

НАЦИОНАЛЬНЫЙ ЦЕНТР
КОГНИТИВНЫХ РАЗРАБОТОК
УНИВЕРСИТЕТА ИТМО

НЦКР

ДУМАТЬ ЧТОБЫ ДЕЙСТВОВАТЬ

Цифровая трансформация научной деятельности: что такое перспективные технологии eScience

Александр Валерьевич Бухановский

Санкт-Петербург, 2021

Приоритет 20А Стратегии научно-технологического развития Российской Федерации – Указ Президента Российской Федерации от 01.12.2016 № 642

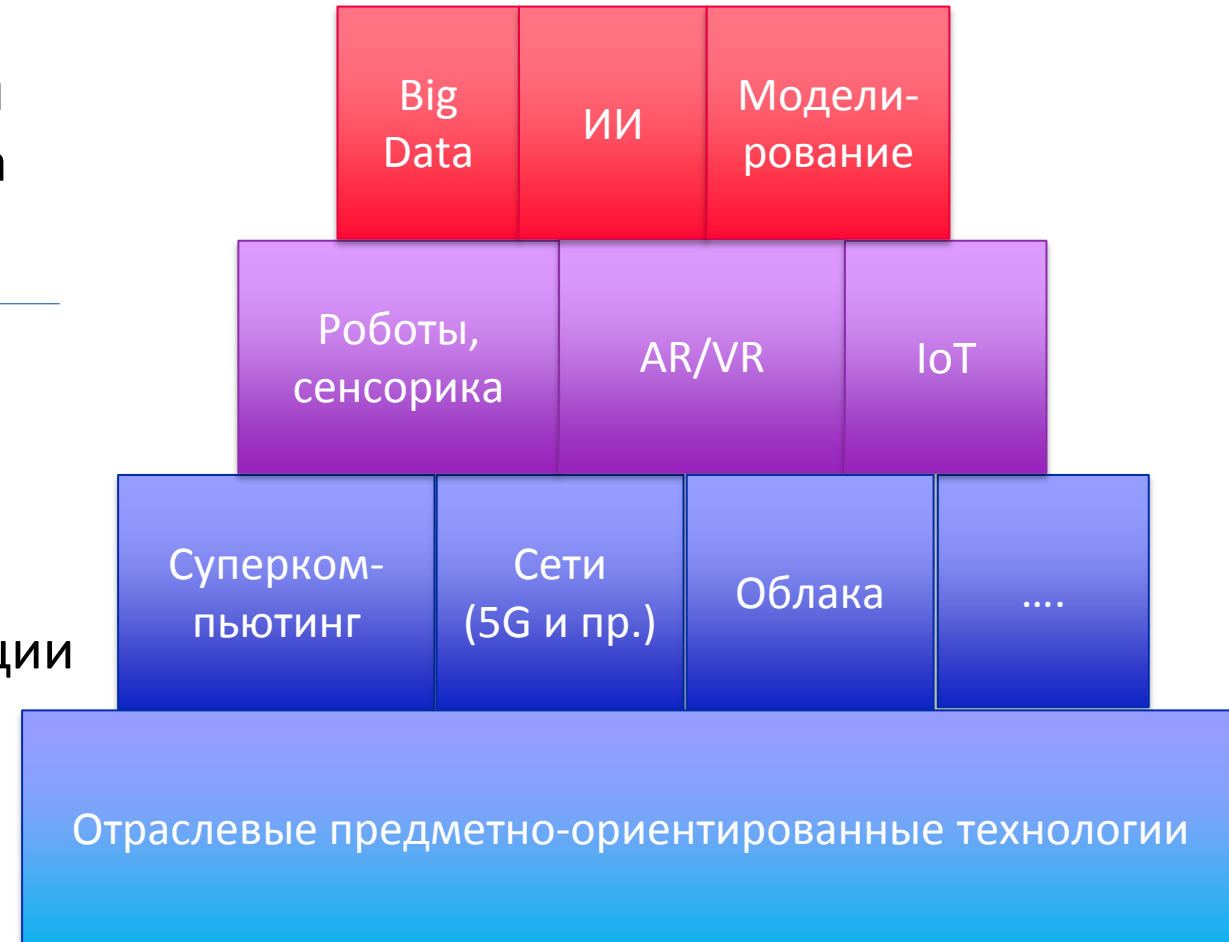
Переход к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, к новым материалам и способам конструирования, создание систем обработки больших объемов данных, машинного обучения и искусственного интеллекта:

- ❖ Системная и нетравматичная цифровая трансформация реального сектора отечественной экономики
- ❖ Повышение производительности труда за счет комплексного использования цифровых технологий
- ❖ Повышение отдачи от капитальных вложений в развитие производства за счет перехода к цифровым технологиям организации производства
- ❖ Повышение устойчивости отечественной экономики за счет обеспечения превентивности управления ею на основе цифровых технологий
- ❖ Обеспечение прозрачности социально-экономических процессов за счет их перевода в цифровую форму, и оптимизация систем управления разных уровней
- ❖ Обеспечение цифрового равенства и цифровой конкурентоспособности граждан

Процесс *глубокого преобразования* всех аспектов жизни человека с помощью цифровых технологий, приводящий к кардинальным изменениям технологий, культуры, операций и принципов создания новых продуктов и услуг, а также рынка труда

- 1) Взрывное развитие средств сбора и хранения данных
- 2) Виртуализация бизнес-процессов
- 3) Ускорение процессов принятия решений
- 4) Невозможность содержательной интерпретации цифровых объектов (сложность)
- 5) Распространение информации со скоростью выше скорости интерпретации человеком
- 6) Порождение новых форм социальной деятельности в цифровом виде

«Пищевая пирамида» технологий цифровой трансформации





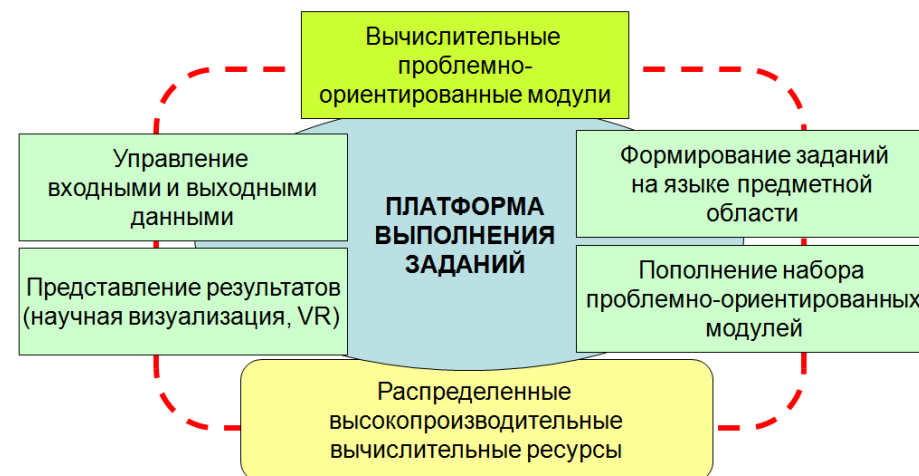
А как это касается ученых мужей? (особенно тех, кто и так занимается «цифрой»)

eScience («электронная» наука), или «киберинфраструктура»

Обобщенное определение eScience - John Taylor, 1999:

*The term **e-Science** (or **eScience**) is used to describe computationally intensive science that is carried out in highly distributed network environments (научные исследования с большим объемом вычислений в распределенных системах), or science that uses immense data sets (большие массивы данных) that require grid computing; (грид-вычисления) the term sometimes includes technologies that enable distributed collaboration (технологии объединения распределенных электронных ресурсов)*

PSE (Problem Solving Environment) – «интегрирующая» технология eScience, ориентированная на использование высокопроизводительных вычислительных ресурсов



Поколение I - цифровые средства коллективного доступа "в ручном режиме" к разнообразным вычислительным ресурсам, сервисам, ПО, базам данных, библиотекам, лабораторному оборудованию и пр. (90е XX века)

