



ИСКИНЫ И НОВЫЕ РИСКИ

ГРАЧЕВ Сергей Павлович
Генеральный директор

НАО "ГК "Генезис знаний"

121205, г. Москва, ИЦ Сколково, Большой бульвар, 42с1, офис 1680

mobile: +7 985 233 0437

email: sg@kg.ru

www: kg.ru



ИСКИНЫ И НОВЫЕ РИСКИ

- Автономные Интеллектуальные Системы
- Стандарты, свод, кодекс
 - ПНСТ 776-2022 «Информационные технологии. Интеллект искусственный. Управление рисками»
 - ISO/IEC 23894:2023 «Information technology – Artificial intelligence – Guidance on risk management»
 - The Artificial Intelligence Risk Management Framework
 - Кодекс этики в сфере искусственного интеллекта
- ИскИн для работы с рисками в области ИИ?

АВТОНОМНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

- Опыт людей по управлению бизнесом оцифровывается и накапливается в **базе знаний**
- Рутинная по формированию и согласованию управленческих решений выполняется программными роботами – **агентами**
- Человек становится **экспертом**, настраивающим взаимодействие агентов



Так же как люди, их агенты конкурируют и сотрудничают, достигая поставленные цели

ЛАНДШАФТ РЕШЕНИЙ ИИ

Подход	Оригинальный процесс	Задача
Экспертные системы	Логический вывод	Поддержка принятия решений
Нейронные сети	Сеть нейронов мозга	Распознавание и классификация объектов
Вывод на фреймах	Психологический механизм построения выводов и предположений о новом на основе опыта	Классификация сложных нечётких (новых) объектов
Генетические алгоритмы	Эволюционный отбор набора признаков	Поиск минимума или максимума в нечёткой задаче
Семантические сети и агенты целей	Конкуренция и кооперация в живой природе	Поиск консенсуса разных меняющихся целей при разных меняющихся ограничениях
Эмерджентный, коллективный ИИ	Поведение роя пчёл, колонии муравьев, группы людей	Решение недетерминированных задач
Гибридные методы		

ТЕХНОЛОГИЧЕСКАЯ ГОНКА — БЫСТРЕЕ ВСЕХ РЕШИТЬ САМУЮ СЛОЖНУЮ ЗАДАЧУ В ГОНКЕ ПОБЕДИТ ТОТ, КТО:

