

Анализ больших данных и генеративный ИИ в социальных и гуманитарных науках

Вячеслав Гойко

Юлия Александрова



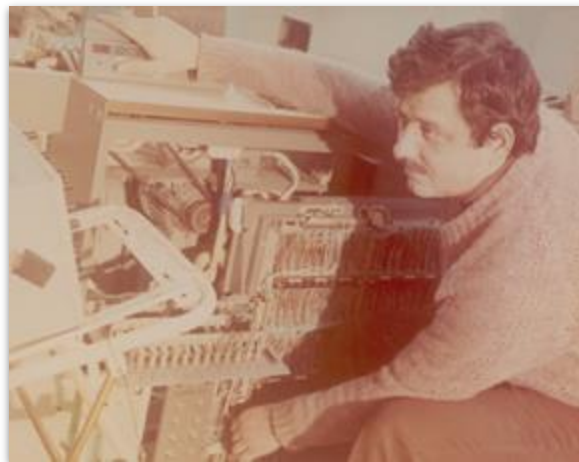
Немного истории

Вековая история вычислительных наук в ТГУ



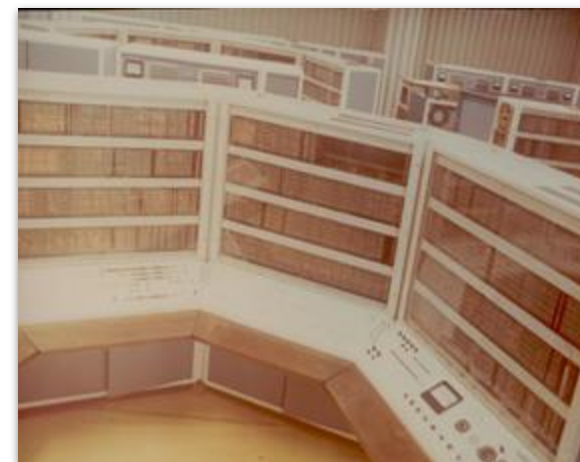
1930

Создание и деятельность при Томском университете вычислительного бюро, а затем и Научно-исследовательского института математики и механики (НИИ ММ)



1941

Преобразование в военные годы НИИ ММ в спецотдел СФТИ. После войны - бурный рост и развитие новых направлений. Внедрение первых аналоговых и цифровых вычислительных машин



1968

Возрождение НИИ ММ как НИИПММ в 1968 г. В 1977 на основе первой в Томской области ЭВМ "БЭСМ-6" построен вычислительный центр при НИИПММ. В 1990 г. Запуск суперкомпьютера "Эльбрус 1-КБ"

1993

Суперкомпьютер СКИФ Cyberia



2007

2024



В 2007 г. на базе вычислительного центра при НИИПММ запущен суперкомпьютер “СКИФ Cyberia”. В 2016 г. расширение спектра задач суперкомпьютера на анализ больших данных и машинное обучение. Подключение вузов участников Университетского консорциума исследователей больших данных

Консорциум

Университетский консорциум исследователей больших данных — одно из самых крупных международных объединений образовательных учреждений в России и СНГ

78 образовательных и научных организаций

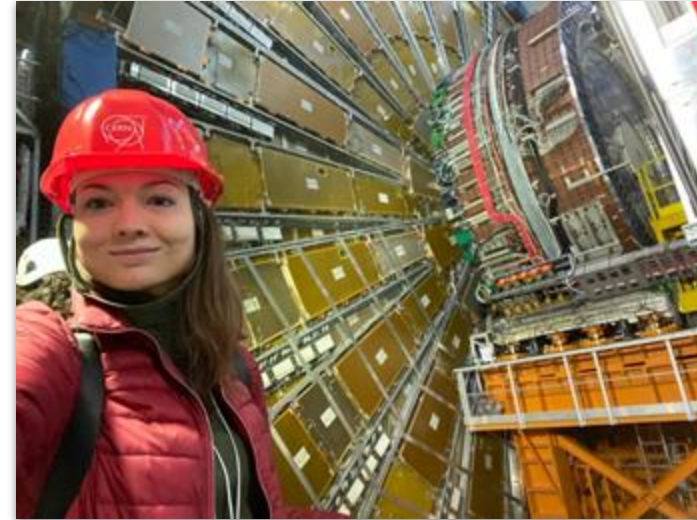


Сообщество объединяет исследователей от Севастополя до Камчатки, от Архангельска до Ташкента

Фундаментальные науки

Исследование элементарных частиц

Анализ данных Большого адронного коллайдера: с помощью суперкомпьютера CYBERIA научный коллектив решал задачу поиска распадов бозона Хиггса на фермионы с использованием машинного обучения и глубоких нейронных сетей.



Прогнозирование погоды

Анализ данных наземного и спутникового зондирования атмосферы инструментами машинного обучения для прогнозирования характеристик и эволюции облаков верхнего яруса



Инженерия

Поиск нефти

Расчет показателей разработки месторождений нефти на полномасштабных моделях с высокой детализацией геологической информации.



Проектирование радиоэлектроники

Расчеты для создания системы мониторинга состояния обогатительной фабрики, исследование НДС рефлектора антенны при воздействии ветровой нагрузки.



Автоматизированное проектирование БПЛА

Создание цифрового двойника и прототипирование беспилотников на основе облачной платформы и мощностей суперкомпьютера





Генеративный ИИ



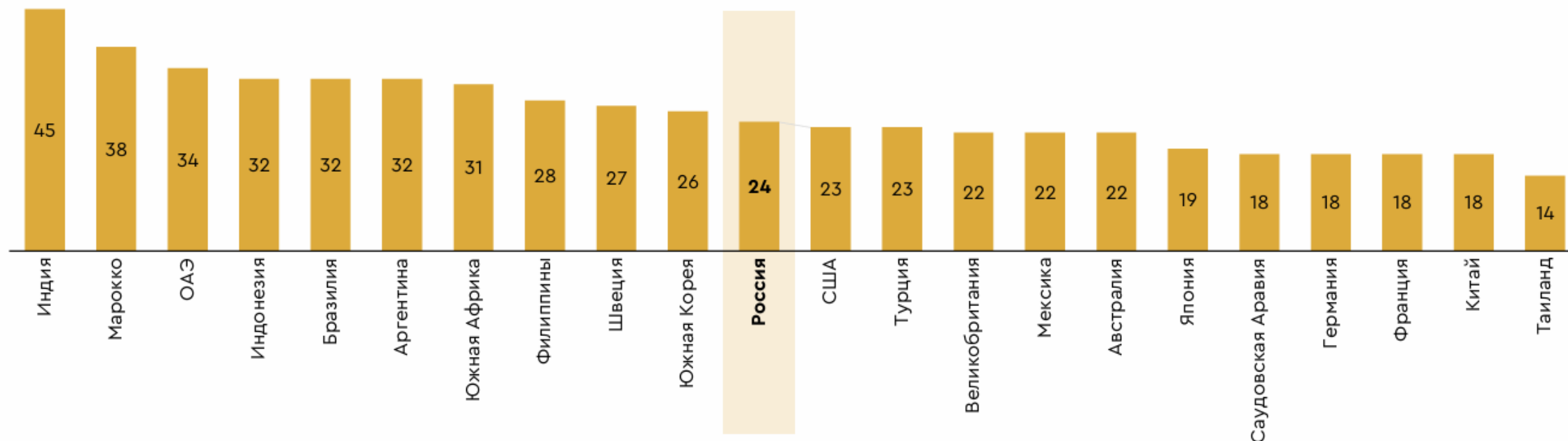
Яков и Партнёры × РОМИР

Новое российское общество: граждане и искусственный интеллект

Август 2024 г.

