

Материалы к сессии Ученого Совета ОФВЭ

Отдел Вычислительных Систем

Направления деятельности

2025

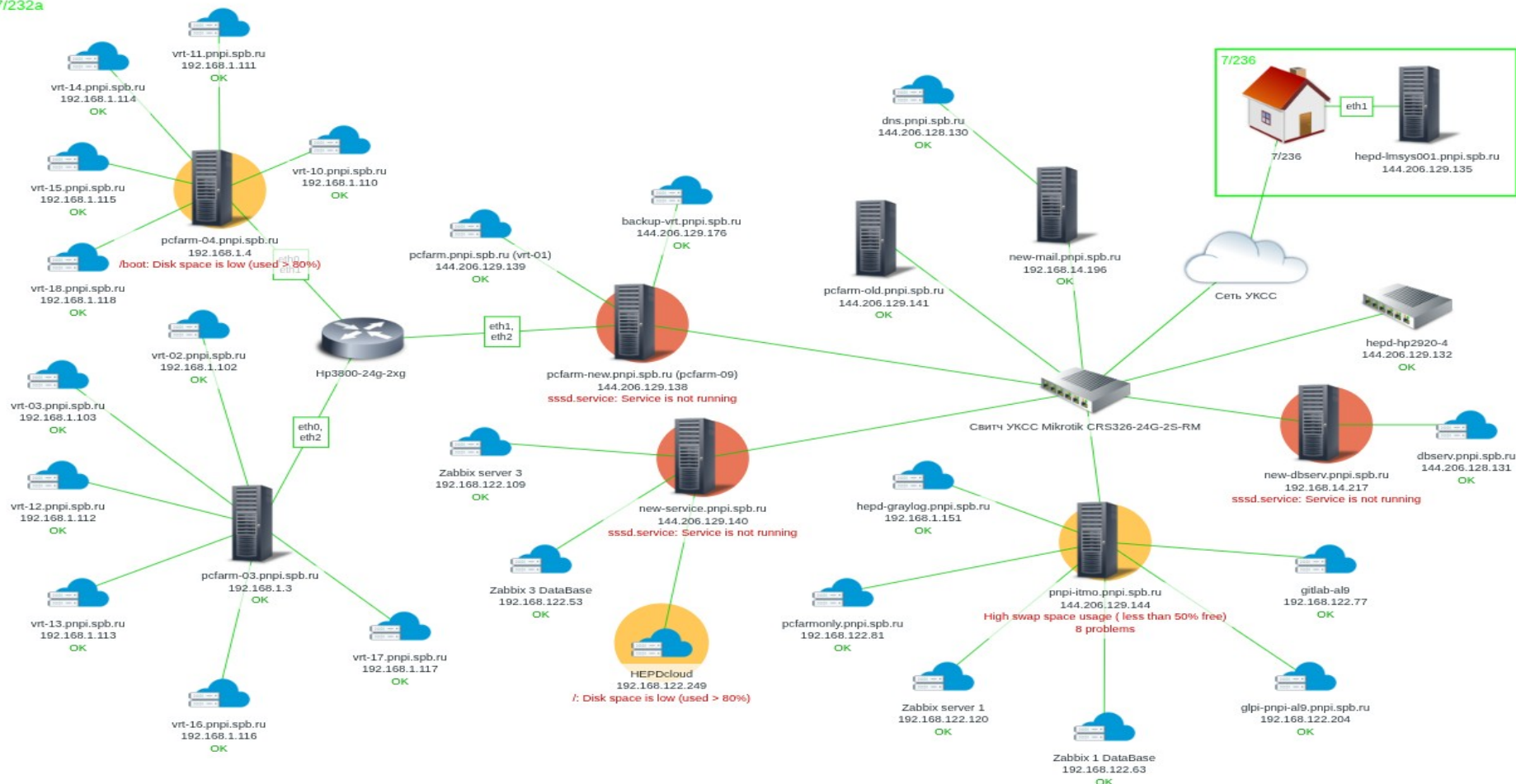
А. Е. Шевель

Основные направления деятельности в 2025

- Разработка эффективных информационно-вычислительных архитектур для соспешествования научным исследованиям ОФВЭ с дальнейшей передачей забот по эксплуатации в централизованные подразделения Института
- **Разработки:**
 - Сценарии применения технологий *машинного обучения* в конкретных условиях.
- Поддержка разработанных ОВС вычислительных систем в форме серверной архитектуры для ОФВЭ:
 - Вэб сервер ОФВЭ.
 - Облачное хранилище данных (<https://lmsys001.pnpi.spb.ru:2180>).
 - Вычислительный микро кластер (pcfarm.pnpi.spb.ru).
 - Сервер JupyterHub (<https://hepd-lmsys001.pnpi.spb.ru:8019/hub>).
 - **Главный и дополнительный DNS серверы** для зоны pnpi.spb.ru, Twiki сервер.
 - Мониторинг серверной сети ОФВЭ (Zabbix) **включая сетевую безопасность**;
 - Полу-автоматическая инвентаризация компьютерного оборудования ОВС (GLPI).
 - Число хостовых и виртуальных серверов на обслуживании составляет 31.