2025



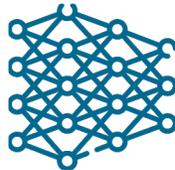
создаст самую большую колонию ИСКИНОВ

2020



соберет самую большую онтологию

2012



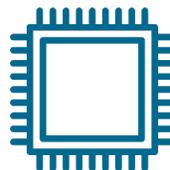
обучит самую большую нейросеть

2000



напишет самую большую программу

1950

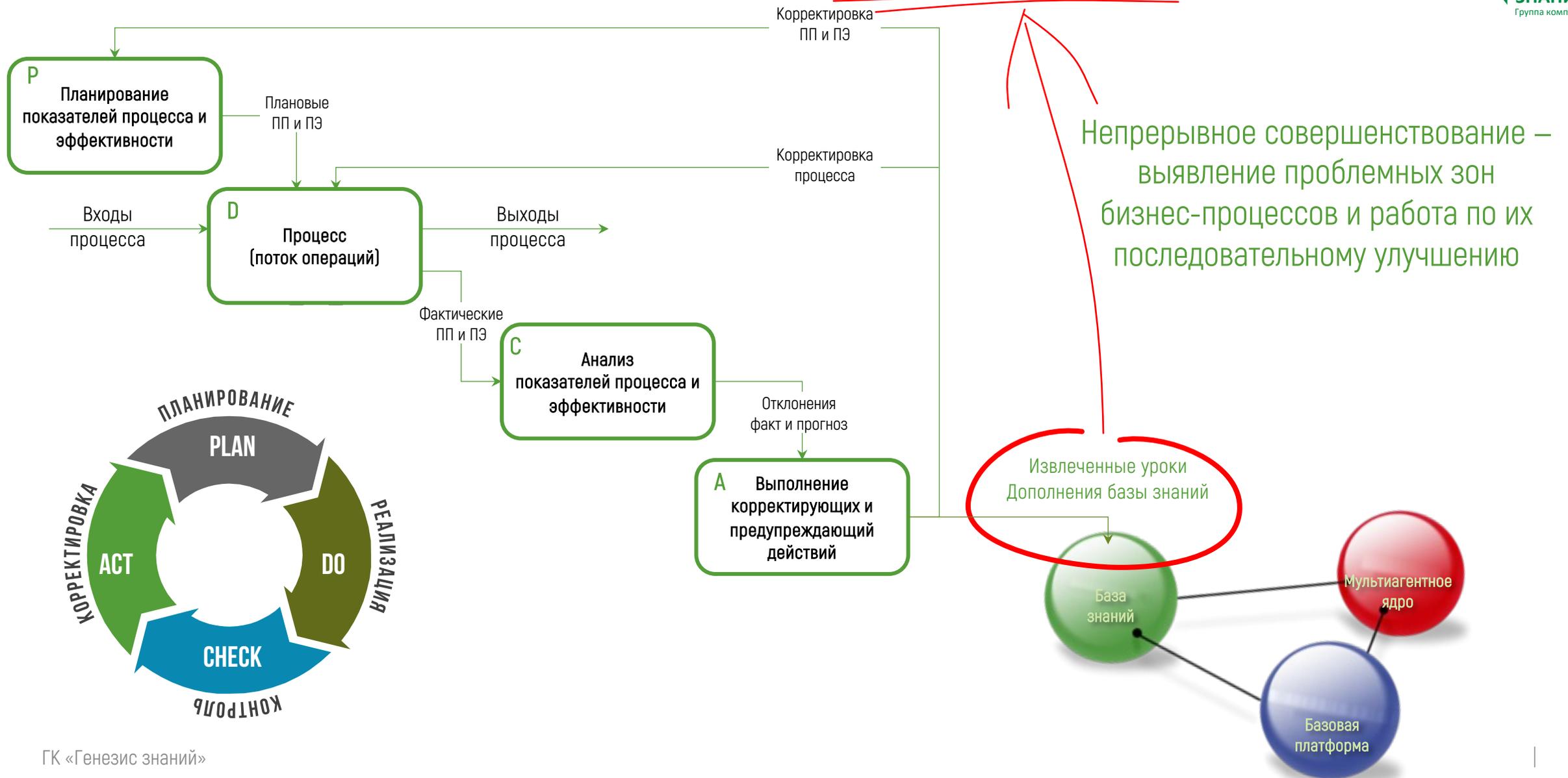


больше всех соберет полупроводниковых
вентилей в одну интегральную микросхему

ТЕХНОЛОГИИ РЕАЛИЗАЦИИ ИННОВАЦИОННОГО ПОДХОДА

Классические методы планирования		Нейронная сеть		База знаний и мультиагентная сеть	
Плюсы	Минусы	Плюсы	Минусы	Плюсы	Минусы
Достижение глобального оптимума по одной целевой функции	Ориентация на интересы центра – не подходит для учета и поиска баланса интересов (консенсуса)	Хорошо подходит для решения узких задач распознавания образов	Требует долгого обучения и полной выборки массивов тестовых данных	Решает сложные задачи управления ресурсами с конфликтами	Высокая сложность и трудоемкость разработки начального решения
Один хорошо изученный метод (алгоритм) для решения любых задач планирования и оптимизации ресурсов	Высокая вычислительная сложность комбинаторного перебора вариантов	Дает устойчивое решение при наличии ошибок в данных, помех и шумов	Зависит от ситуации во внешней среде и учителя: при изменении надо начинать заново	Возможность развития и учета индивидуальных особенностей заказов и ресурсов	При изменении ситуации во внешнем мире требует коррекции и базы знаний
Возможность купить готовый программный продукт на рынке	Не возможность работы по событиям в адаптивном режиме	Возможность купить готовый программный продукт на рынке	Не работает в адаптивном режиме по событиям	Работает в адаптивном режиме - быстрого ответа на событие	Усложнение диалога с пользователем (проактивность, может не согласиться и т.д.)
Легко вкладываются в существующие бизнес-процессы	Большая трудность настройки на решение практических задач	Позволяет использовать накопленные исторические данные предприятия	Большая трудность настройки на решение практических задач	Учет семантики предметной области в базе знаний	Трудности инновации и первопроходцев, смена бизнес-процессов