По использованию генеративного ИИ Россия не уступает другим странам

Доля респондентов, использующих генИИ, по странам, % (сводные данные разных опросов)¹



Если судить по результатам опроса о ChatGPT, который провела компания BCG, то жители России неплохо осведомлены о генИИ и применяют его довольно активно: российские показатели выше, чем в США, Великобритании и Германии

Причем в развивающихся странах исследование BCG охватило преимущественно обеспеченных респондентов, поэтому результаты, полученные в этих государствах, могут быть завышены

Источник: BCG, BCG CCI Global Consumer Sentiment Survey 2023

Вопрос: «Which of the following statements apply to you regarding ChatGPT?»

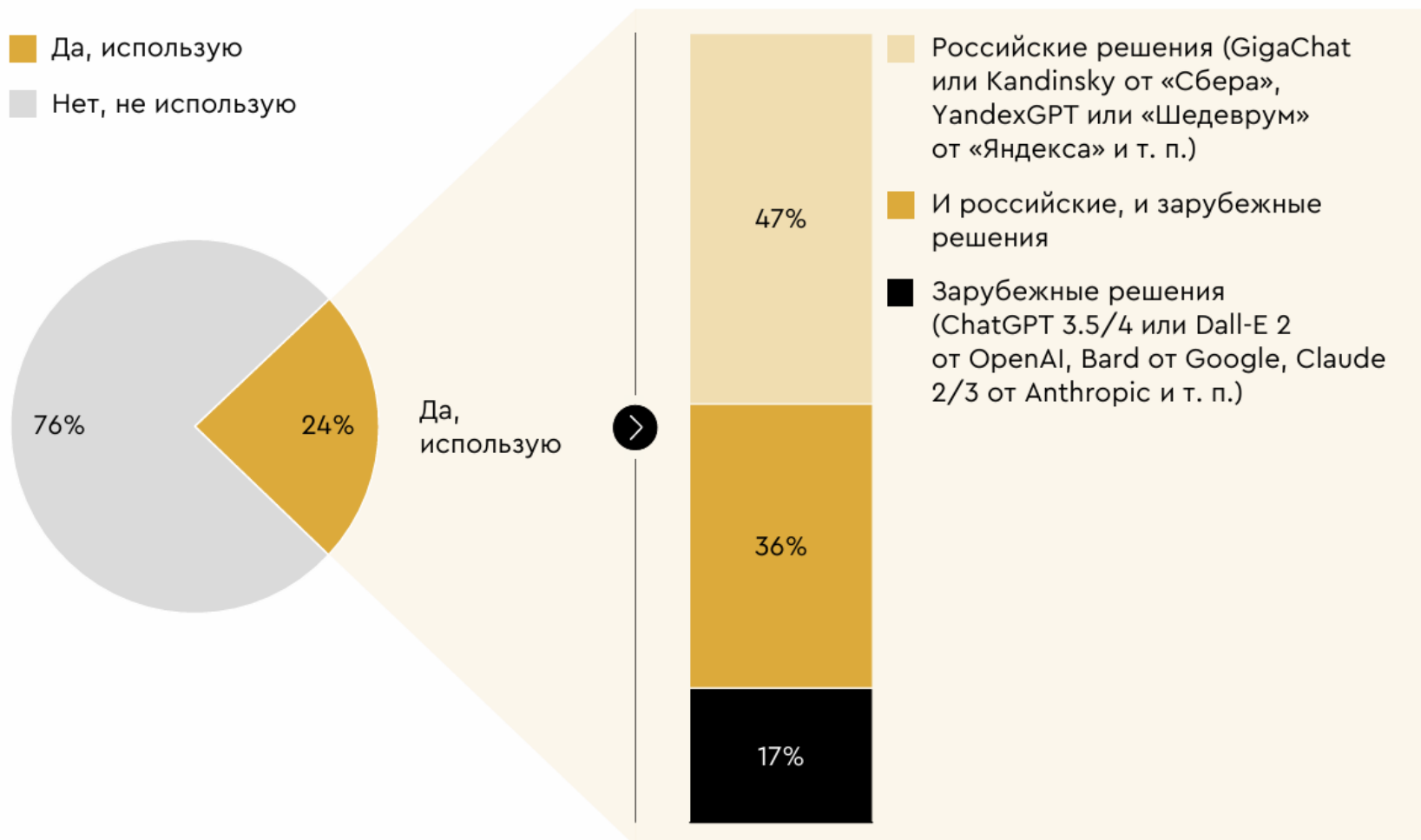
Варианты ответа: «I have never heard about it», «I have heard about it», «I have used it»

In developing nations, BCG survey captured more affluent consumers

Каждый четвертый респондент использует генеративный искусственный интеллект

«Используете ли Вы в личных или рабочих целях генеративный искусственный интеллект (ИИ)?», %

■ Да, использую
■ Нет, не использую



24% респондентов используют генеративный ИИ (генИИ), при этом каждый второй из них (47%) пользуется российскими решениями

Предпочтение зарубежным продуктам отдают 8–23% пользователей по возрастным группам

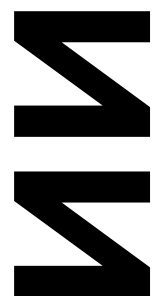
Треть респондентов используют и российские, и зарубежные решения

ГенИИ практически одинаково часто используют в столицах и в малых городах (27% и 24% соответственно)

Респонденты с высоким уровнем дохода в 3 раза чаще (12% против 3–4%) выбирают западные решения в области генИИ

Источник: «Срез потребительских настроений в РФ» – совместное исследование компании «Яков и Партнёры» и холдинга «РОМИР», май 2024 г.

Вопрос: «Используете ли Вы в личных или рабочих целях генеративный искусственный интеллект (ИИ)?»



- **комплекс технологических решений**, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их (**Указ Президента РФ от 10.10.2019 N 490**)
- **техническая система**, которая для определенных человеком целей может делать прогнозы, выдавать рекомендации или принимать решения, влияющие на реальную или виртуальную среду. (**H.R.6216 — National Artificial Intelligence Initiative Act of 2020**)
- **системы**, демонстрирующие разумное поведение, которое заключается в анализе окружающей среды и совершении действий – с некоторой степенью автономии – для достижения поставленных целей. (**EUROPEAN COMMISSION. COMMUNICATION FROM THE COMMISSION. Artificial Intelligence for Europe**)

Этапы развития ИИ

Золотой век ИИ, появление компьютерного зрения, первого нейрокомпьютера и развитие математики для ИИ

Возвращение интереса к ИИ. Появление экспертных систем и многослойных нейронных сетей

