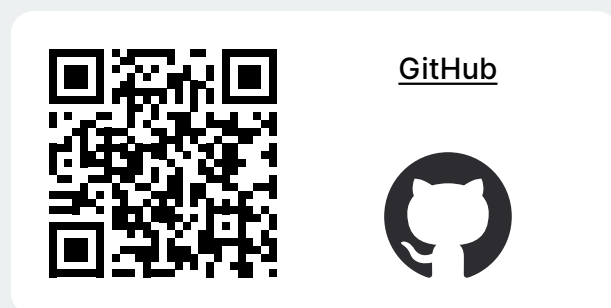
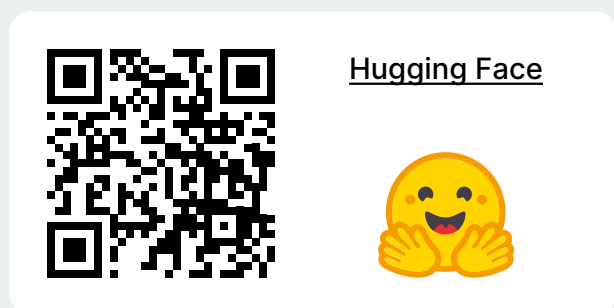


[Ссылка на источник](#)

Публикации AIRI



Open source репозитории AIRI





Мероприятия
и выступления

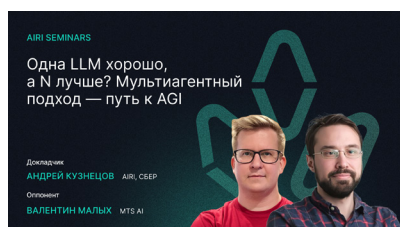
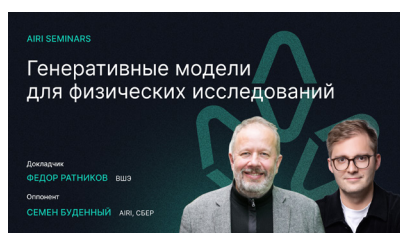


AIRI Seminars

AIRI Seminars — это научный диалог на равных и знакомство профессионального сообщества с достижениями в области искусственного интеллекта.

Семинар призван популяризировать и распространять в профессиональной среде принципы и ценности, которых придерживается Институт, а также продвигать идеи, реализующие миссию AIRI: создание универсальных систем искусственного интеллекта, решающих задачи реального мира.

В качестве докладчиков или оппонентов на семинар приглашаются ведущие специалисты в области искусственного интеллекта из России и из-за рубежа, которые рассказывают и конструктивно критикуют исследовательские работы. В этом году семинары проводились не только онлайн, но и офлайн. В 2023 году был проведен 21 семинар.



Научные семинары AIRI доступны для всех желающих на YouTube-канале AIRI.



Статистика

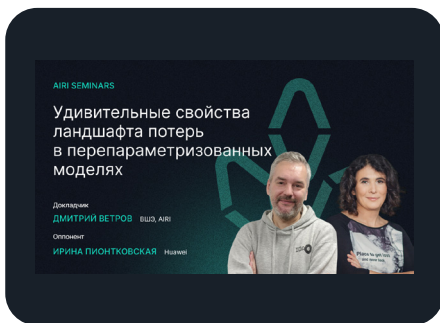
21
семинар

23
организации

12 758
просмотров

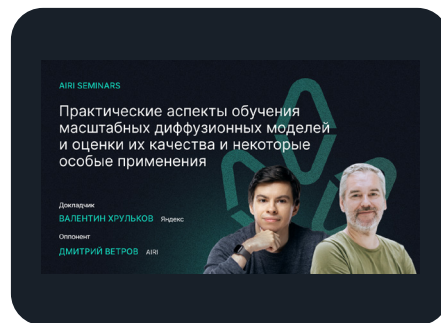
Самые просматриваемые семинары

Удивительные свойства ландшафта потерь в перепараметризованных моделях



[Ссылка на семинар](#)

Обучение больших диффузионных моделей и оценки качества



[Ссылка на семинар](#)