АВТОМАТИЗАЦИЯ УПРАВЛЕНИЯ НА ОСНОВЕ АИС ДЕЛАЕТ НЕПРЕРЫВНОЕ СОВЕРШЕНСТВОВАНИЕ БИЗНЕСА НЕИЗБЕЖНЫМ



ЭФФЕКТ ОТ ВНЕДРЕНИЯ АВТОНОМНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ДЛЯ УПРАВЛЕНИЯ РЕСУРСАМИ ПРЕДПРИЯТИЯ В РЕАЛЬНОМ ВРЕМЕНИ



Руководитель становится экспертом, настраивающим бизнес-процессы

- Фокус руководителя смещается от рутины процесса управления к настройке процесса



Руководители применяют сценарное моделирование для прогнозирования результатов реализации планов

- Неограниченный сценарный анализ «что-если?»
- Ранжирование достаточного количества вариантов управленческих решений
- Переход от выбора «лучшего из имеющихся» к формированию оптимального варианта решения



Повышается устойчивость и безопасность системы управления

- Снижается вовлеченность людей в рутину процесса управления, операционную деятельность
- Повышается автономность отдельных узлов/процессов



Непрерывное совершенствование бизнеса становится неизбежным

- Дополнение БЗ происходит на каждом цикле управления бизнес-процессом
- Использование «извлеченных уроков» становится возможным уже на следующем цикле управления

Одна расчетная модель бизнес-процесса



- Одна целевая функция, набор ограничений
- Push/pull планирование
- Единая модель всего процесса



Набор онтологических моделей с настраиваемыми функциональными связями



- Неограниченное количество целевых функций, ограничений и функциональных зависимостей
- Интеграция моделей участников процесса путем переговоров

ИСКИНЫ ПОМОГАЮТ ЧЕЛОВЕКУ ДОСТИЧЬ ТОГО, О ЧЕМ ОН МЕЧТАЛ

ПРИЯТИЕ И НЕПРИЯТИЕ АИС ЛЮДЬМИ

- «Я не смогу объяснить руководству чем отличаются два варианта плана, сформированных АИС»
- «А кто сядет?»
- «Мои знания – мои, не отчужу!»
- Сопротивление изменениям
- ...

ПНСТ 776-2022

«ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. ИНТЕЛЛЕКТ ИСКУССТВЕННЫЙ. УПРАВЛЕНИЕ РИСКАМИ»

ISO/IEC 23894:2023

«INFORMATION TECHNOLOGY — ARTIFICIAL INTELLIGENCE — GUIDANCE ON RISK MANAGEMENT»

AI RMF

THE ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK

КОДЕКС ЭТИКИ

КОДЕКС ЭТИКИ В СФЕРЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

МЕНЕДЖМЕНТ РИСКОВ ИИ

ОБЕСПЕЧЕНИЕ ХАРАКТЕРИСТИК ИИ

- компетентность систем ИИ;
- доступность и качество тестирования данных;
- воздействие на окружающую среду;
- справедливость;
- пригодность к ремонту;
- конфиденциальность;
- прочность;
- надежность;
- безопасность;
- прозрачность и объяснимость.

ОСНОВНЫЕ ФАКТОРЫ РИСКА

- недостаточная прозрачность и объяснимость;
- уровень автоматизации;
- факторы риска, связанные с машинным обучением;
- проблемы системной аппаратуры;
- проблемы жизненного цикла системы;
- технологическое развитие.

ОСНОВНЫЕ ОТЛИЧИЯ РИСКОВ ИИ ОТ ТРАДИЦИОННЫХ

- непредвиденные изменения во время обучения;
- высокая степень сложности прогнозирования побочных эффектов.

Risk assessment

- This entails the identification and evaluation of potential AI system risks, such as bias, errors, and security vulnerabilities.

