

ПСМ¹

2026

ЗАЩИЩЁННЫЙ ЛОКАЛЬНЫЙ AI-КОНТУР

AI Base

Инженерная база

AI Cluster

Вычисления

AI Factory

Прикладной AI



Локальный защищённый AI-контур нужен не всегда

Выбор зависит от от данных, ИБ нагрузки и площадки

Обработка данных допустима вне периметра сети
ИБ и контроль доступа
AI-контур связан с КИИ
Операционные процессы зависят от AI
Постоянная предсказуемая нагрузка
Минимальные задержки ответа
Удалённая / энергодефицитная площадка

Облако
V – подходит
X
X
X
X
! – подходит при условиях
! – подходит при условиях

Локальный контур
X – не рекомендуется
V
V
V
V
V
V

AI-архитектура ПСМ обеспечивает

Защиту

Автономность

Непрерывность

Три уровня готовности AI-контура и бизнес-сценарии

ИТ-контур

Прикладной результат

Конечные сценарии для клиента
Реализуем AI-контур под ключ
или входим проект на нужном вам уровне

AI Factory

3

Управляемая AI-платформа

«железо»

AI Cluster

2

Локальный вычислительный контур

AI Base

1

Площадка под AI-мощности
Сетевое исполнение. Собственная генерация – опционально

ПСМ AI Base

Инженерная база без вычислений

Base

Cluster

Factory

Модульный AI ready ЦОД

без GPU, хранения данных и AI-платформы

Жидкостное охлаждение
liquid-ready / direct-to-chip
для класса нагрузки high-density

База под высокоплотные
ИТ-стойки

Электроснабжение

Мониторинг

Резервирование

Безопасность
и пожаротушение



ПСМ AI Cluster

Вычислительный контур без прикладных сценариев

Base

Cluster

Factory

На базе AI ready ЦОД

без прикладных сценариев

GPU, хранение и ПО подбираются по профилю нагрузки и доступности оборудования

Процессоры CPU / GPU

Структурированная кабельная система

Система хранения данных

Сеть передачи и управления данными

Виртуальная или контейнерная среда запуска AI-сервисов

Базовый мониторинг вычислительного контура

ПСМ. Локальный AI-контур



powerunit.ru

Диапазон AI-нагрузки

Подбор GPU и охлаждения

Класс нагрузки	Профиль нагрузки	Нагрузка на стойку*	Класс GPU*	Инженерная логика
Корпоративный AI-ready	RAG, документы Inference	до 25-32 кВт	RTX PRO / L40S A100-класс	Стандартная инженерия
Промышленный High-density AI	Мультимодал, видео Массовый Inference	32-45 кВт	H100 / H200 GB200-класс**	Охлаждение direct-to-chip (к GPU напрямую на уровне чипа)
Сверхмасштабируемый Extreme-density AI	Rack-scale GPU, высокая параллельная нагрузка, обучение моделей	более 45 кВт, в т.ч. 80-130+ кВт	GB200 NVL-class**	Отдельный проект с CDU (жидкостный контур охлаждения, усиленная энергосистема)

*Проектные ориентиры. Итоговая конфигурация рассчитывается индивидуально

**Проектное исполнение

Базовый модельный контур. Проектные прикладные сценарии

Base

Cluster

Factory

Управление вычислительными ресурсами и нагрузкой:

- Вход и маршрутизация
- Работа с данными (RAG, OCR, база знаний)
- Контроль и качество (журналирование, метрики)

От IT-команды заказчика: данные, бизнес-процессы, экспертиза, требования ИБ

Результаты:

Оптимизация «операционки»

Базы знаний

Планирование

Диспетчеризация

Предиктив

Цифровые двойники

Контроль качества и безопасности

Модели и маршруты обработки

Управляемая архитектура

Единая точка входа AI Gateway ➤ Маршрутизация запроса ➤ Модель ➤ AI-ответ

С чем работаем?	Пример запроса	Маршрут	Примеры моделей
База знаний Поиск ответа на запрос	Как правильно оформить рекламацию?	Text / RAG: Embeddings / Retrieval	LLM + база знаний GigaChat / YandexGPT / Qwen / GLM-класс / российские модели
Быстрые задачи Обработка заданных текстов	Напиши письмо по этим данным: <...>	Быстрый маршрут	Компактные LLM
Договоры и тендеры Сканы, таблицы, структурированные документы	Сравни версии договора и выдели риски	OCR / Parsing → Text / RAG	LLM + правила извлечения данных
Схемы, чертежи, макеты Визуально-текстовые материалы	Сравни презентацию с ТЗ и выяви отклонения	Мультимодальный маршрут	VLM + LLM-анализ
Видеоаналитика	Опиши инцидент на кадре	Отдельный событийный контур	CV + VLM / LLM

