Цифровой блок НИУ ВШЭ
Отдел суперкомпьютерного моделирования

Москва,
08 апреля 2025 г.

Высокопроизводительные вычисления в НИУ ВШЭ

Костенецкий П.С, Козырев В.И., Чулкевич Р.А., Саликова М.Т., Мишенин Р.М.

Характеристики суперкомпьютера *cHARISMa* НИУ ВШЭ
(Computer of HSE for Artificial Intelligence and Supercomputer Modelling)

- **10 место в ТОП 50 СНГ**
- Пиковая производительность для ИИ (FP16) **51.1 Петафлопс**
- Пиковая производительность (FP64): **2.12 Петафлопс**
  (2.12 квадриллиона операций в секунду над числами с двойной точностью)
- LINPACK-производительность: **927.4 Терафлопс**
- **48** вычислительных узлов
  - **6** узлов **с 1 ТБ** ОЗУ**, 8 GPU A100 80 ГБ SXM**
  - **10** узлов с **1,5 ТБ** ОЗУ, 4 GPU V100 32 ГБ
  - **19** узлов с **768 ГБ** ОЗУ, 4 GPU V100 32 ГБ
  - **11** узлов с **384 ГБ** ОЗУ для расчётов на CPU
  - **2** узла с **512 ГБ** ОЗУ, **2 GPU H100 80 ГБ**
- **2** управляющих узла
- **48 GPU NVIDIA A100 80 ГБ**
- **116 GPU NVIDIA V100 32 ГБ**
- **4 GPU NVIDIA H100 80 ГБ**
- **2648** ядер центральных процессоров
- Оперативная память: **53,9 ТБ RAM + 7.9 ТБ GPU Memory**
- Дисковая память: **1,26 ПБ**
  - параллельная СХД на базе Lustre **848 ТБ**
  - локальные диски **128 ТБ**
  - сервер резервного копирования **280 ТБ**
- Коммуникационная сеть: **2 x InfiniBand EDR**
  (**2x100 Гбит/с**, топология **FatTree**)