Risk treatment

- Once hazards have been identified, organisations must determine the most effective course of action to manage them. This could involve mitigating the hazards, transferring them to another party, or accepting them.

Risk communication

- Organizations must inform relevant stakeholders, such as consumers, customers, and regulators, of the risks associated with AI systems.

Risk monitoring and review

- To ensure that the risk management framework remains effective, organisations must monitor the performance of AI systems and conduct regular reviews of the risk management framework.

СТАДИИ ФУНКЦИОНИРОВАНИЯ И РАБОТЫ С РИСКАМИ ИИ



СТАДИИ ФУНКЦИОНИРОВАНИЯ ИИ

1. Проектирование
2. Разработка
3. Внедрение
4. Оценка
5. Использование

СТАДИИ РАБОТЫ С РИСКАМИ ИИ

- Управление
- Планирование
- Измерение
- Менеджмент

МЕНЕДЖМЕНТ РИСКОВ ИИ



УПРАВЛЕНИЕ

- выработка политики управления рисками;
- прозрачное описание процессов прогнозирования, выявления и управления рисками;
- оценка потенциального воздействия;
- определение ответственных лиц;
- ...

ПЛАНИРОВАНИЕ

- установление контекста определения рисков, связанных с системой ИИ;
- категоризация систем ИИ;
- оценка возможностей ИИ, его целевого использования, ожидаемых выгод и затрат;
- сопоставление рисков и преимуществ для всех компонентов системы ИИ;
- характеристика воздействия на отдельных лиц, группы, сообщества, организации и общество в целом.

ИЗМЕРЕНИЕ

- приоритизация рисков ИИ;
- выработка стратегий максимизации преимуществ ИИ и минимизации негативных последствий;
- контроль рисков и преимуществ ИИ от сторонних организаций;
- обработка рисков.

КОДЕКС ЭТИКИ В СФЕРЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА



1. ГЛАВНЫЙ ПРИОРИТЕТ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ В ЗАЩИТЕ ИНТЕРЕСОВ И ПРАВ ЛЮДЕЙ И ОТДЕЛЬНОГО ЧЕЛОВЕКА

- Человеко-ориентированный и гуманистический подход
- Уважение автономии и свободы воли человека
- Соответствие закону
- Недискриминация
- **Оценка рисков и гуманитарного воздействия**

2. НЕОБХОДИМО ОСОЗНАВАТЬ ОТВЕТСТВЕННОСТЬ ПРИ СОЗДАНИИ И ИСПОЛЬЗОВАНИИ ИИ

- **Риск-ориентированный подход.**
- Ответственное отношение
- Предосторожность
- Непричинение вреда
- Идентификация ИИ в общении с человеком
- Безопасность работы с данными
- Информационная безопасность
- Добровольная сертификация и соответствие положениям Кодекса
- **Контроль рекурсивного самосовершенствования СИИ**

3. ОТВЕТСТВЕННОСТЬ ЗА ПОСЛЕДСТВИЯ ПРИМЕНЕНИЯ СИИ ВСЕГДА НЕСЕТ ЧЕЛОВЕК

- Поднадзорность
- Ответственность

4. ТЕХНОЛОГИИ ИИ НУЖНО ПРИМЕНЯТЬ ПО НАЗНАЧЕНИЮ И ВНЕДРЯТЬ ТАМ, ГДЕ ЭТО ПРИНЕСЕТ ПОЛЬЗУ ЛЮДЯМ

- Применение СИИ в соответствии с предназначением
- Стимулирование развития ИИ

5. ИНТЕРЕСЫ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ ВЫШЕ ИНТЕРЕСОВ КОНКУРЕНЦИИ

- Корректность сравнений СИИ
- Развитие компетенций
- Сотрудничество разработчиков

6. ВАЖНА МАКСИМАЛЬНАЯ ПРОЗРАЧНОСТЬ И ПРАВДИВОСТЬ В ИНФОРМИРОВАНИИ ОБ УРОВНЕ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ, ИХ ВОЗМОЖНОСТЯХ И РИСКАХ