Команда



Александр
Панов



Алексей
Скрынник



Александра
Бройтман



Екатерина
Мамонтова



Юрий
Яровиков

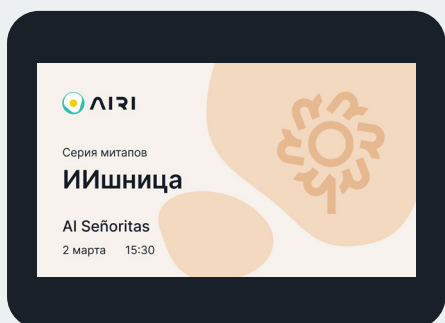


Николай
Кущенко

Серия митапов ИИшница

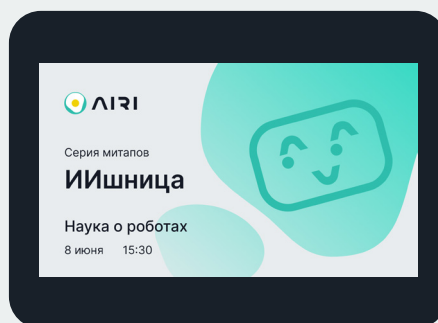
В 2023 году мы развивали спецпроект «ИИшница» — это серия онлайн-митапов, где ученые рассказывают про искусственный интеллект в рамках 20-ти минутного научного доклада.

В этом году темами ИИшницы стали: AI Señoritas, «Наука о роботах», «Диффузионные модели и их расширения» и новогодний выпуск про статьи на конференции NeurIPS 2023.



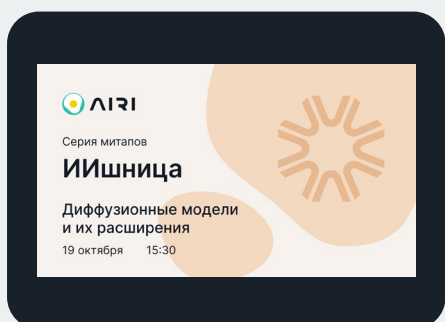
2 марта
AI Señoritas

[Ссылка на митап](#)



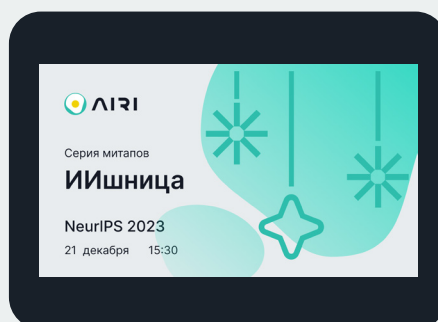
8 июня
Наука о роботах

[Ссылка на митап](#)



19 октября
Диффузионные
модели и их
расширения

[Ссылка на митап](#)

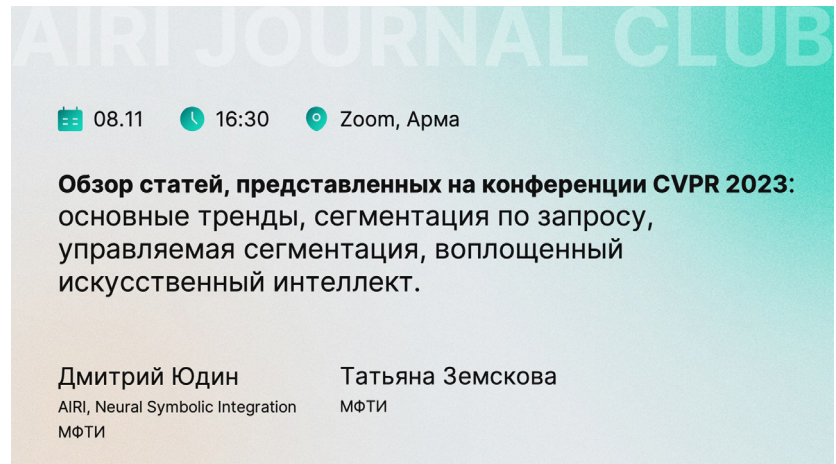


21 декабря
NeurIPS 2023

[Ссылка на митап](#)

Журнальный клуб

Исследователи AIRI проводят журнальный клуб, где обсуждают последние научные статьи в области искусственного интеллекта. В 2023 году было проведено 13 журнальных клубов.



AIRI JOURNAL CLUB

📅 08.11 🕒 16:30 📍 Zoom, Арма

Обзор статей, представленных на конференции CVPR 2023:
основные тренды, сегментация по запросу, управляемая сегментация, воплощенный искусственный интеллект.

Дмитрий Юдин
AIRI, Neural Symbolic Integration
МФТИ

Татьяна Земскова
МФТИ

Образовательные школы

Исследователи AIRI приняли участие в образовательных школах, включая летнюю школу по ML в биоинформатике ВШЭ, Летнюю школу машинного обучения SMILES и Летнюю школу РАИИ 2023.