Суперкомпьютер *cHARISMa* в цифрах

10 МЕСТО
В ТОП-50

2.24 ПЕТАФЛОПС
ПИКОВАЯ ПРОИЗВОДИТЕЛЬНОСТЬ

51.1 AI-ПЕТАФЛОПС
ПИКОВАЯ ПРОИЗВОДИТЕЛЬНОСТЬ ДЛЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

168 ГРАФИЧЕСКИХ ПРОЦЕССОРОВ
NVIDIA A100, V100 И H100

2648 ПРОЦЕССОРНЫХ ЯДЕР
INTEL XEON GOLD / AMD EPYC

54/1256 ТБАЙТ
ОПЕРАТИВНАЯ / ДИСКОВАЯ ПАМЯТЬ

БОЛЕЕ **440**
АКТИВНЫХ ПОЛЬЗОВАТЕЛЕЙ

109 НАУЧНЫХ ПРОЕКТОВ
СЕЙЧАС В РАБОТЕ

БОЛЕЕ **380** НАУЧНЫХ СТАТЕЙ
ОПУБЛИКОВАНО ПОЛЬЗОВАТЕЛЯМИ

64 ПОДРАЗДЕЛЕНИЯ
ИСПОЛЬЗУЮТ СУПЕРКОМПЬЮТЕР

БОЛЕЕ **600** НАУЧНЫХ ПРОЕКТОВ
ВЫПОЛНЕНО НА КЛАСТЕРЕ

БОЛЕЕ **165** НАУЧНЫХ СТАТЕЙ
В ЖУРНАЛАХ ИЗ I КВАРТИЛЯ SCOPUS И КОНФЕРЕНЦИЯХ CORE A*

2 200 000 ЗАДАЧ
ВЫПОЛНЕНО ПОЛЬЗОВАТЕЛЯМИ

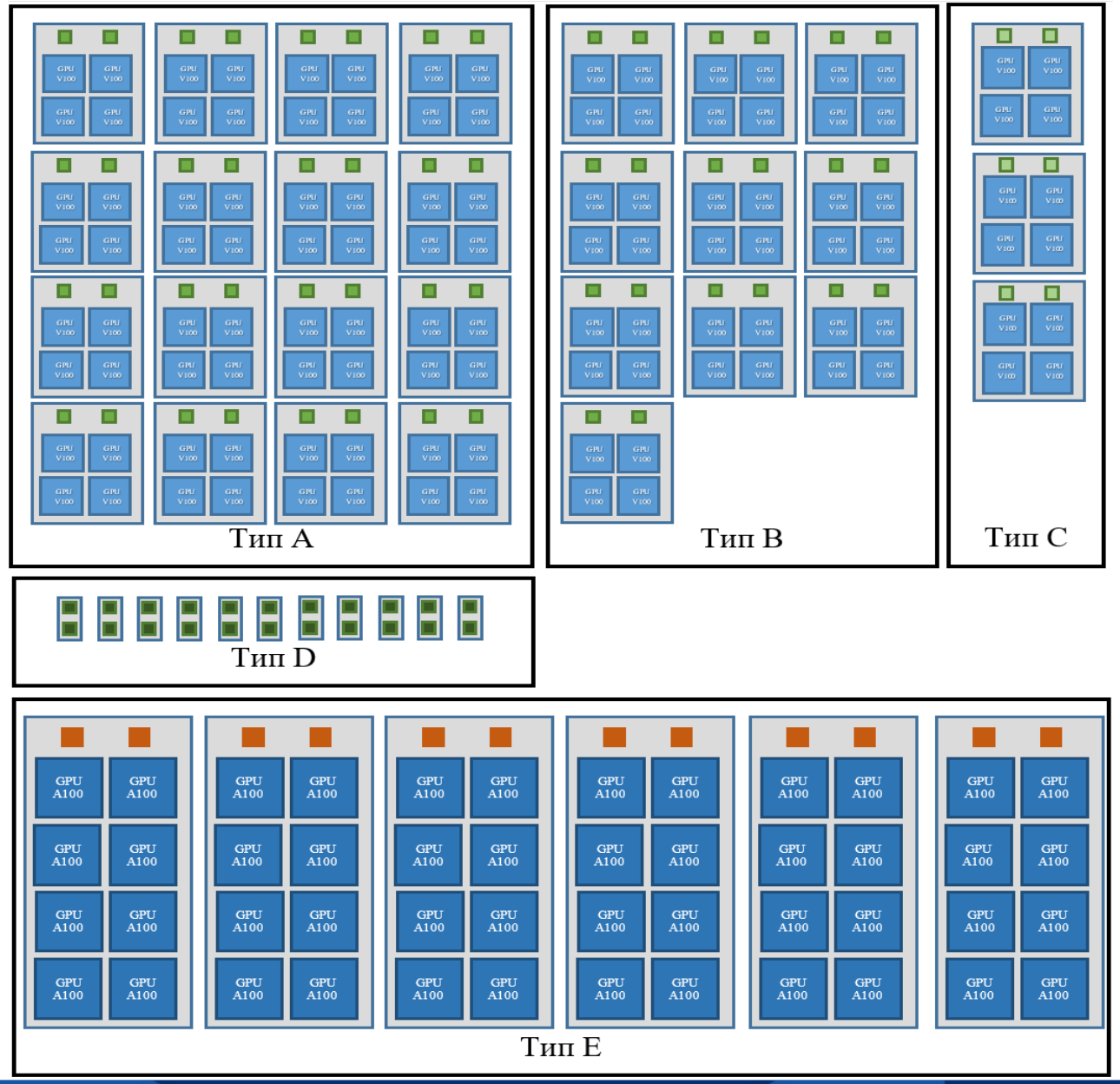
24/7