- Достоверность информации о СИИ
- Повышение осведомлённости об этике применения ИИ

https://a-ai.ru/wp-content/uploads/2021/10/Кодекс_этики_в_сфере_ИИ_финальный.pdf

... **круг лиц ... принимающих участие в жизненном цикле СИИ** ... включая предоставление товаров и оказание услуг. К таким лицам относятся, в том числе:

- разработчики, создающие, обучающие, тестирующие модели/системы ИИ и разрабатывающие, реализующие такие модели/системы, программные и/или аппаратные комплексы и принимающие на себя ответственность в отношении их конструкции;
- заказчики (лицо или организация), получающие продукт или услугу;
- поставщики данных и лица, осуществляющие формирование наборов данных для применения их в СИИ;
- эксперты, осуществляющие измерение и/или оценку параметров разработанных моделей/систем;
- изготовители, осуществляющие производство СИИ;
- эксплуатанты СИИ, на законном основании владеющие соответствующими системами, использующие их по назначению и непосредственно реализующие решение прикладных задач с использованием СИИ;
- операторы (лицо или организация), осуществляющие работу СИИ;
- лица, принимающие участие в регуляторном воздействии на сферу ИИ, в том числе разработчики нормативно-технических документов, руководств, различных регуляторных положений, требований и стандартов в области ИИ;
- иные лица, действия которых потенциально могут повлиять на результаты действий СИИ или лиц, принимающих решения с использованием СИИ.

https://a-ai.ru/wp-content/uploads/2021/10/Кодекс_этики_в_сфере_ИИ_финальный.pdf

«АИС ЗНАЕТ О ПРОЕКТЕ, КОТОРЫМ УПРАВЛЯЕТ»

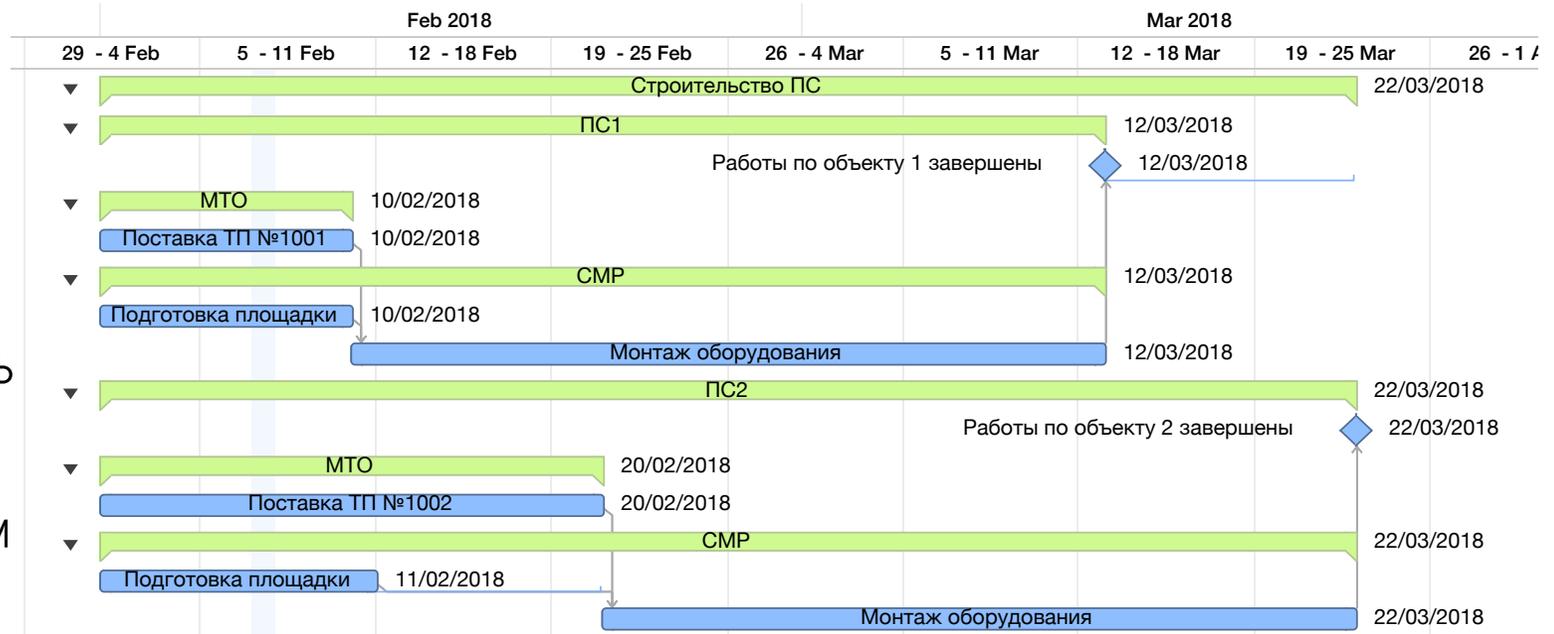
ИСКИН ДЛЯ РАБОТЫ С РИСКАМИ В ОБЛАСТИ ИИ?