 - **Централизованный цветной принтер (корпус 7).**
- В отделе 3 постоянных сотрудника +2 высококвалифицированных сотрудника (завершивших аспирантуру ИТМО) на ½ ставки.

Схема серверной инфраструктуры ОФВЭ

7/232a



2025-12-18 09:49:04

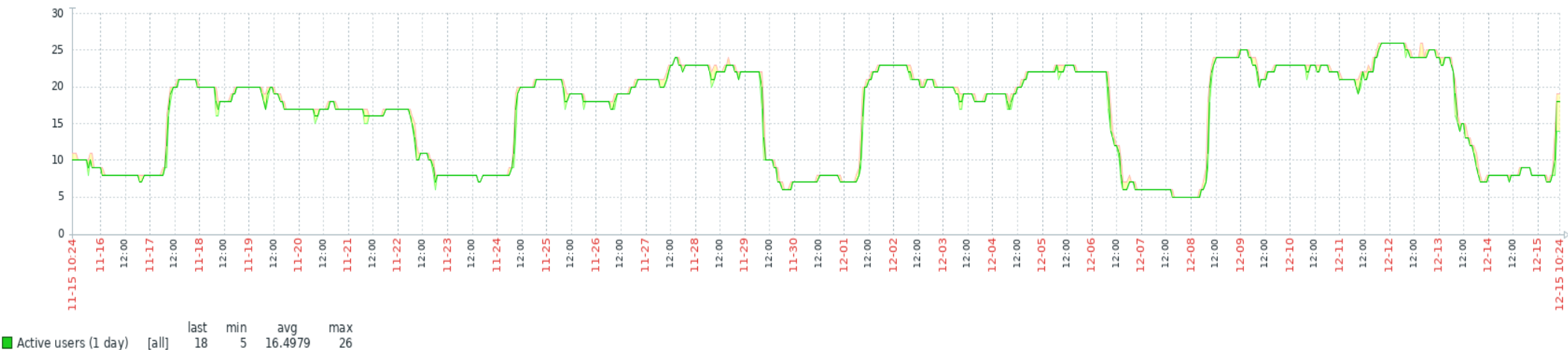
Страны обозреватели вэб сервера ОФВЭ (с 1994) в ноябре 2025

#	Country	Hits
1	<u>V</u> iet Nam	46712
2	Brazil	6986
3	Argentina	1572
4	Russian Federation	1342
5	British Indian Ocean Territory	1230
6	Mexico	606
7	Uruguay	418
8	Paraguay	356
9	Ecuador	352
10	Colombia	258
11	Uzbekistan	236
12	Anguilla	222

13	Kenya	214
14	Nicaragua	192
15	United Arab Emirates	168
16	South Africa	166
17	Pakistan	156
18	Jordan	148
19	Bolivia	128
20	Oman	128
21	St. Helena	128
22	China	122
23	Kyrgyzstan	116
24	Kazakhstan	110
25	Germany	105

Ежедневное число пользователей облачного хранилища ОФВЭ (с 2017)

в ноябре-декабре 2025



Стабильное использование:
в среднем 16 человек каждый день.

Ежедневное число пользователей микро кластера ОФВЭ (с 1998 года)

в ноябре-декабре 2025



Стабильное использование:

в среднем 10 человек каждый день.