1944

1956 - 1974

1974 - 1980

1980 - 1987

1987 - 1994

1994 –

Предложена идея искусственных нейронных сетей



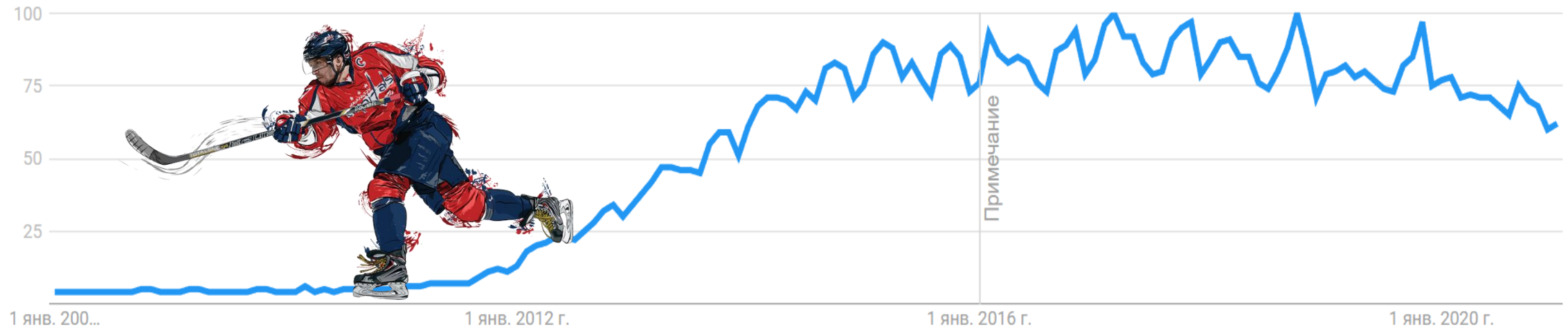
Н.В.



Большие данные

Упоминаемость термина «Big Data»

Динамика популярности ?



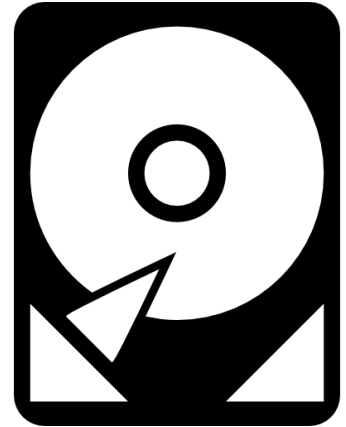
01



02



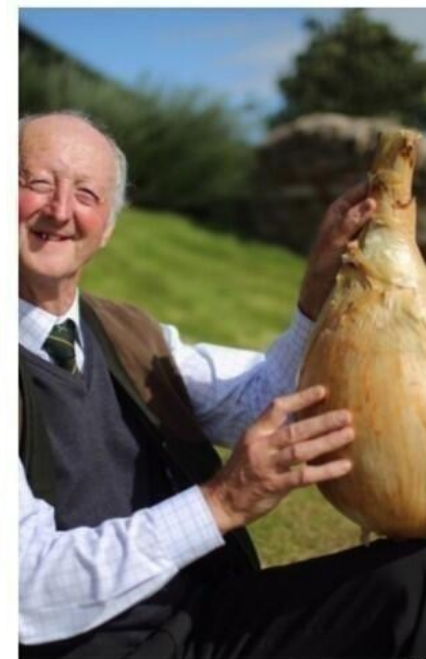
03





Концепция 3V

Volume (Объем)



Volume (Объем)

Velocity (Скорость)

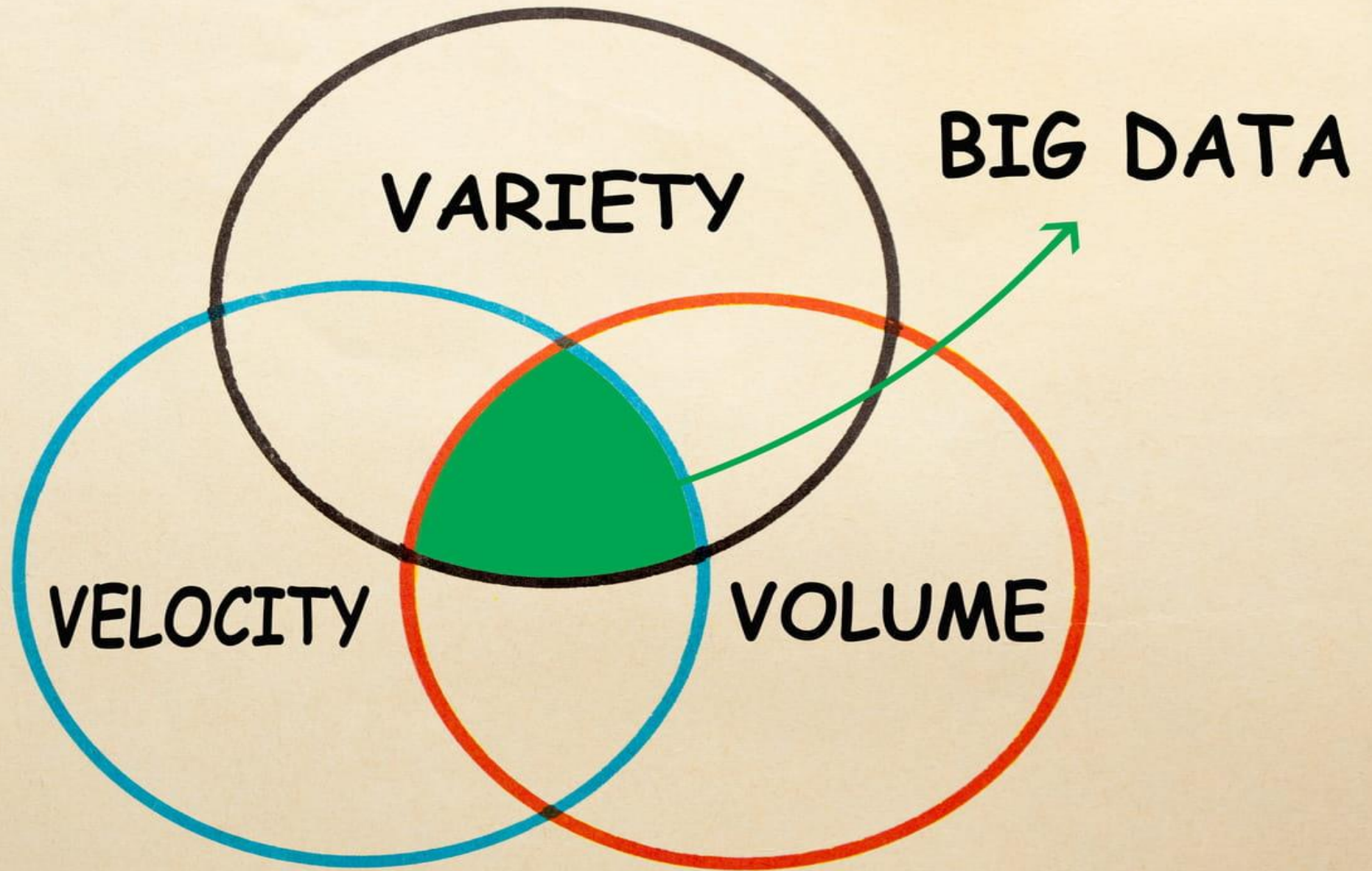



Volume (Объем)

Velocity (Скорость)

Variety (Разнообразие)