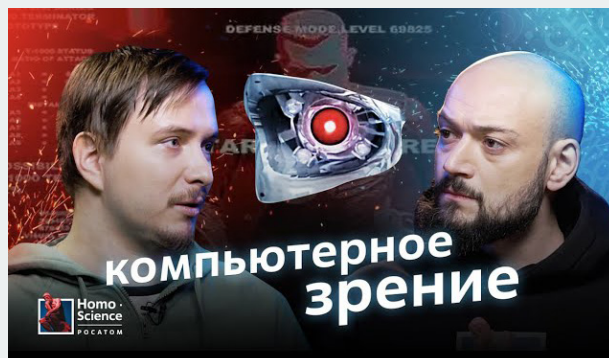


Ученые AIRI приняли участие в подкастах



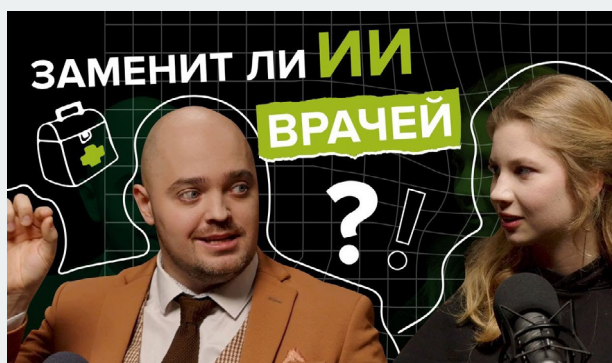
[Ссылка на подкаст](#)

Алексей Осадчий вместе с ведущим подкаста «Сергея и микрофон» рассуждают про силу мысли и расширение возможностей мозга.



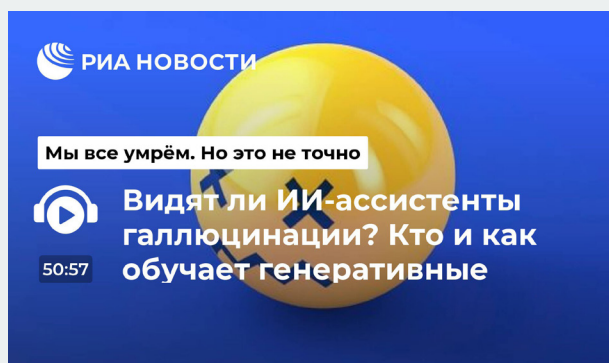
[Ссылка на подкаст](#)

Выпуск подкаста «Мозги есть» с руководителем группы «AI в промышленности» Ильей Макаровым.



[Ссылка на подкаст](#)

Выпуск подкаста от Академии AI с ведущим научным сотрудником Виктором Гомболовским про медицину и medtech-стартапы.



[Ссылка на подкаст](#)

О том, как по запросу «включить экспертизу» AI, рассказал руководитель научной группы Fusion Brain Андрей Кузнецов.

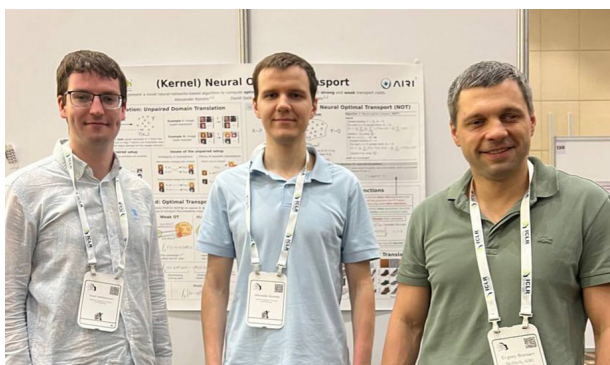
Ученые AIRI представили свои результаты более чем на 25 научных конференциях