Интеграция AI в ИТ-ландшафт и контур ИБ строго по требованиям клиента

Контур ИБ*

Управление ролями и доступами

Приложения / Интерфейсы

AI-платформа
Управление моделями (ML)

Хранение данных

Вычислительные мощности

Инженерная инфраструктура



SIEM/SOC, IAM, DLP



ERP, CRM, Bitrix24, BI



ЕСМ, ЭДО, архивы, базы знаний



MES, SCADA, датчики АСУ ТП



ITSM, Service Desk, почта, мессенджеры, API, webhooks

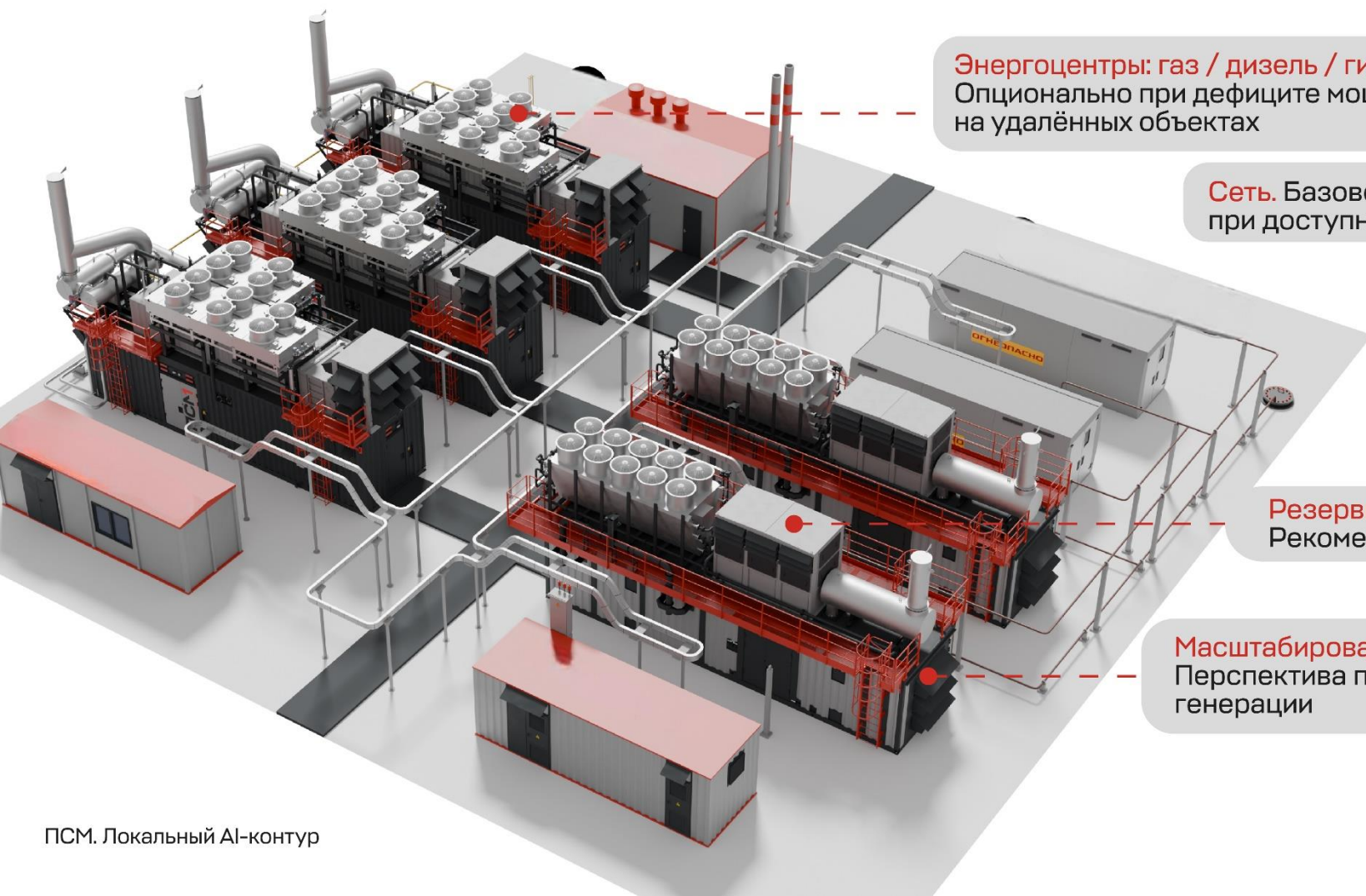
Контур данных заказчика

- данные остаются в согласованном контуре
- доступы наследуются из систем заказчика
- запросы и ответы фиксируются в журнале