БЕСПЕРЕБОЙНАЯ РАБОТА СУПЕРКОМПЬЮТЕРА

ПО **2335** ЗАЯВКАМ ОТ ПОЛЬЗОВАТЕЛЕЙ
ВЫПОЛНЕНА ТЕХПОДДЕРЖКА

Мощнейшие компьютеры СНГ

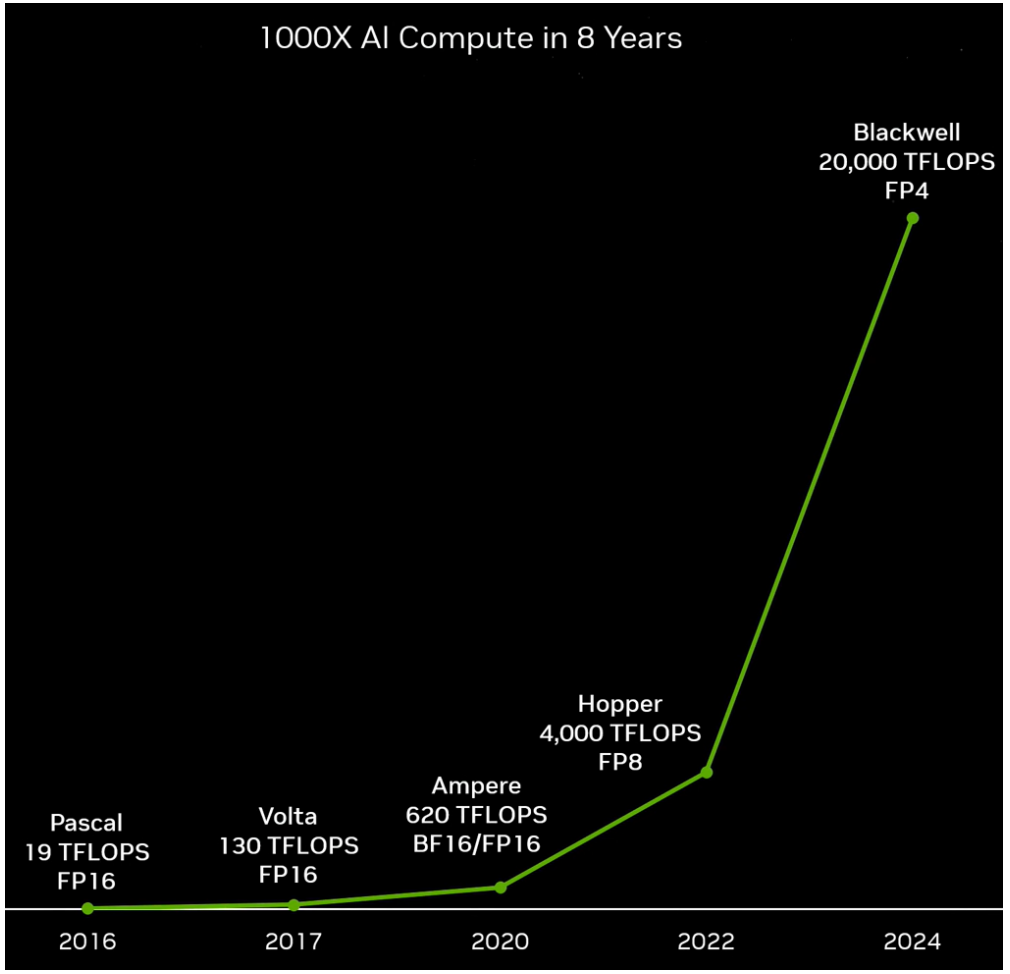
№	Место установки / Название	Linpack (Тфлоп/с)
1	Яндекс / "Червоненкис"	21 530
2	Яндекс / "Галушкиг"	16 020
3	Яндекс / "Ляпунов"	12 810
4	SberCloud / "Кристофари Нео"	11 950
5	SberCloud / "Кристофари"	6 669
6	МГУ / "Ломоносов-2"	2 478
7	МТС / "GROM"	2 258
8	ФГБУ "ГВЦ Росгидромета"	1200
9	Политех, Санкт-Петербург / "Политехник - РСК Торнадо"	971
10	**НИУ Высшая школа экономики / "cHARISMa"**	**927**
11	МСЦ РАН / "МВС-10П"	759
12	НИЦ "Курчатовский Интститут" / "МВС-10П"	756
13	Сколтех, Москва / "ZHORES CDISE Cluster"	496
14	Компьютерные системы, Новосибирск / "PetaNode Cluster"	420
15	Тинькофф Банк / "Колмогоров"	419