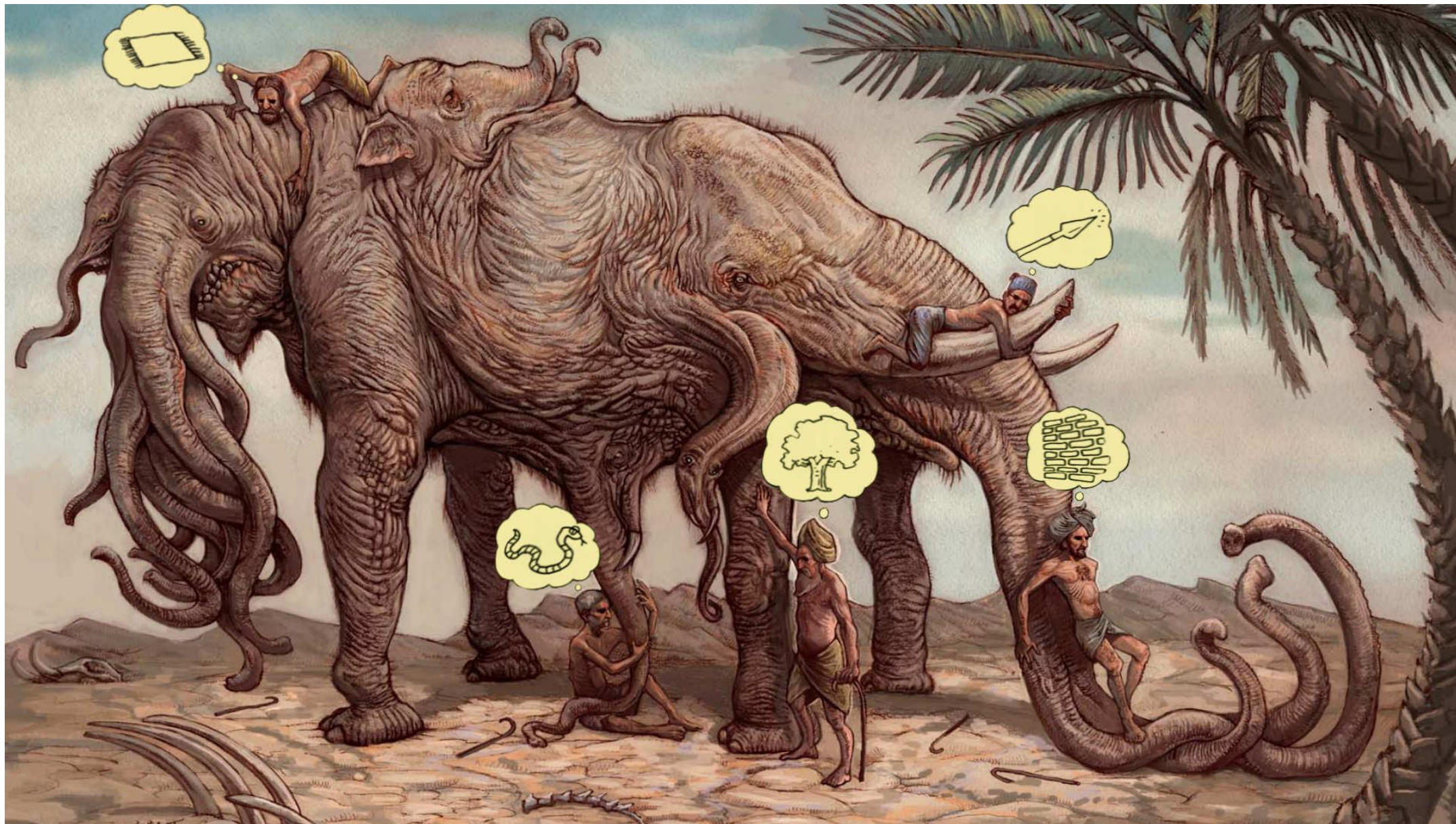
Поколение II - механизмы бесшовной агломерации объектов коллективного доступа (в том числе, на основе интеллектуальных технологий), облегчающие процессы взаимодействия между пользователями и совместного использования ресурсов для решения собственных задач (WMS, семантические сервисы, облачные планировщики и пр.). (2000е)

Поколение III - механизмы, реализующие логику т.н. "4-й парадигмы науки" - автоматизация процессов получения новых знаний на основе данных (извлекаемых или синтезируемых на решениях I и II поколения) (2010е).

Поколение IV – механизмы цифровой трансформации научной деятельности посредством интеллектуальных технологий (2020е).

Как отличить цифровую трансформацию от обычной автоматизации в науке?

(трансформируется ли подход к исследованию vs. есть ли ИИ++?)



Искусственный интеллект – ЭТО:

цифровое воспроизведение процессов **сознательной активности** человека и социума в целом

в части **творческой** обработки и рассуждений на основе нетривиально формализуемой информации

в условиях временных и ресурсных **ограничений неопределенности и неполноты** исходных данных,

создающее кибернетические объекты, способные самостоятельно **ставить цели и их достигать** с качеством не ниже среднего специалиста,