АИС УПРАВЛЕНИЯ КРУПНЫМИ ПРОЕКТАМИ – КЕЙС (1/3)



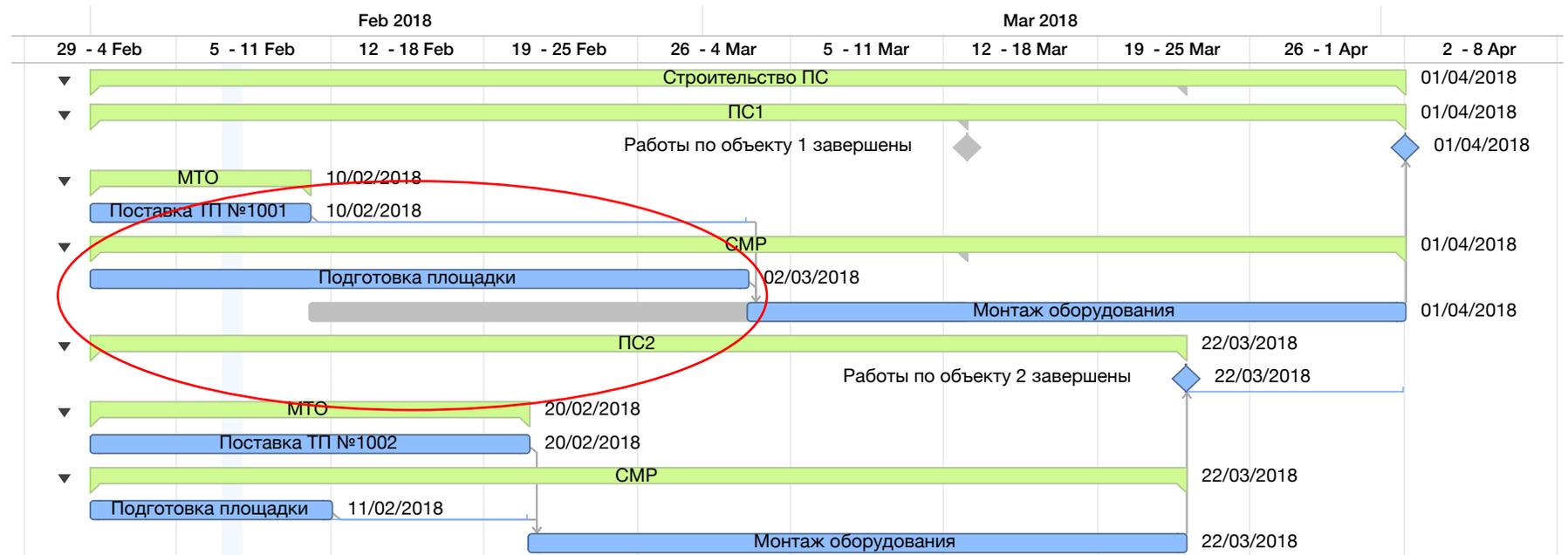
- На одной строительной площадке одновременно идет строительство 2-х электрических подстанций ПС1 и ПС2. График строительства каждого объекта состоит из 3-х работ:
 - Доставка Трансформаторной подстанции;
 - Подготовка площадки;
 - Монтаж оборудования.
- Заказчиком проекта поставлена задача «Как можно быстрее завершить строительство всех объектов»

- Монтаж оборудования возможен только после завершения работ 1 и 2. Работы 1 и 2 выполняются параллельно и не зависят друг от друга.
- На объекте «ПС1» поставка оборудования и подготовка площадки имеют одинаковую продолжительность и заканчиваются одновременно.
- На объекте «ПС2» поставка оборудования происходит позднее чем завершается подготовка площадки.