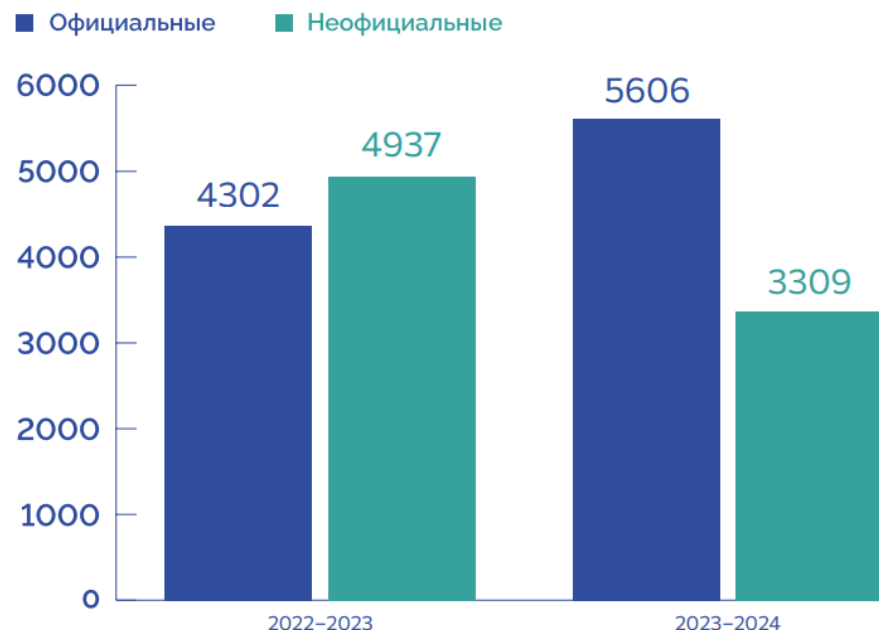
**Как найти мнения
студентов и не утонуть
в спаме?**



COVID-19

Источники данных

В рамках исследования проанализировано 1.8 млн постов и 3.5 млн. комментариев из 8 915 пабликов российских университетов в социальной сети «ВКонтакте». Данные собраны за период: май 2023г. - июнь 2024г.



Соотношение официальных и неофициальных пабликов которые были задействованы в исследовании



На исследовательско-коммуникационной платформе «Неравнодушный человек» с мая по июнь 2024 года проводился онлайн-опрос «Качество образования» среди студентов, преподавателей и административного состава российских государственных вузов, в котором были территориально охвачены 8 федеральных округов и новые регионы.

Даты онлайн-опроса: май–июнь 2024 года

111 365

Количество респондентов, прошедших опрос

Разметка и очистка данных

90% времени в проекте – очистка и ручная категоризация

Классификация по темам:

- Важные аспекты учебного процесса
- Социально-бытовые условия
- Внеучебная деятельность
- Финансы
- Лояльность к университету
- Стресс

А также определение тональности сообщений, позволяющих понять отношение автора о высказываемой теме

Data sets in tutorials



Data sets in the wild



Трудоемкий препроцессинг

Важно - не срезать релевант, 0.5% - 1% релевантных постов и комментариев из всего объема.

По общей структуре мусор и релевант могут быть очень похожи друг на друга.

Кого-нибудь отчислили из-за посещаемости?

Как принимает экзамен Иванов?

Сколько стоит общежитие в месяц ???

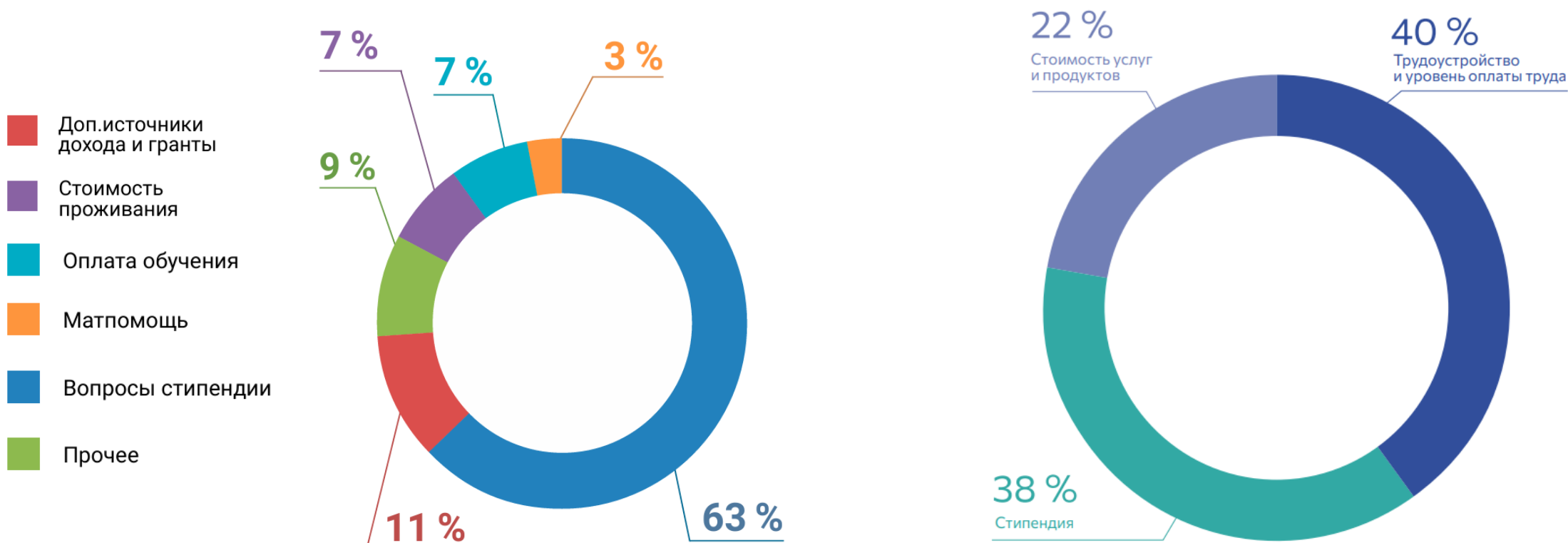
Когда уже туалеты откроют???
Я устал в женский ходить.

Преподаватель душнила страшный,
что делать?

Кто тут учился, Тот в цирке
не смеется!

Финансы студентов

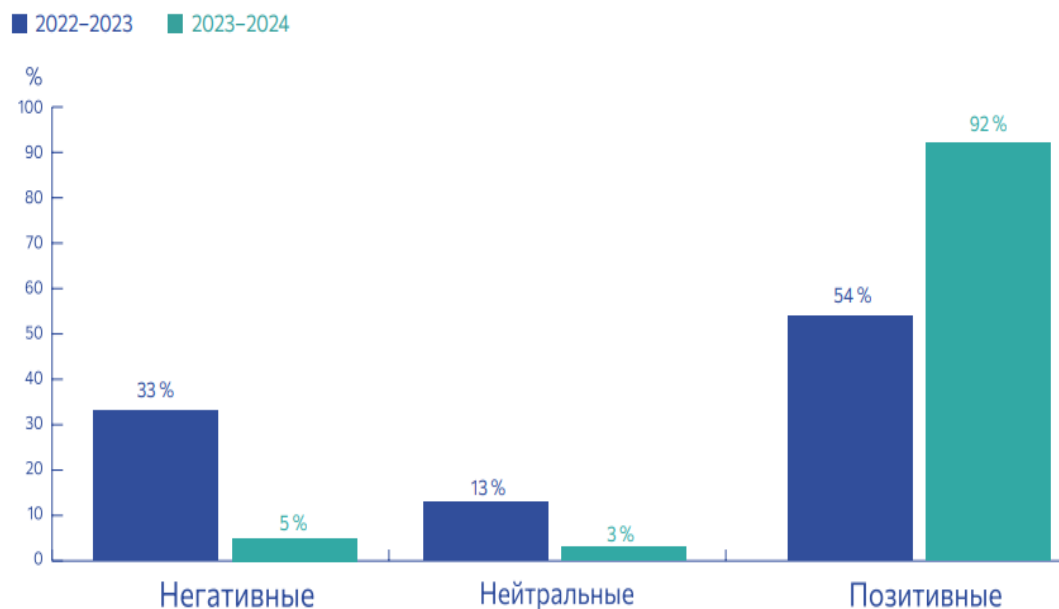
В прошлом учебном году 63% студентов активно обсуждали вопросы, связанные со стипендиями, сейчас же на первый план вышли темы трудоустройства и уровня оплаты труда (40%), стипендия продолжает обсуждаться 38% обучающихся.