Статистика централизованного цветного принтера ОФВЭ

01-24 декабря 2025



В декабре 2025 в среднем около 30 страниц в день

Изменения в компьютерной инфраструктуре ОФВЭ в 2025

- Количество акаунтов локальной облачной системы хранения документов ОФВЭ *превысило 70*. Суммарная ёмкость хранящихся документов более 200 GB.
- На большинстве серверов обновлены дистрибутивы Linux до AlmaLinux 9.x.
 - *На сервере возрастом 15+ лет переход на версию 9.x оказался невозможен из-за старой архитектуры процессора.*
- Используются сценарии машинного обучения. На *pcfarm* доступна платформа *ollama* для изучения разных моделей машинного обучения.

Технологии машинного обучения

- Применение технологий машинного обучения в форме Больших Языковых Моделей является в настоящее время таким же массовым как использование математики — нельзя найти область, где не используются те или иные модели.
- Критерии эффективности использования модели в конкретных условиях (конкретный сценарий использования модели):
 - Имеются средства (методы) верификации ответов моделей в конкретных условиях;
 - Необходимые результаты достигаются с значительно меньшими усилиями, чем без использования моделей.

Сценарий использования машинного обучения

- ОВС разработал конкретный сценарий машинного обучения для разработчиков. Технология - использование Markov Decision Process в контексте Reinforcement Learning. Основная цель – **сократить время на** подготовку статей, описаний систем и/или проектов.
- Цикл уточнения текста описания в соответствии с разработанным сценарием:
 1. Преобразование текста в векторную базу данных (embedding) для использования совместно с моделью.
 2. Подготовка и ввод проверочных вопросов по содержанию текста.
 3. Получение ответов модели и оценка ответов.
 1. Если ответ модели неверен, определить части текста, где необходимы коррекции.
 4. Если внесены коррекции, то пункт 1. Если нет, то пункт 2 или **Завершение цикла уточнения, что определяется командой разработчиков.**
 - **Замечание:** в пунктах 2 и 3 могут возникнуть новые идеи в отношении документа.

Ближайшее будущее

- Разработанный сценарий удобен для открытых публикаций или отчётов, т. к. можно пользоваться инстансами хороших моделей, которые доступны в Интернете: Gemini, ChatGPT, etc. Эти модели имеют достаточно большую базу разнообразных цитат (много миллиардов цитат), поскольку модели тренированы на открытых текстах.
- Однако для локального применения доступных в Интернете моделей в конкретных условиях необходимо эти модели тренировать на локальных конкретных данных, которые не всегда возможно передавать за пределы локальной сети.
- Для любой крупной установки: компьютерной, физической/механической, организационной имеется масса конкретных технических деталей, которые совершенно неизвестны моделям в Интернет.
- **Прогноз:** в ближайшие годы станет стандартом использование локальной системы машинного обучения, которая тренирована на локальных данных для любого крупного технического изделия, особенно уникального. Такая система машинного обучения станет дополнением к обычному техническому описанию крупной технической установки.

Инструменты Машинного Обучения в Интернет и локально

- Для поиска ответов на вопросы, выполнения обзоров, переводов, преобразований текста и др. (список популярных моделей на следующем слайде)
- Инструменты для локального использования
 - Ollama
 - LmStudio
 - Open WebUI
- А также на страницах <https://hepd.pnpi.spb.ru/CSD/>