Схематическое представление вычислительного поля суперкомпьютера



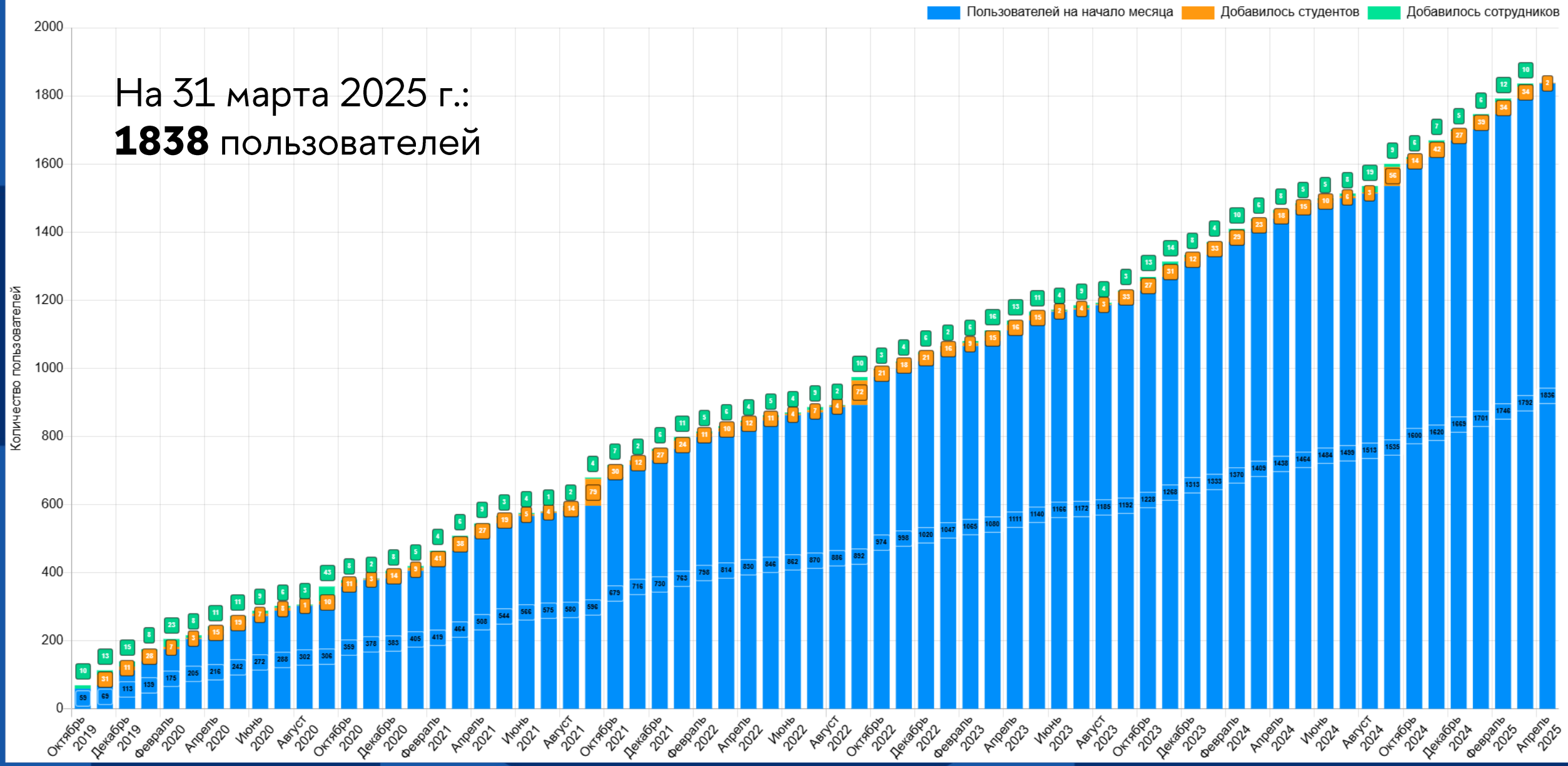
Различные типы GPU для различных задач

Parameter	**H100 80GB PCIe**	**A100 80GB SXM**	**A40 48GB PCIe**	**V100 32GB SXM**	**P40 24GB PCIe**
Architecture	Hopper	Ampere	Ampere	Volta	Pascal
FP64 (TFlops)	25.6	9.7	-	7.8	-
FP64 Tensor Core (TFlops)	51.2	19.5	-	-	-
Real LINPACK performance FP64 (TFlops)	25.71	13.58	-	6.11	-
FP32 (TFlops)	51.2	19.5	10.752	15.7	12
TF32 Tensor Core (TFlops)	378	156	74.8	-	-
BF16 (TFlops)	102.4	39	-	-	-
BF16 Tensor Core (TFlops)	756	312	149.7	-	-
FP16 (TFlops)	102.4	78	37.42	31.4	183.7
FP16 Tensor Core (TFlops)	756	312	149.7	125	-
FP8 Tensor Core (TFlops)	1513	-	-	-	-
INT32 (TOps)	25.6	19.5	5.376	15.7	-
INT8 Tensor Core (TOps)	1513	624	299.3	62	47
INT4 Tensor Core (TOps)	3026	1248	598.7	-	-
GPU Memory (GB)	80	80	48	32	24
GPU Memory Bandwidth (GB/s)	2039	1555	696	900	346
FP32 Cores (Qty.)	14592	6912	-	5020	-
FP64 Cores (Qty.)	7296	3456	-	2560	-
Tensor Cores GPU (Qty.)	456	432	336	640	-
Max Thermal Design Power (TDP)	350W	400W	300W	300W	250W
Multi-Instance GPU (MIG)	7x10 GB, 4x20 GB, 2x40 GB	7x10 GB, 3x20 GB, 2x40 GB	-	-	-
Form Factor	PCIe5	SXM4	PCIe4	SXM2	PCIe3
Interconnect	NVLink Bridge: 600 GB/s, PCIe Gen5: 128 GB/s	HGX-8xGPU NVLink: 600 GB/s, PCIe Gen4: 64 GB/s	PCIe Gen4: 64 GB/s	HGX-4xGPU NVLink: 300 GB/s, PCIe Gen3: 32 GB/s	PCIe Gen3: 32 GB/s