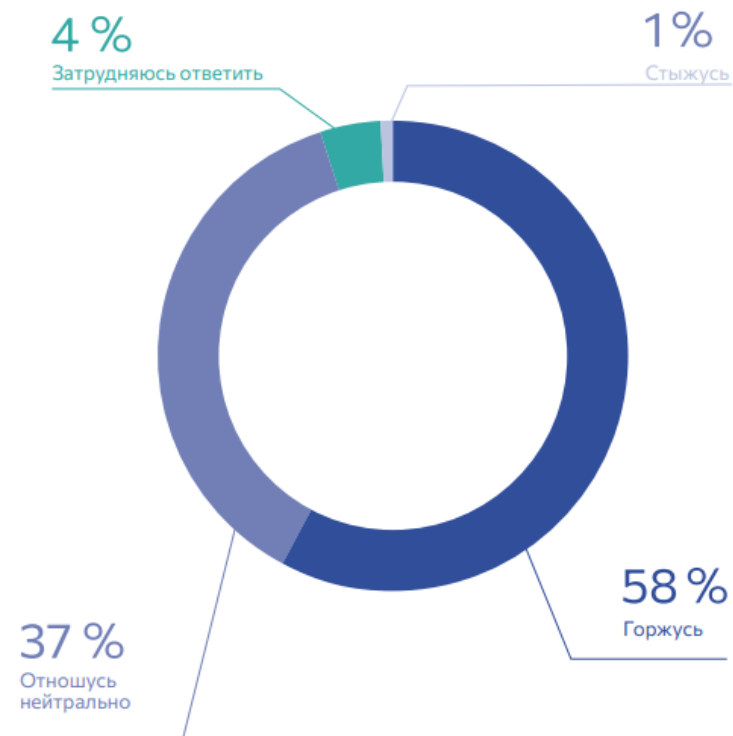
Лояльность студентов к вузу

Снижение доли негативных мнений более чем в 6 раз по отношению к прошлому учебному году, в 82 % случаев - публикации в официальных пабликах, где контент активно модерируется представителями администрации вуза.

По данным ВЦИОМ доля положительных оценок студентами своих вузов выросла на 4 п. п. с 60 % в 2023 году до 64 % в 2024. Гордость за свой вуз, как и в прошлом году, испытывают 58 % студентов.

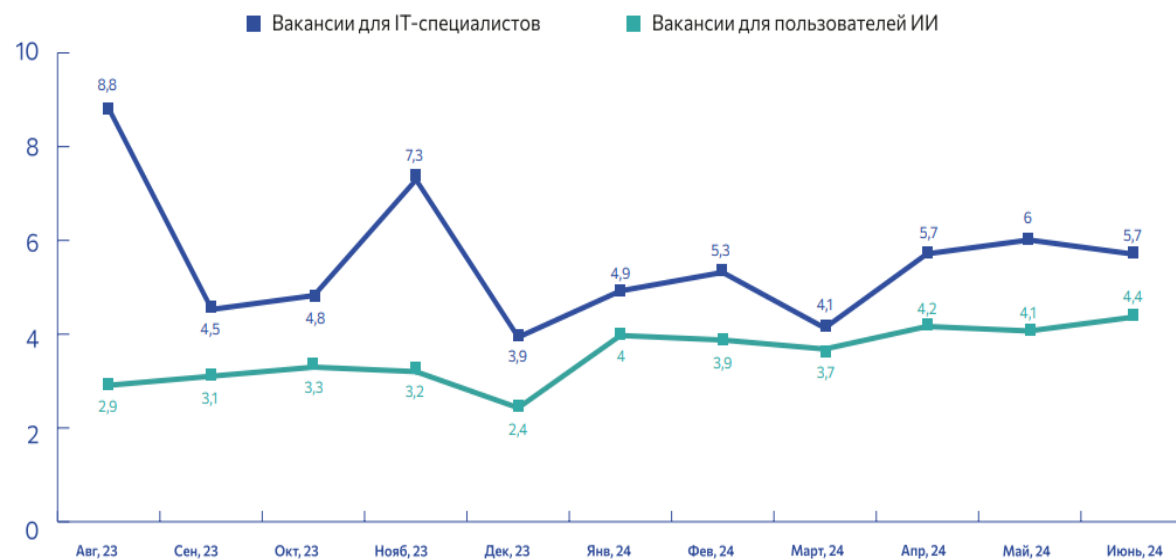


Динамика количества вакансий для IT-специалистов и пользователей ИИ с высшим образованием