IJCNLP-AAACL 2023



EMNLP в Сингапуре



Конференция ICLR



Международный симпозиум IEEE по смешанной и дополненной реальности ISMAR



ACL 2023 в Торонто



ICML 2023 на Гавайских островах



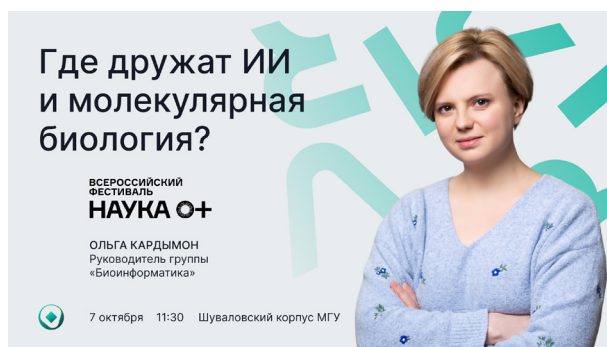
Interspeech 2023 в Дублине



Fall into ML 2023



OpenTalks AI 2023 в Ереване

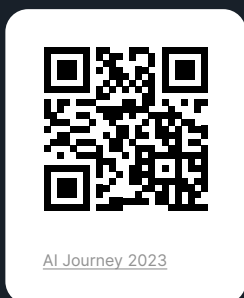
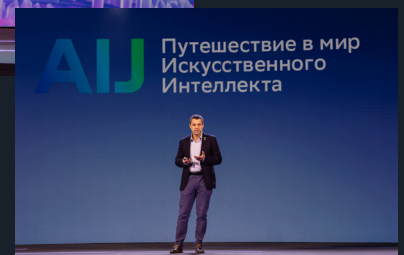


Наука 0+

AI Journey 2023

17

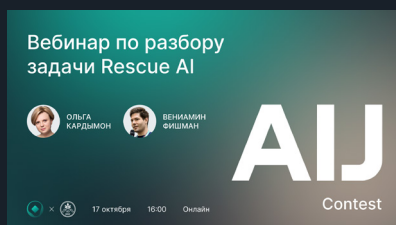
исследователей
выступили
на конференции
с докладами





14

ученых
представили
свои постеры



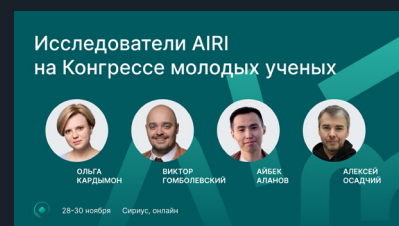
Исследователи AIRI подготовили две задачи на соревнование AIJ Contest

Strong Intelligence — задача про мультимодальных чатботов, Rescue AI — задача про детекцию хромосомных перестроек по Hi-C картам.



Наука 0+ в Китае

Иван Оселедец, Евгений Бурнаев, Семен Буденный, Константин Яковлев, Андрей Кузнецов выступили с докладами в сессии AIJ на фестивале науки в Китае.



Конгресс молодых ученых

29 и 30 ноября на Конгрессе молодых ученых исследователи AIRI приняли участие в нескольких дискуссиях и представили 3 доклада.



Лето с AIRI 2023

Лето с AIRI

Летняя школа Института искусственного интеллекта AIRI для студентов и аспирантов — это глубокое погружение в работу с широким спектром современных методов искусственного интеллекта и машинного обучения.

Участники провели две недели с ведущими учеными из AIRI, МФТИ, ВШЭ, Сколтеха и других авторитетных научно-исследовательских организаций и ВУЗов.

Применить полученные в ходе лекций знания участники смогли на практических семинарах и в проектной деятельности, по итогам которой представили доклады о результатах работы.



700

заявок

35

преподавателей

80

студентов

2

недели

33

проекта

132

академических
часа

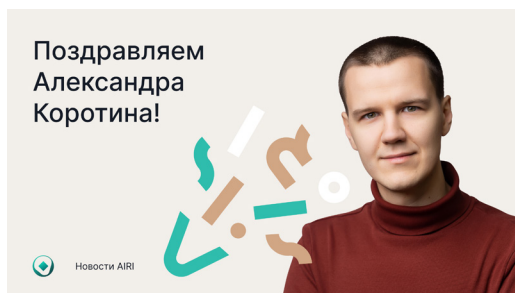
70

научных
постеров



Партнеры
проекта





Награды

- Команда исследователей из AIRI, МФТИ и ФИЦ ИУ РАН под руководством Александра Панова и Алексея Староверова заняла первое место в международном соревновании по обучению роботов навигации внутри помещений.
- Научный сотрудник AIRI и руководитель исследовательской группы в Сколтехе Александр Коротин стал лауреатом премии за расширение границ познаний об искусственном интеллекте.
- Лауреатами премии Yandex ML Prize стали руководитель группы «AI в промышленности» Илья Макаров, научный сотрудник группы «Вероятностные методы машинного обучения» Айбек Аланов и научный сотрудник группы FusionBrain Антон Разжигает.