Популярные Большие Языковые Модели

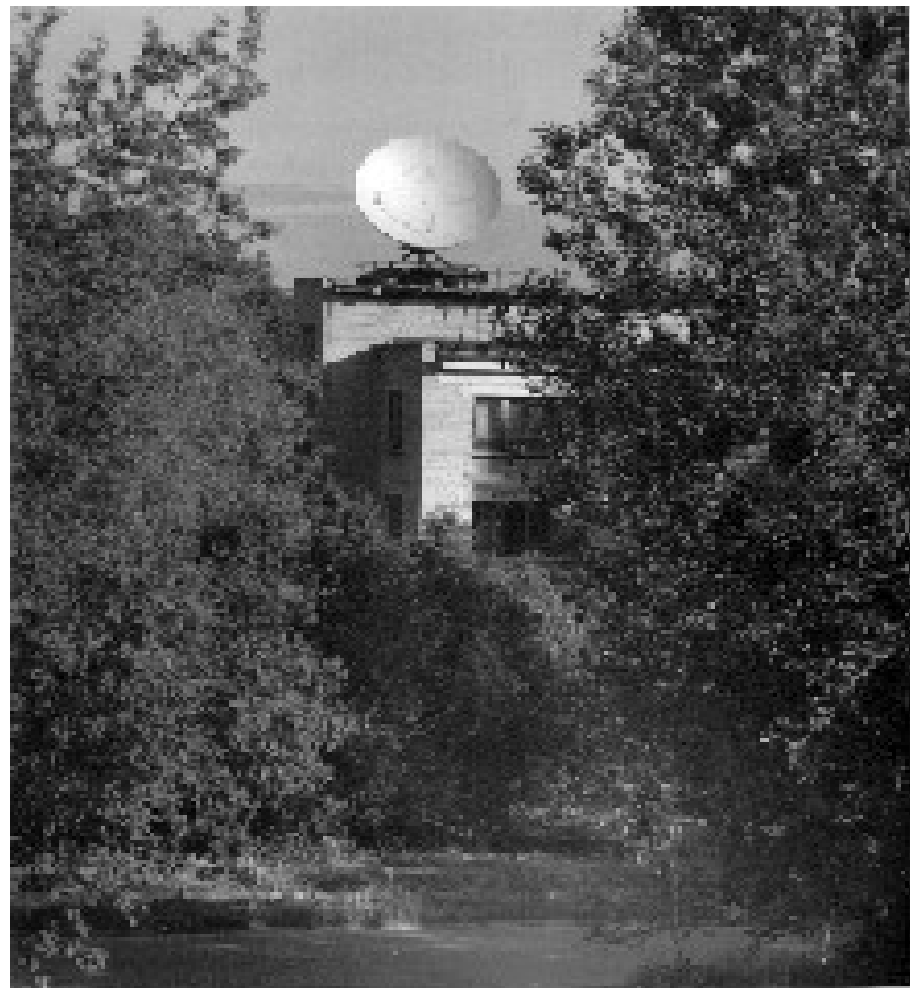
<u>Имя модели</u>	<u>Разработчик</u>	<u>Год выпуска</u>	<u>Примерное число пользователей в месяц</u>
GPT-5	<u>OpenAI</u>	Aug 2025	769 M
<u>Doubao</u>	<u>ByteDance</u>	2025	159 M
<u>Qwen 3</u>	Alibaba	Apr 2025	152 M
Gemini 2.5 Pro	Google	Mar 2025	77 M
DeepSeek-V3	<u>DeepSeek</u>	2025	72 M
Grok 5	<u>xAI</u>	Jul 2025	55 M
Llama 4 Scout	Meta	Apr 2025	50 M
Claude 4.1	<u>Anthropic</u>	Aug 2025	35 M

ОВС: Публичные выступления и подготовленные документы

- **Препринт:** «The machine learning platform for developers of large systems»
<https://arxiv.org/abs/2501.13881>
- **Выступление** на международном совещании HEPIX Spring с сообщением «Machine learning for developers and administrators»
 - <https://indico.cern.ch/event/1477299/contributions/6370765/>
- **Семинар** ОФВЭ «Обзор международного совещания (Hepix-Spring-2025 March 31, 2025 to April 4, 2025) - 2025-05-15»
 - Презентация - https://hepd.pnpi.spb.ru/hepd/events/abstract/2025/Survey_Hepix_2025-05-13_0959.pdf
 - Видео - https://hepd.pnpi.spb.ru/hepd/events/abstract/2025/HEPD_Seminar_Shevel_A.Ye._13.05.2025.mp4
- **Пленарный доклад** «*Large Language Models in High Energy Physics (succinct survey) and directions of future developments*» на конференции «11th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2025)»
 - <https://indico.jinr.ru/event/5170/contributions/31950/>
- **Ежегодный доклад** о развитии ИТ инфраструктуры ОФВЭ на ближайшие 5 лет
https://hepd.pnpi.spb.ru/CSD/CSD_Docs/HepdForecast-2025-08-16.pdf

“Круглые” даты ОВС

- 30 лет (1995)
- А.Е. Шевелю удалось привлечь из гранта INTAS ~\$150K+ на запуск спутникового канала в Интернет. Этим каналом пользовались свыше тысячи сотрудников ПИЯФ в период 1995-2000.



Антенна на крыше корпуса 2а

Семинар ОВС посвящённый запуску спутникового канала октябрь 1995 (большой зал, корпус 7)



Семинар ОВС посвящённый запуску спутникового канала октябрь 1995 (большой зал, корпус 7)



Спасибо за внимание!
Далее справочные слайды

Информация для потребителей

Любые компьютерные ресурсы ФГБУ ПИЯФ предназначены только для санкционированного использования зарегистрированными лицами в целях, описанных в уставе Института.