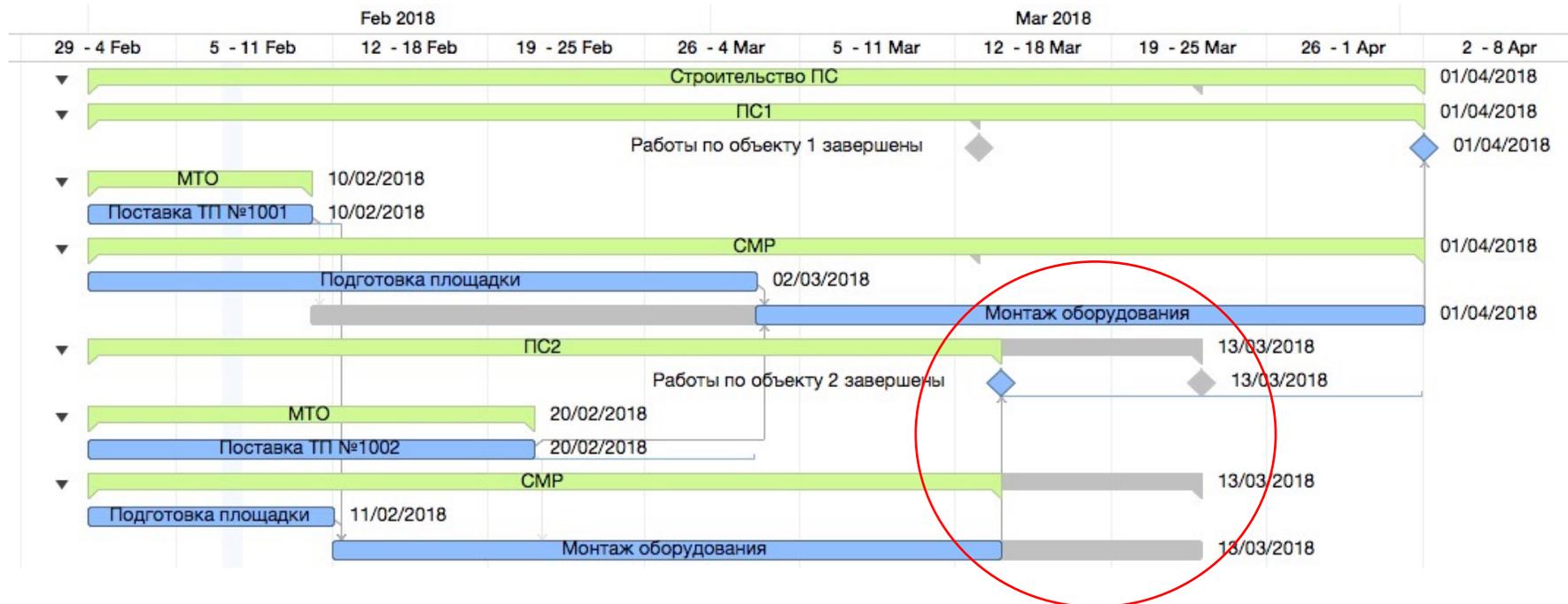
АИС УПРАВЛЕНИЯ КРУПНЫМИ ПРОЕКТАМИ – КЕЙС (2/3)

- В ходе производства работ на объекте «ПС1» произошла поломка техники, в следствии чего, интенсивность работы подрядчика снизилась в 3 раза.
- Продолжительность работы «Подготовка площадки» на объекте «ПС1» увеличилась в 3 раза.