*По требованиям 152-ФЗ, 187-ФЗ и ФСТЭК по применимости. Статус соответствия фиксируется в ТЗ и проектной документации

Энергосистема для запуска AI

Сетевое исполнение или собственная генерация



Энергоцентры: газ / дизель / гибрид.
Опционально при дефиците мощности на удалённых объектах

Сеть. Базовое решение при доступном лимите

Резервирование. Рекомендуемая опция

Масштабирование. Перспектива при собственной генерации



Форматы входа в проект

AI Factory в инфраструктуре заказчика

Результат

внедрение модельного контура на инженерной и вычислительной базе клиента

Состав

AI Factory – только модели, маршруты и прикладные сценарии

Пилотный контур

Результат

проверка сценариев и нагрузок

Состав

AI Base / AI Cluster – ограниченный объём вычислений

Промышленный контур

Результат

регулярная эксплуатация AI-сервисов в защищённом контуре

Состав

AI Cluster / AI Factory – расширенное хранение и управление

Автономный контур

Результат

размещение на удалённой / энергодефицитной площадке

Состав

AI Cluster / AI Factory + собственная генерация

Матрица ответственности*

Зависит от модели сотрудничества и роли ПСМ

ПСМ масштабирует участие от инфраструктуры до полного управления AI-контуром

	Поставщик	Интегратор	Оператор
Инженерная инфраструктура	● ПСМ отвечает	● ПСМ отвечает	● Совместно с заказчиком
Энергетика / генерация	● ПСМ отвечает	● ПСМ отвечает	● Совместно с заказчиком
Вычислительное оборудование	● ПСМ отвечает	● ПСМ отвечает	● Совместно с заказчиком
Интеграция платформы управления AI	● ПСМ поддерживает	● ПСМ отвечает	● ПСМ отвечает
Прикладные сценарии	● ПСМ поддерживает	● Совместно с заказчиком	● Совместно с заказчиком
Мониторинг, обновление, сопровождение	● ПСМ поддерживает	● Совместно с заказчиком	● ПСМ отвечает
Данные и бизнес-процессы заказчика	● ПСМ поддерживает	● Совместно с заказчиком	● Совместно с заказчиком

*Фиксируется в договоре и модели эксплуатации

ПСМ. Локальный AI-контур

powerunit.ru

Ответственность после запуска

Управление ответами – через источники, тестовые наборы, журналирование, регламенты.
Человеческий контроль для критичных сценариев.

Корректность AI-ответов управляется, но не может гарантироваться полностью

Заказчик эксплуатирует
самостоятельно

1

ПСМ сопровождает по SLA

2

ПСМ ведёт управляемый
сервис

3

Эксплуатация:

- Мониторинг
- Доступы
- Обновления
- Развитие сценариев
- Регламентные работы

Фиксируем на старте:

- SLA
- ИБ
- Ответственность
- Регламенты
- Масштабирование

Старт работы

Утвердим концепцию AI-контура и рассчитаем проект

[Уровень решения]

ПСМ AI Base
инженерная инфраструктура

ПСМ AI Cluster
инфраструктура + вычисления

ПСМ AI Factory
полный локальный контур

[Энергосистема]

Внешняя сеть

Собственная генерация

[Эксплуатация]

Силами заказчика

При поддержке ПСМ

Управляемый сервис

Следующие шаги

1 Обследование площадки

2 Профиль нагрузок

3 Требования ИБ

4 Конфигурация AI-контура

5 Расчёт проекта:

концепция, CAPEX/OPEX,
план пилота / масштабирования

Референсная база и проектные проработки

[Энергетика]

20 лет

опыта в энергетическом машиностроении

Инженерный лидер

интеграции энергосистем в отрасли ЦОД



[AI-экспертиза]

ПСМ UNLIM

собственная ИТ-компания и продукты автоматизации

Внутренние разработки моделей

RAG, OCR, AI Gateway, маршруты, AI-агенты



[Проектная проработка]

AI Base проект

класс нагрузки high-density, охлаждение direct-to-chip

Концепция

8 модулей, 4 блока по 800 кВт каждый, GPU-нагрузки класса NVIDIA GB200 NVL