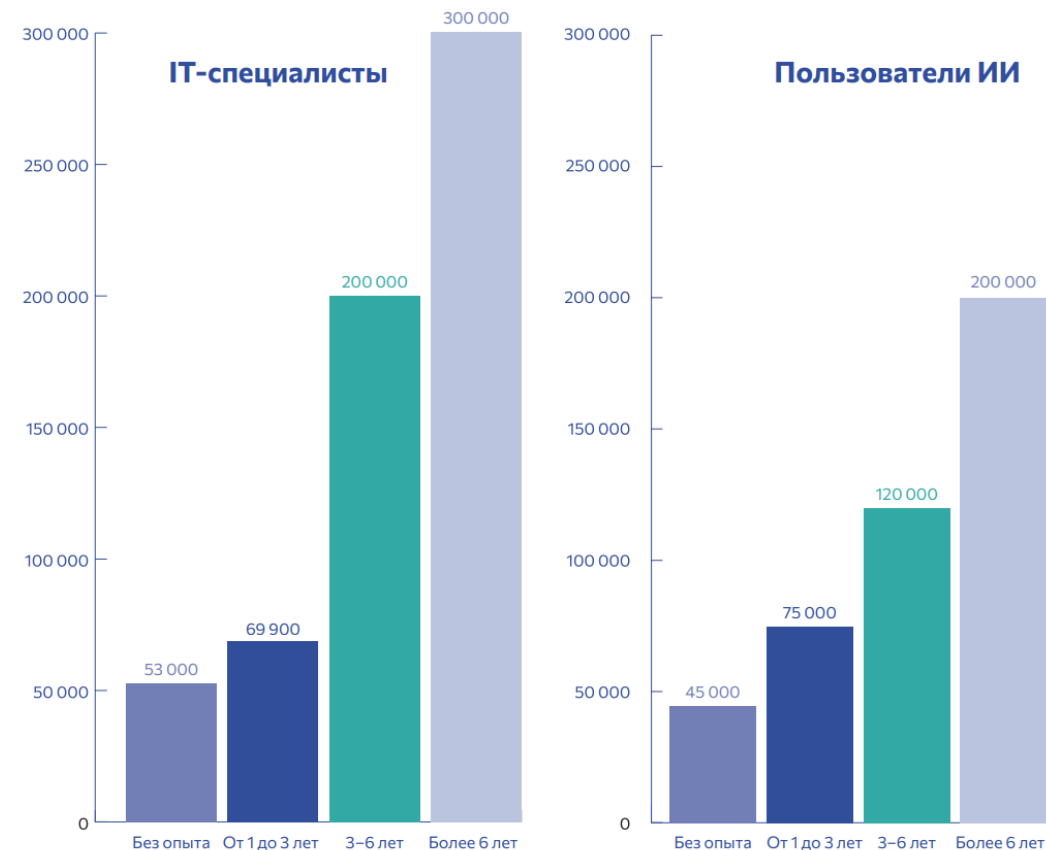


Рынок труда и ИИ

Выгружены 8 671 вакансии (специальности и навыки работы с ИИ, требование высшего образования) из hh.ru, SuperJob, Работа в России, Работа.ру. Данные собраны за период: май 2023г. – июнь 2024г.



Динамика количества вакансий для IT-специалистов и пользователей ИИ с высшим образованием



Распределение медианной заработной платы специалистов в сфере ИИ по опыту работы

Вклад в развитие сообщества

Открытые отчеты

По результатам исследования ежегодно публикуются аналитические отчеты с основными выводами и результатами.

<https://opendata.university/library>




Открытые наборы данных

Dataset containing posts and comments from university publics on the social media VKontakte
<https://data.mendeley.com/datasets/fcyfn32mv6/1>

Dataset containing posts and comments from university publics on the social media VKontakte (2022-2023)
<https://data.mendeley.com/datasets/fvz9mrnjzy/1>

Dataset containing posts and comments from university publics on the social media VKontakte (2023-2024)
<https://data.mendeley.com/datasets/kf3s4xf33j/1>



Как создавать востребованные образовательные программы



РосНавык

мониторинг рынка труда

Веб-сервис для выявления востребованных навыков на российском рынке труда для проектирования образовательных программ и профориентации

4

источника

40

млн вакансий*

89

субъектов РФ



*по данным на июнь 2025 г.

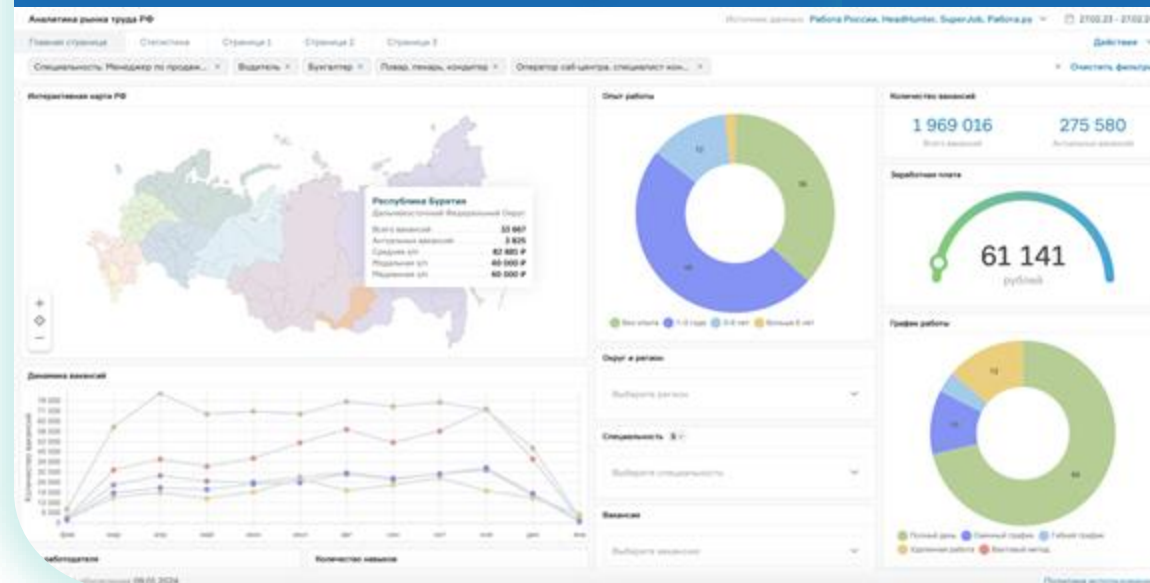
Наши технологии

Автоматический сбор вакансий ежедневно, агрегация в единую базу данных

Определение профессий и требований с применением технологий ИИ

AI помощник для анализа рынка труда и сборки образовательных продуктов

Интерактивный аналитический дашборд для визуализации результатов



Как работает РосНавык

01

Сбор данных

С помощью API и открытых наборов данных собираем вакансии из разных источников

02

Агрегация данных

Объединяем все вакансии в одну базу данных, настраиваем общие переменные, проверяем их качество

03

Подключение моделей

Модели автоматической классификации извлекают из текстов вакансий специальности и навыки

04

Классификация данных

Выделенные специальности и навыки модели объединяют в классы на основе экспертных оценок

05

Анализ данных

Извлекаем из данных полезную информацию о состоянии рынка труда по регионам, отраслям и профессиям

06

Визуализация данных

Представляем результаты в виде карт, графиков и диаграмм в дашборде

Навыки в тексте вакансий

Проблема

У каждого работодателя свой вариант написания навыков и компетенций

Владение компьютером

Знание компьютера

Базовые знания офисных программ

Быть опытным пользователем ПК

Владение ПК уровень бог

Умение пользоваться компьютером

Знание ПК

Знание компьютерных программ



Базовые знания ПК



Извлечение навыков из описания вакансии

Проблема

Работодатели не указывают требуемые навыки в специальных полях на сайтах-агрегаторах

Решение

- Использование LLM
- Использование лингвистических правил и поисковых запросов
- Разработка собственной модели

Использование LLM	Правила и запросы
Долго	Быстро обрабатывают
Высокие требования к вычислительным и денежным ресурсам	Простота реализации
Низкая полнота данных	Низкая полнота данных
Тяжело контролировать	Легко контролировать

ИИ-помощники в разработке образовательных программ

01

AI Аналитик

Проведет комплексный анализ рынка труда по отрасли, покажет ключевые тренды, запросы и предложения работодателей

Инструмент для принятия стратегических решений о запуске и содержании программы.

Аналитический отчет по рынку труда



02

AI Методист

Автоматизирует разработку программы, учитывая ваши цели и аудиторию. Упростит работу методиста.

Сокращение сроков разработки программы на 30-40%.

Проект образовательной программы

03

AI Маркетолог

Поможет выделить вашу программу на основе анализа конкурентов и подготовить эффективные рекламные материалы.

Создание контента для продающего лендинга.

Рекомендации по продвижению программы



РосНавык – инструмент для оценки качества образовательных программ в нацпроекте «Кадры»

ТГУ выдал лицензионный доступ к сервису всем образовательным партнерам. Теперь образовательные программы проходят обязательную проверку на предмет востребованности на рынке труда.