АИС УПРАВЛЕНИЯ КРУПНЫМИ ПРОЕКТАМИ – КЕЙС (3/3)

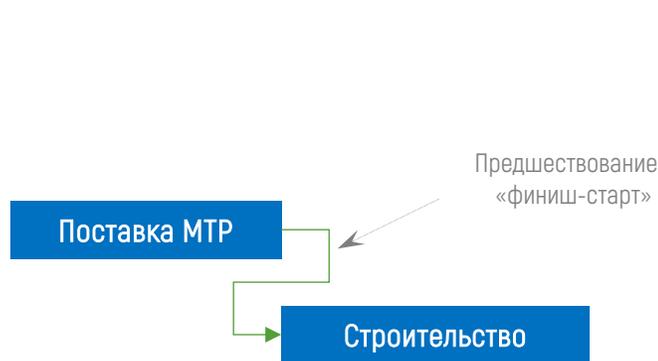
- В ходе производства работ на объекте «ПС1» произошла поломка техники, в следствии чего, интенсивность работы подрядчика снизилась в 3 раза.
- Продолжительность работы «Подготовка площадки» на объекте «ПС1» увеличилась в 3 раза.
- **Руководитель проекта принял предложение МАС поменять поставку оборудования между объектами и таким образом сократить сроки по объекту «ПС2».**



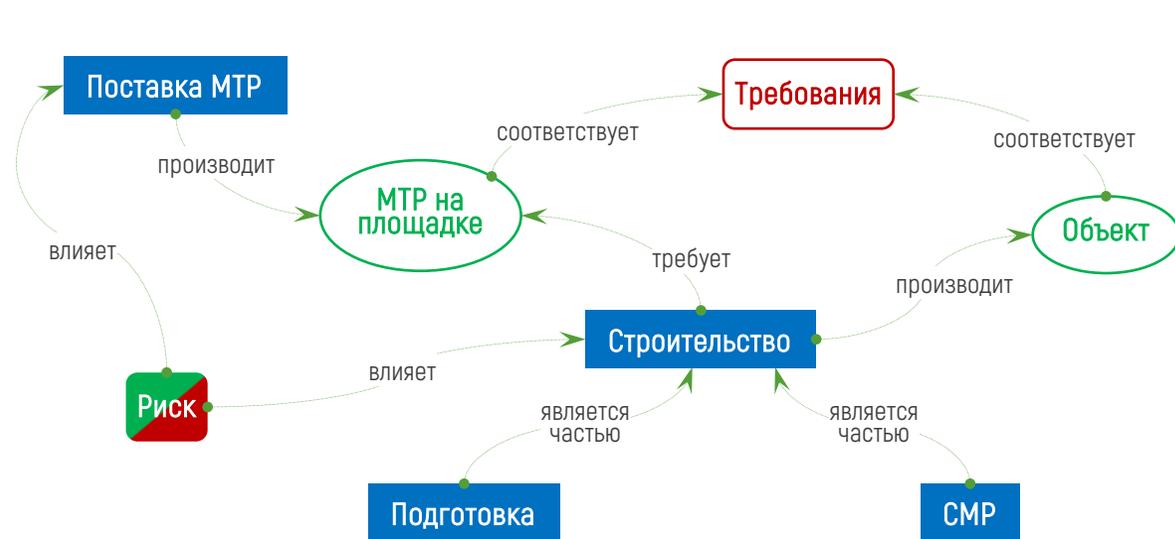
ПРИМЕР АИС УПРАВЛЕНИЯ ПРОЕКТАМИ

- Цель – эффективное управление реализацией проектов (минимальные сроки и стоимость, максимальное соответствие требованиям к результату)
- Гибкое формирование и исполнение графика проекта за счет:
 - Вариативности формирования календарно-сетевой модели проекта за счет перехода от жестких связей предшествования к отношениям в семантической сети – связи «работа»-«результат» и «ресурс»-«работа»
 - Вариативности представления работ в календарно-сетевой модели проекта как «одна большая работа» или «несколько составляющих большой работы»
 - Вариативности требований
 - Вариативности способов предотвращения и реагирования на риски проекта

«как есть»



«как будет»





CALL TO ACTION ?

ГРАЧЕВ Сергей Павлович
Генеральный директор

НАО "ГК "Генезис знаний"

121205, г. Москва, ИЦ Сколково, Большой бульвар, 42с1, офис 1680

mobile: +7 985 233 0437

email: sg@kg.ru

www: kg.ru