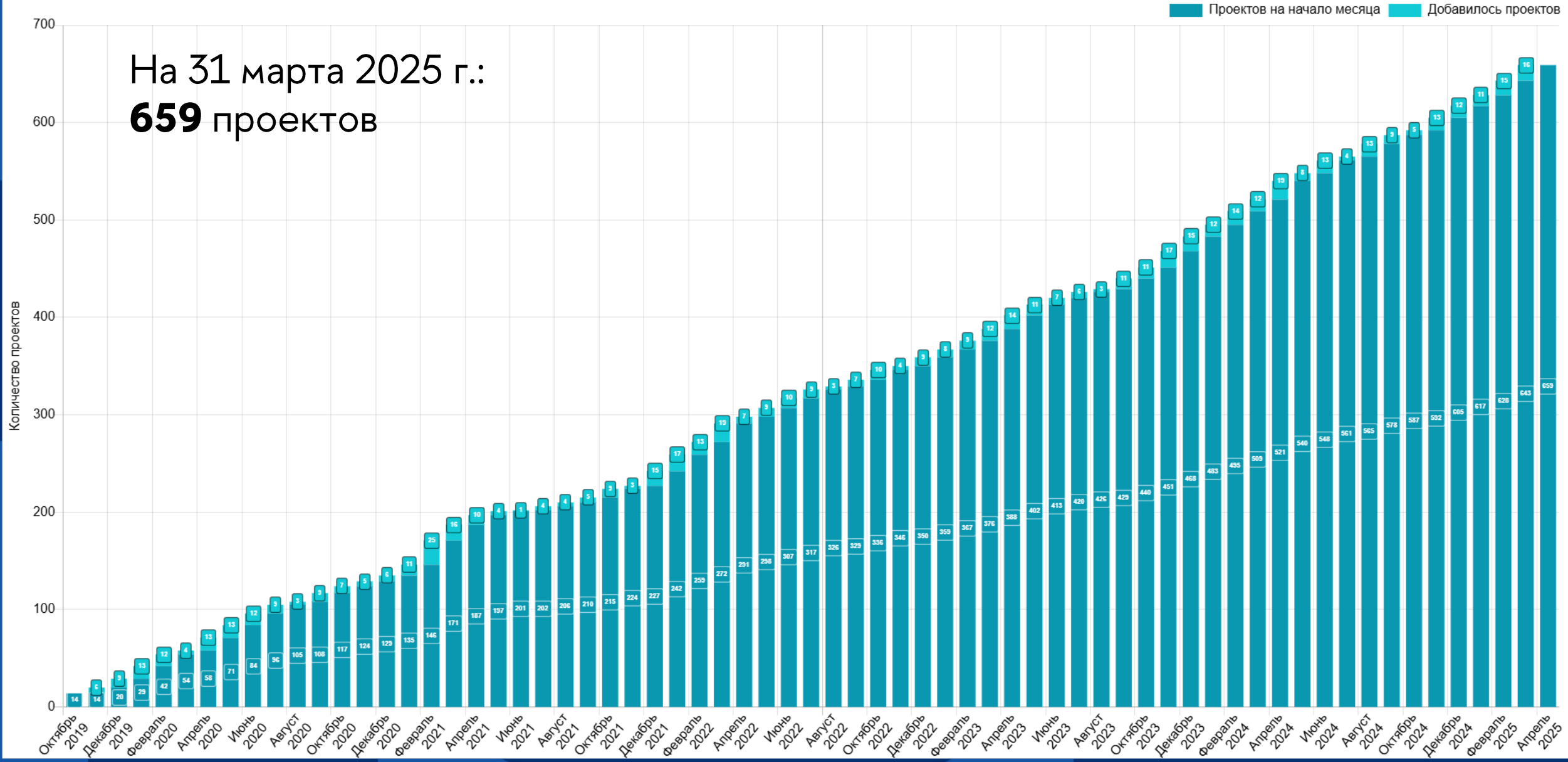
Рост количества пользователей

На 31 марта 2025 г.:
1838 пользователей



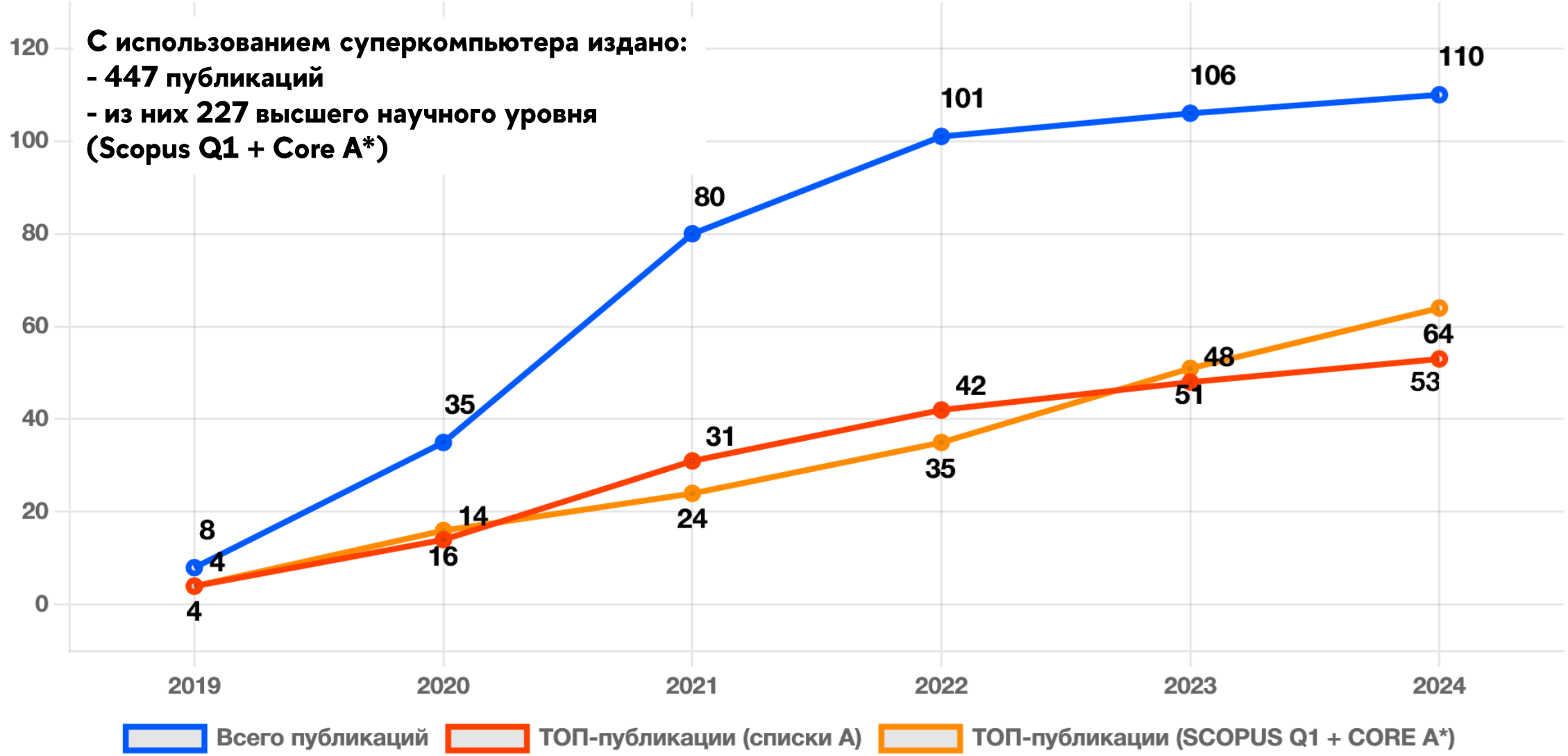
Рост количества проектов

На 31 марта 2025 г.:
659 проектов



Высокорейтинговые публикации пользователей суперкомпьютера НИУ ВШЭ

С использованием суперкомпьютера издано:
- 447 публикаций
- из них 227 высшего научного уровня (Scopus Q1 + Core A*)