За всё, что выполняется в рамках зарегистрированного аккаунта, отвечает только владелец аккаунта.

Любое использование компьютерных ресурсов, любые файлы, передачи данных, выполнение команд могут быть скопированы, инспектированы и переданы официальным уполномоченным лицам и/или организациям.

Несанкционированное или неправильное, т.е. вне целей определённых уставом Института, использование компьютерных ресурсов может привести к административным и другим последствиям.

Если вы не согласны с такими условиями использования компьютерных ресурсов -

НЕМЕДЛЕННО ПРЕКРАТИТЕ их ИСПОЛЬЗОВАНИЕ.

Справочные ссылки

- https://hepd.pnpi.spb.ru/CSD/CSD_SupportedProjects.shtml - проекты ОВС
- <http://hepd.pnpi.spb.ru/VIDEO/> - оборудование видео конференций
- <http://hepd.pnpi.spb.ru/CSD/> - инфо об ОВС
- <https://hepd-lmsys001.pnpi.spb.ru:8019-> JupyterHub
- Облачное хранилище <https://lmsys001.pnpi.spb.ru:2180>
- Машинное обучение в ОФВЭ
https://hepd.pnpi.spb.ru/CSD/CSD_MachineLearning.shtml

Пользовательский сегмент серверной сети ОФВЭ

- Число акаунтов в облачной системе хранения ОФВЭ около 70 - <https://lmsys001.pnpi.spb.ru:2180>
- JupyterHub - <https://hepd-lmsys001.pnpi.spb.ru:8019>
- Web сервер ОФВЭ - <https://hepd.pnpi.spb.ru>
- Сервер логирования в микрокластер - [pcfarm.pnpi.spb.](https://pcfarm.pnpi.spb.ru)
- **Замечание:** на всех серверах выполняется регулярное резервное копирование данных домашних директорий, а также данные Web сервера ОФВЭ и конфигурационные файлы всех служебных виртуальных машин.

Свойства мульти пользовательского портала JupyterHub

- различные языки программирования (kernels)
- интерактивный программный код
- визуализация результатов
- редактирование кода в браузере, с подсветкой синтаксиса, автоотступами и автодополнением
- запуск кода в браузере
- отображение результатов вычислений с медиа представлением (схемы, графики)
- работа с языком разметки Markdown и LaTeX

Платформа интерактивной разработки ОФВЭ JupyterHub

- Jupyter — это проект с открытым исходным кодом, платформа, которая помогает максимально просто получить собственную web-среду разработки, не задумываясь о локальных пакетах и развертывании. С помощью Jupyter можно не только осуществлять обработку данных, но и делиться результатами с другими.
- <https://hepd-lmsys001.pnpi.spb.ru:8019>

Свойства Zabbix (Мониторинг)

- Распределённый мониторинг — до нескольких тысяч узлов. Конфигурация младших узлов полностью контролируется старшими узлами, находящимися на более высоком уровне иерархии
- Сценарии на основе мониторинга
- Автоматическое обнаружение
- Централизованный мониторинг журналов
- Веб-интерфейс для администрирования и настройки
- Отчётность и тенденции
- SLA-мониторинг
- Поддержка высокопроизводительных агентов (zabbix-agent) практически для всех платформ
- Комплексная реакция на события
- Поддержка SNMP v1, 2, 3
- Поддержка SNMP-ловушек
- Поддержка IPMI
- Поддержка мониторинга JMX-приложений
- Поддержка выполнения запросов в различные базы данных без необходимости использования сценарной обвязки
- Расширение за счёт выполнения внешних скриптов
- Гибкая система шаблонов и групп
- Возможность создавать карты сетей

Система автоматизированной инвентаризации (GLPI)

- Инвентаризацию компьютеров, периферийного оборудования, сетевых принтеров и связанных компонентов через интерфейс с OCS Inventory или FusionInventory.
- Управление заявками и инцидентами
- Управление лицензиями, договорами (по стандарту ITIL)
- Управление деловой и финансовой информацией (договоры)
- Управление статусом объектов
- Поддержка базы знаний и Часто задаваемых вопросов (FAQ)
- Генераторы отчетов
- Поддерживаются базы данных MySQL/MariaDB
- Поддержка UTF8
- Система оповещения о событиях