AIRI LEGO клуб

В этом году неформальный клуб по интересам сотрудников AIRI продолжил активно развиваться. Коллеги собирали Lego, знакомились, приглашали друзей и играли в квизы.



Портретиист

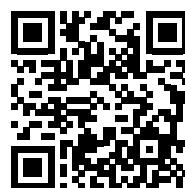
Взгляд нейросети, стиль мастеров

Благодаря методу группы «Вероятностные методы машинного обучения» можно обучить AI 5 тысячам новых стилей, потратив на это столько же ресурсов, сколько раньше заняло бы знакомство модели только с одним типом картинок. В этом году мы воплотили научное достижение в простой интерфейс и предложили пользователям прикоснуться к алгоритмам и математическим конструкциям с помощью Портретииста. В течение 30 дней любой желающий мог сгенерировать видео с портретом в стиле великих мастеров.

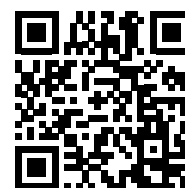


7 353

видео
сгенерировано
за ноябрь



[Статья](#)



[Модели](#)

О нас пишут и говорят

ТАСС, Forbes, РБК, Коммерсантъ, Российская Газета, Газета.Ру, N+1, Хайтек, Код Дурова, РИА, RTVI, Радио «Маяк», НТВ, Москва 24 и многие другие

Прочсть и послушать:



БиоМолекула

История развития
искусственного интеллекта
и его пришествия
в биологию



Российская Газета

Иван Оселедец о границах
возможностей AI



Forbes

Дмитрий Ветров
о регулировании AI



RTVI

Евгений Бурнаев
в передаче «900 секунд»



РБК

Илья Макаров о нейросетях
в повседневной жизни



Газета.Ру

Ольга Кардымон
о биоинформатике



N+1

Это Нейробаза!
AI-тезаурус



Подкаст «Сереза и микрофон»

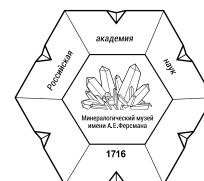
Алексей Осадчий
о силе мысли и других
возможностях мозга



Проект «Топ Технологии»

Андрей Кузнецов
об AI

Партнерства и коллаборации



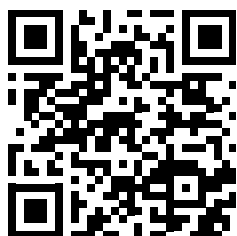
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ИМ. Н.Э. БАУМАНА

АССОЦИАЦИЯ
ЦИФРОВЫЕ ТЕХНОЛОГИИ
В ПРОМЫШЛЕННОСТИ





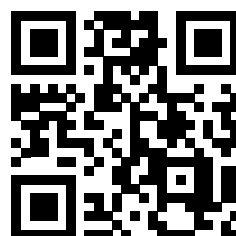
Ученые AIRI В СОЦИАЛЬНЫХ СЕТЯХ



Ivan Oseledets'
Channel



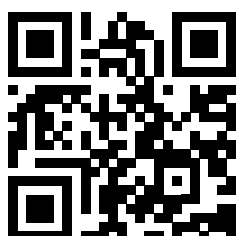
Telegram-канал
Ивана Оселедца



Manvel's Channel



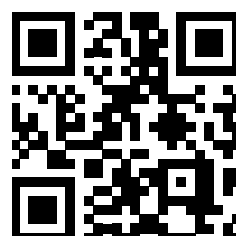
Telegram-канал
Манвела Аветисяна



AI for biology



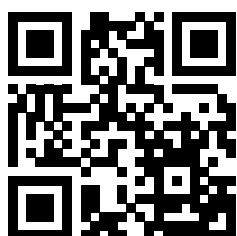
Telegram-канал
Ольги Кардымон



Complete AI



Telegram-канал
Андрея Кузнецова



AbstractDL



Telegram-канал
Антон Разжигаяев

Контакты

Сайт

airi.net

Соцсети



[airi_research_institute](https://t.me/airi_research_institute)



[artificial-intelligence-research-institute](https://www.linkedin.com/company/artificial-intelligence-research-institute)



[AIRIInstitute](https://www.youtube.com/AIRIInstitute)



[AIRI_inst](https://x.com/AIRI_inst)



[Airi_institute](https://vk.com/Airi_institute)

Адрес

Москва, Кутузовский пр-кт, д. 32, корп. 1, помещ. 4.В.08

Москва, Нижний Сусальный пер. 5, стр. 19