Центр Искусственного интеллекта

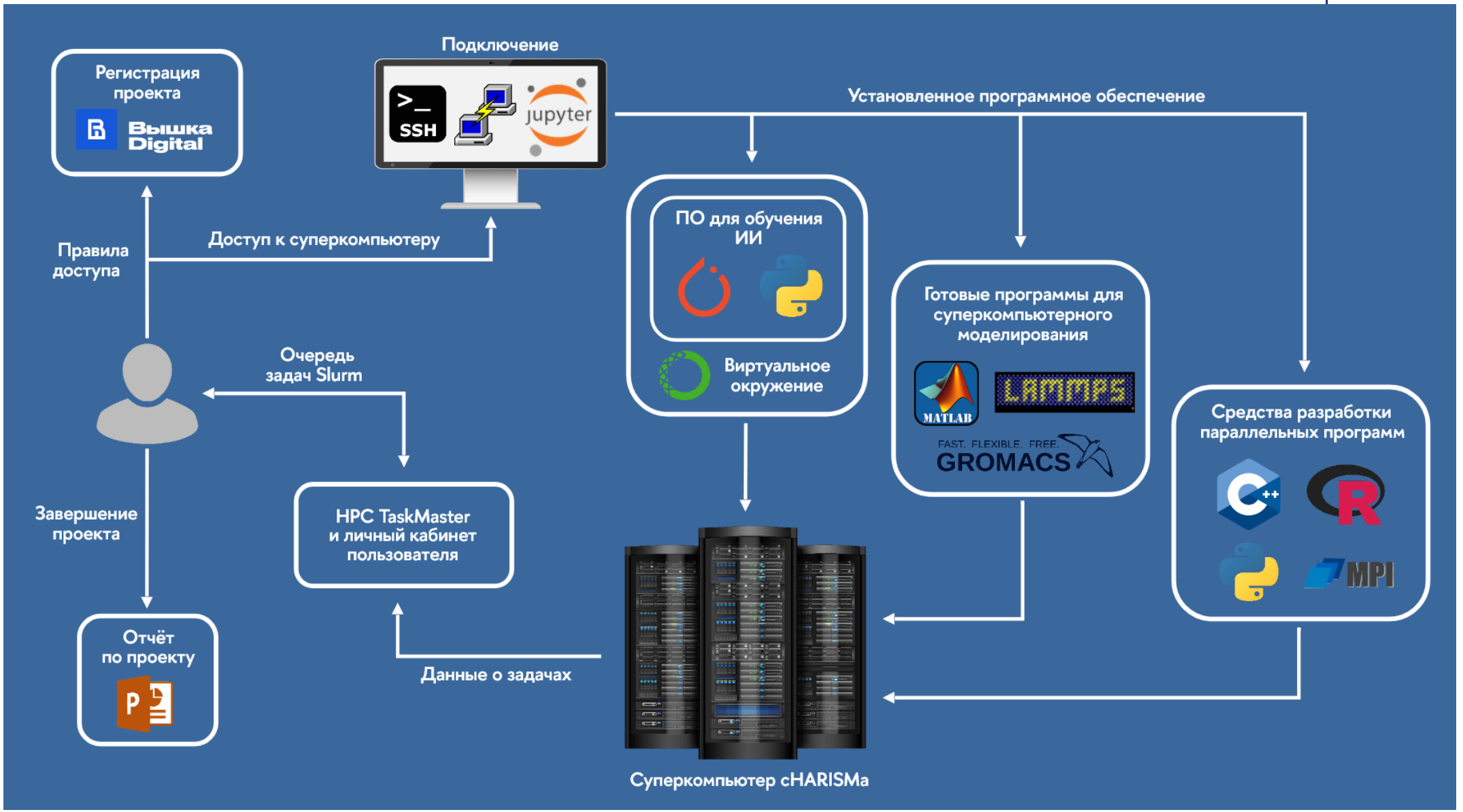
Один из топовых пользователей суперкомпьютера

- Более 40 Междисциплинарных проектов в сфере ИИ.
- Более 30 РИД (датасеты, репозитории, решения с открытым кодом).