«Для участников нацпроекта мы сделали новый функционал — автоматическое формирование аналитических отчетов по заданным фильтрам. Такой отчет вузы и РОИВы передают федеральному оператору (ТГУ), чтобы подтвердить актуальность своих программ и запрос на специалистов в регионах»

Уже участвуют

185 университетов
45 региональных
органов
исполнительной
власти

Новая модель запуска образовательных программ

Образовательные партнеры – оценка актуальности программ в регионах



РОИВы – оценка кадровой потребности регионов



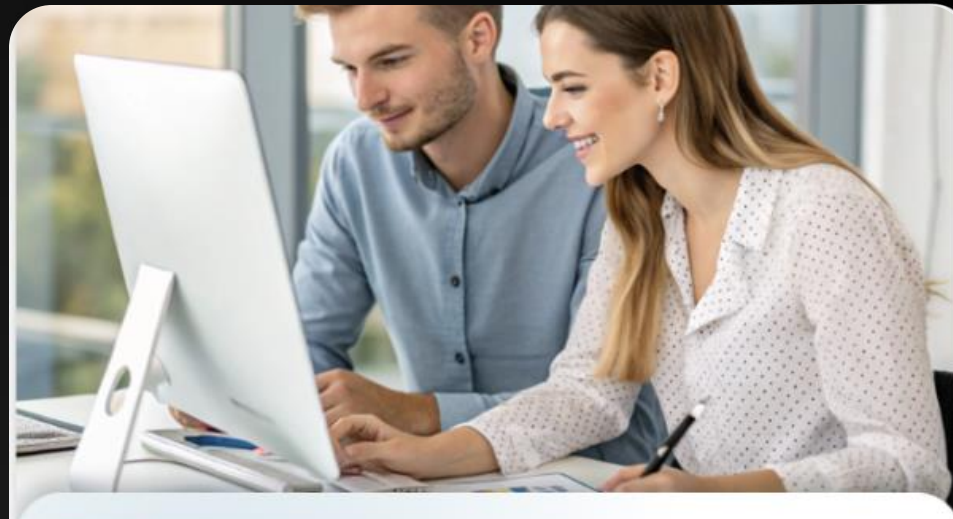
Отчет по рынку труда



Национальный исследовательский
Томский государственный университет

Анализ и верификация предложений по организации обучения от РОИВов
Анализ актуальности предложенных образовательных программ от партнеров

Решение о запуске программы




РосНавык
мониторинг рынка труда

Составлен на основе индивидуального запроса пользователя

**Аналитика рынка труда
на основе запросов
работодателей**

Источники данных:
HeadHunter, Работа.ру, Работа в
России, SuperJob

Период: 09.03.23 — 09.03.25
Дата создания: 14.03.25



Анализ востребованности специалистов в области химии

Методология сбора данных

514 401 вакансии

для специалистов различных направлений, опубликованные компаниями, которые входят в список в подотраслей(компании группы А)

63 414 вакансий

для специалистов с химическим образованием, опубликованные компаниями, которые не входят в список подотраслей (компании группы Б)

Период выгрузки: май 2022 – июнь 2025 гг.

- 1

Экспертное формирование списка подотраслей

 - Формирование списка компаний подотраслей на основе ОКВЭД
- 2

Поиск и валидация компаний в базе вакансий сайтов-агрегаторов HH.ru, Работа России, Работа.ру и SuperJob :

 - ИНН (при наличии)
 - Название (совпадение по косинусному расстоянию)
- 3

Выгрузка профильных вакансий на основе лингвомаркеров, встречающихся в названиях профессий, требующих хим. образования

Подотрасль	Количество компаний
Резинотехнические изделия (РТИ)	395
Изделия из пластмасс	348
Бытовая химия	228
Фармацевтическая химия	211
Лакокрасочные материалы (ЛКМ)	190
Промышленные газы	92
Химические средства защиты растений (ХСЗР)	81
Химволокна	72
Нефтехимические продукты (НХЗ)	71
Неорганические соединения	68
Химия для пищевой промышленности (ХПП)	64
Химия для кормовых продуктов животноводческого комплекса (ХКП)	64
Пластмассы в первичных формах	61
Смазочные материалы	53
Удобрения	36
Катализаторы, сорбенты	29
Шины	17
Лесохимия	13
Синтетический каучук	6
Метанол	3

Структура занятости в хим.отрасли

126 764 вакансии

для специалистов, где требуется химическое образование

63 414 вакансий из них

относятся к компаниям группы Б

Медианные зарплаты **специалистов компаний группы А** превышают для аналогичных вакансий **компаний группы Б** (исключения – химик-технолог, наладчик, научный сотрудник)

Соотношение профильных вакансий по химическому образованию в компаний групп А и Б



Динамика вакансий

Прирост рассчитан относительно прошлого периода, в %

44,94%

прирост в 2024
году вакансий с
хим. образованием
в группе Б

15,97 %

прирост в 2025
году вакансий с
хим. образованием
в группе Б

	Прирост всех вакансий 2024	Прирост вакансий с хим.образованием-24	Прирост всех вакансий 2025	Прирост вакансий с хим.образованием-25
Бытовая химия	37.76	39.31	-10.11	-7.12
Изделия из пластмасс	41.49	44.10	-4.22	-1.32
Катализаторы, сорбенты	68.83	37.07	15.11	20.75
Лесохимия	19.81	67.26	34.99	49.21
ЛКМ	47.86	30.93	-10.94	-12.12
Метанол	30.95	47.06	-7.27	-16.00
Неорганические соединения	83.12	50.66	-2.07	6.99
НХЗ	42.22	33.01	10.35	23.38
Пластмассы в первичных формах	71.67	39.01	32.10	42.90
Промышленные газы	43.08	39.64	111.39	3.58
РТИ	45.99	-23.71	26.68	18.46
Синтетический каучук	-8.68	-64.17	70.15	73.13
Смазочные материалы	33.52	53.21	-5.67	-13.80
Удобрения	31.30	33.84	-0.50	48.72
Фармацевтическая химия	30.97	29.79	3.59	2.41
Химволокна	39.66	37.44	6.19	-4.10
ХКП	65.42	59.85	-9.78	-2.61
ХПП	26.72	33.99	1.53	24.93
ХСЗР	31.61	45.62	-7.39	-1.92
Шины	37.53	26.05	9.26	-0.67

Динамика медианных зарплат

Прирост рассчитан относительно прошлого периода, в %

9%

прирост в 2024 году
медианной зарплаты
в вакансиях с хим.
образованием
в группе Б

13,32 %

прирост в 2025 году
медианной зарплаты
в вакансиях с хим.
образованием
в группе Б

	Прирост зарплат 2024	Прирост зарплат с хим.образованием-24	Прирост зарплат 2025	Прирост зарплат с хим.образованием-25
Бытовая химия -	20.00	20.00	23.33	25.00
Изделия из пластмасс -	20.00	20.00	16.67	16.67
Катализаторы, сорбенты -	31.58	28.57	12.00	15.56
Лесохимия -	12.00	21.30	8.93	3.05
ЛКМ -	20.00	30.00	25.00	15.38
Метанол -	20.93	29.40	11.35	13.79
Неорганические соединения -	8.51	6.67	9.80	12.50
НХЗ -	10.90	11.11	21.73	30.00
Пластмассы в первичных формах -	13.74	-8.33	13.33	13.64
Промышленные газы -	20.00	7.55	16.67	26.04
РТИ -	4.14	23.53	19.79	14.29
Синтетический каучук -	11.14	3.00	36.67	17.48
Смазочные материалы -	27.27	20.00	14.29	26.62
Удобрения -	25.65	20.00	3.30	15.00
Фармацевтическая химия -	10.68	13.54	18.18	21.10
Химволокна -	22.22	30.95	20.00	24.73
ХКП -	20.00	22.22	11.67	14.55
ХПП -	20.00	33.33	16.67	13.17
ХСЗР -	26.67	15.64	18.42	22.13
Шины -	31.41	33.69	13.14	7.00