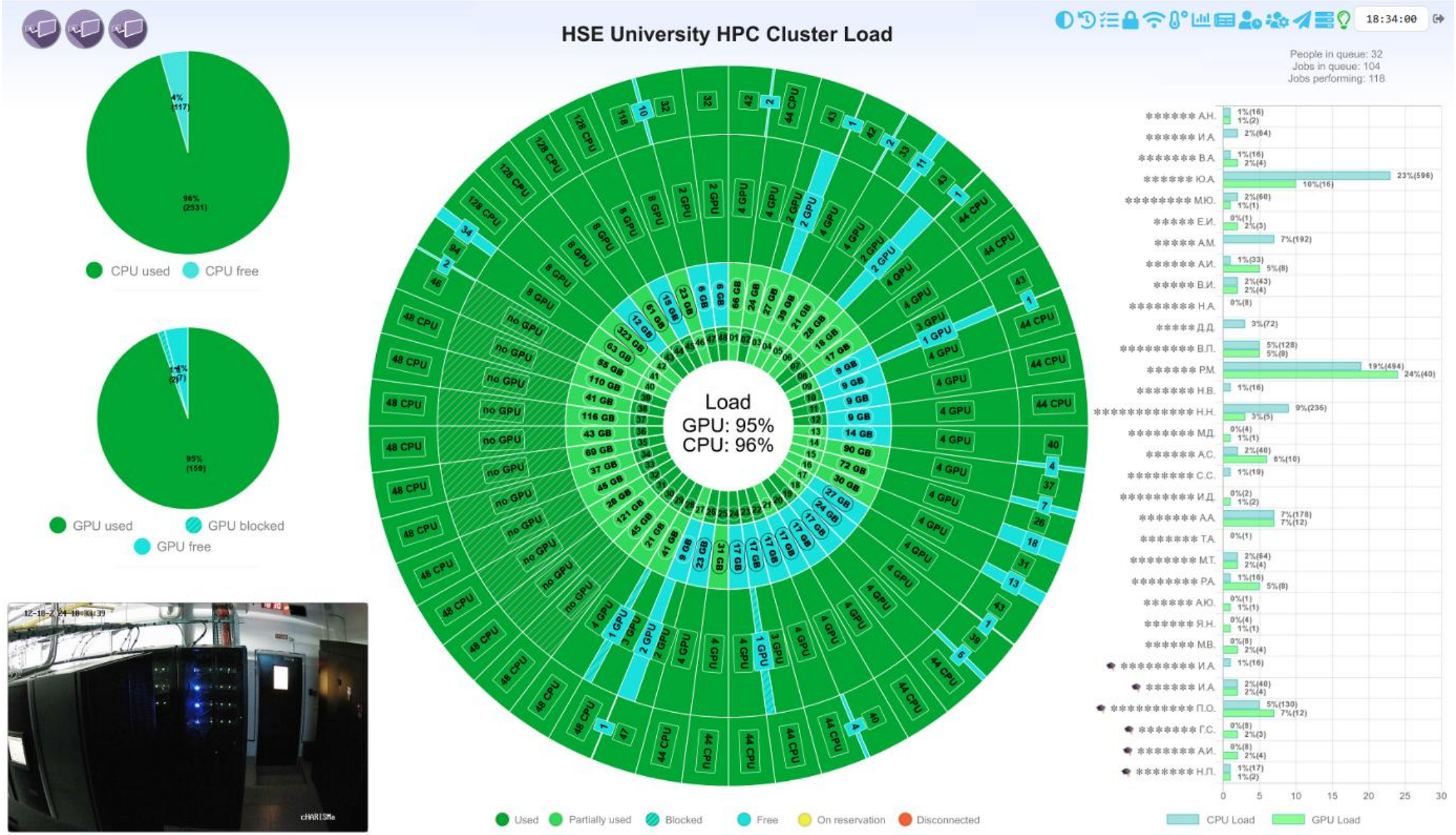
Проекты для индустрии

- Платформа предиктивной маркетинговой аналитики.
- Выявление проблем и потребностей на основе обращений клиентов
- Новый подход для генерации текста (совмещаются диффузионные модели и GPT-модели)
- В рекомендательные системы по подбору продуктов экосистемы внедрены алгоритмы обучения с подкреплением
- Суммаризация текстов для выделения поручений
- Разработана методика для выявления неоднозначного контента на основании тематик.

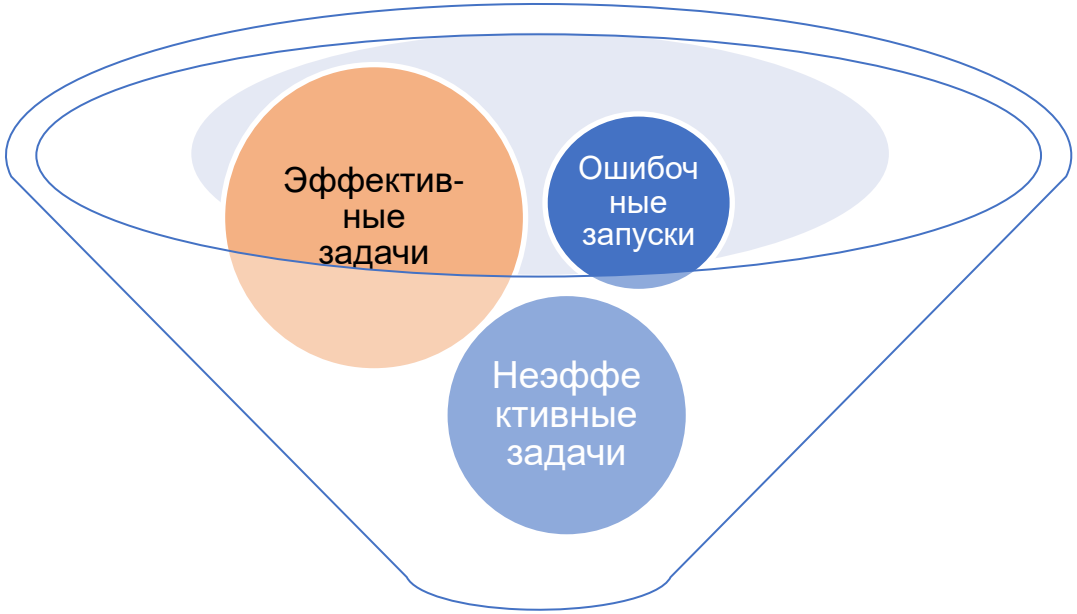
Взаимодействие пользователей с суперкомпьютером