Выпускники хим.направлений: высшее образование

84 178 выпускника с 2019 по 2024
70 435 из них трудоустроены
в различных сферах и отраслях

Распределение количества выпускников и
трудоустроившихся из них по годам



ТОП-10 выполняемых трудовых функций (среди
определенных) выпускников, которые трудоустроены
в различных сферах и отраслях, в %

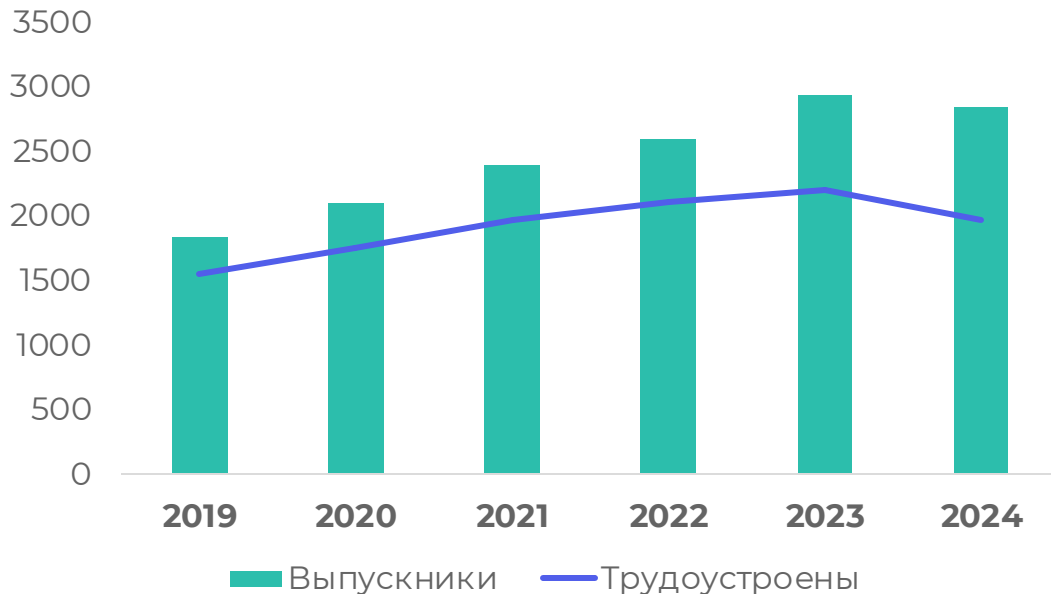


Выпускники хим.направлений: СПО

14 741 выпускника с 2019 по 2024

11 567 из них трудоустроены в
различных сферах и отраслях

Распределение количества выпускников и
трудоустроившихся из них по годам



ТОП-10 выполняемых трудовых функций (среди
определенных) выпускников, которые трудоустроены в
различных сферах и отраслях, в %



Химик–сновидец

В архиве с 8 мая 2025

от 250 000 ₽ за месяц, на руки

Опыт работы: 1–3 года

Частичная занятость

График: 5/2

Рабочие часы: 2 или 3

Формат работы: на месте работодателя



Компания ARSKA — ведущая инжиниринговая компания, создающая химические производства и решения, оптимизирующие деятельность в химической промышленности. Мы всегда в поисках свежих идей, которые упрощают рабочие процессы и влияют на улучшение результатов.

Ссылка на информацию о компании и новости:

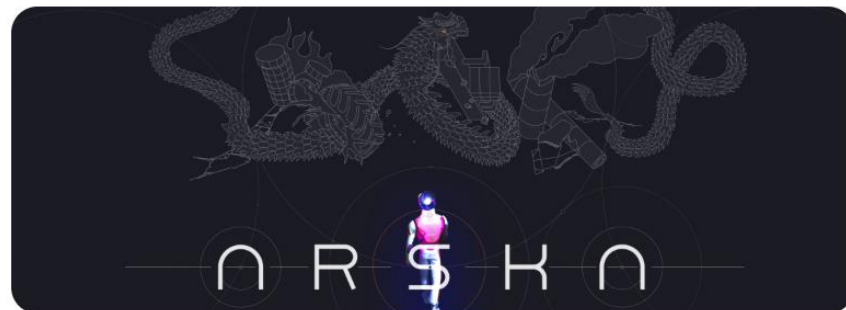
<https://t.me/arskateka>

Мы вдохновились историями ученых, которые увидели свои открытия во сне. Среди них:

- Дмитрий Иванович Менделеев, открывший периодическую таблицу химических элементов;
- Томас Эдисон, запатентовавший более 3000 изобретений, среди которых лампа накаливания;
- Джеймс Уотсон, изобразивший форму и структуру ДНК;
- Нильс Бор, создавший планетарную модель строения атомов, за которую ему вручили Нобелевскую премию,
- и другие.

В связи с этим мы решили запустить экспериментальный проект по поиску гениальных решений во сне.

Для реализации проекта мы ищем талантливых химиков и инженеров, начинающих ученых, готовых поучаствовать в поиске нестандартных подходов и решений в области химии.



Обязанности:

- Спать на территории работодателя по 2-3 часа по будням (краткосрочный сон с 4-6 пробуждениями под наблюдением специалиста по сновидениям)
- После всех пробуждений, пересказывать свой сон специалисту для его интерпретации
- Взаимодействовать со специалистами ARSKA, участвовать в решении инжиниринговых задач
- Фиксировать все свои сны в специальный ежедневник
- Записывать короткие видео с обратной связью по некоторым снам (10 раз в месяц)

Требования:

- Химическое образование (*начинающий* специалист или студент)
- Понимание химических процессов, отличное знание химии
- Умение грамотно доносить свои мысли
- Уверенная речь
- Психическая уравновешенность
- Без хронических болезней

Условия:

- Проект рассчитан минимум на 3 месяца. После этого будет принято решение о его пролонгировании или завершении
- График работы 5/2, на 2-3 часа в вечернее время
- Дружный коллектив
- Бесплатное питание
- Медицинская страховка
- Психолог
- Достойная оплата труда
-

Ключевые навыки

химический инжиниринг

Английский — A1 — Начальный



Возможности консорциума



ПРЕЗИДЕНТСКАЯ
АКАДЕМИЯ



Национальный
исследовательский
Томский
государственный
университет



Университетский
консорциум
исследователей
больших данных

гравитация

2025

МЕЖДУНАРОДНАЯ УНИВЕРСИТЕТСКАЯ ПРЕМИЯ В ОБЛАСТИ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И БОЛЬШИХ ДАННЫХ



ТГ-канал Консорциума

Делитесь **своими исследованиями и разработками в сфере ИИ и больших данных** с широким университетским сообществом. Приглашаем вас публиковать информацию о ваших научных достижениях, проектах и инновациях!

Тут должна быть
вдохновляющая цитата,
но, к сожалению, её нет