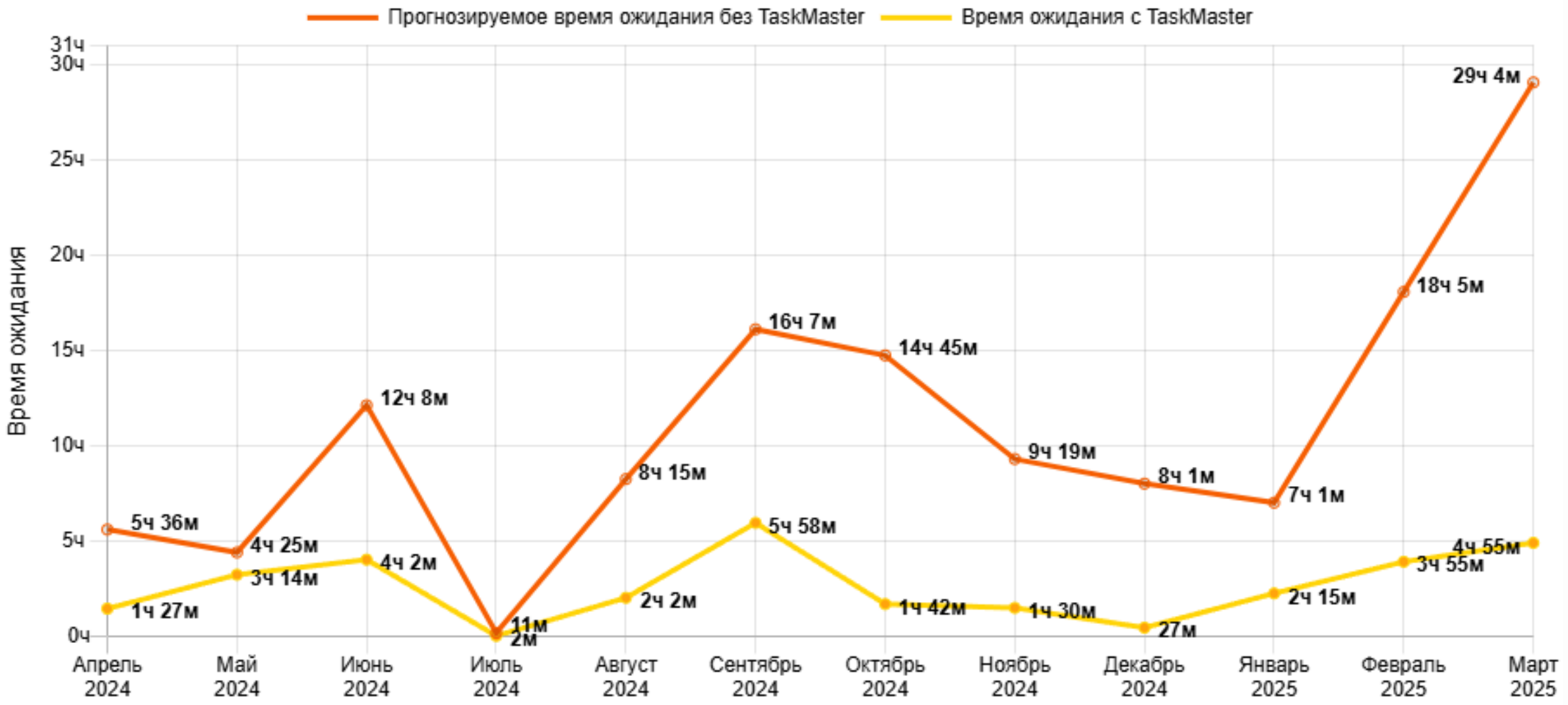
Разработана система мониторинга



Фильтрация неэффективынх задач



- 1838 пользователей выполняют свои расчеты на суперкомпьютере *cHARISMa*
- Пользователи могут ошибочно запускать неэффективные задачи, которые могут замедлить работу остальных пользователей
- Под неэффективными подразумеваются задачи которые используют не все выделенные ресурсы
- Необходимо автоматически определять неэффективных задач, а также собирать статистику пользователей для обнаружения у них проблем с запусками и проведения консультаций
- Разработанная система позволяет экономить до 20% дорогостоящего машинного времени суперкомпьютера

ЕДИНЫЙ РЕЕСТР РОССИЙСКИХ ПРОГРАММ ДЛЯ ЭЛЕКТРОННЫХ ВЫЧИСЛИТЕЛЬНЫХ МАШИН И БАЗ ДАННЫХ

Эффект от HPC TaskMaster