способное в перспективе **заменить** существующие виды деятельности и профессии

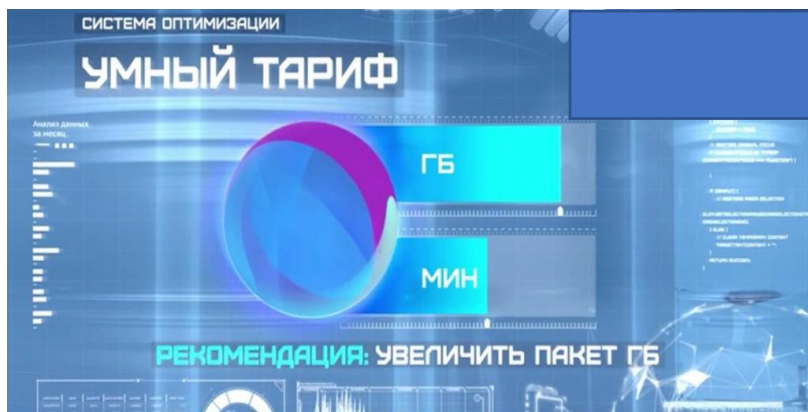
Он нужен ДЛЯ:

- 1) освобождение человека от рутинных действий (**нет сил**);
- 2) поддержка принятия решений в экстремальных условиях (**нет времени**);
- 3) устранение коммуникационных барьеров (**нет понимания**);
- 4) автоматизация получения новых фундаментальных знаний (**нет... ума...**).

ИИ = «сознание» + «цели» + «неопределенность» + «творчество»

ИИ = «сознание» + «цели» + «неопределенность» + «творчество»

Системы на основе
формальной обработки
данных



А ГДЕ ТВОРЧЕСТВО?

Системы на основе
математического
моделирования

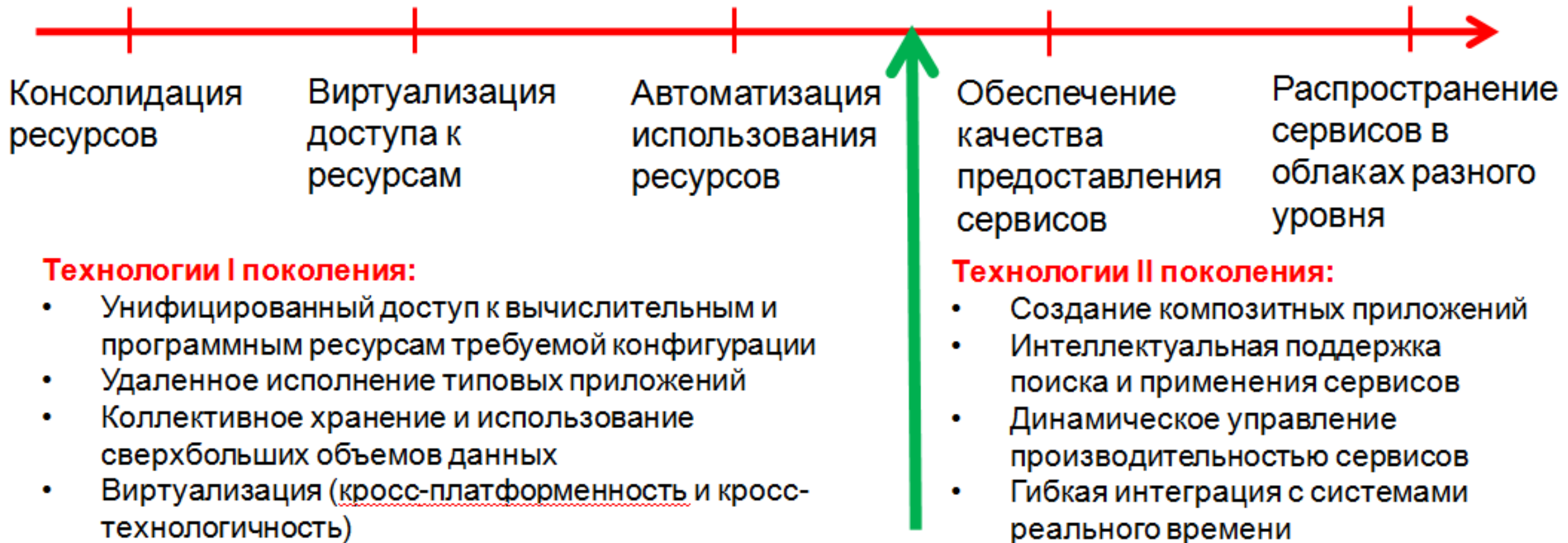


ИИ работает с задачами, для которых **не придуманы** формальные алгоритмы решения!



ТРАНСФОРМАЦИЯ ПРОЦЕССА РАБОТЫ С ВЫЧИСЛИТЕЛЬНОЙ ИНФРАСТРУКТУРОЙ

Cloud Computing Maturity Model – CCMM



Организация взаимодействия через систему управления потоками заданий (Workflow Management Service)

Формально основой традиционной **WMS** является НАГ (направленный ациклический граф), хотя в настоящих решениях существуют циклы и условные блоки.

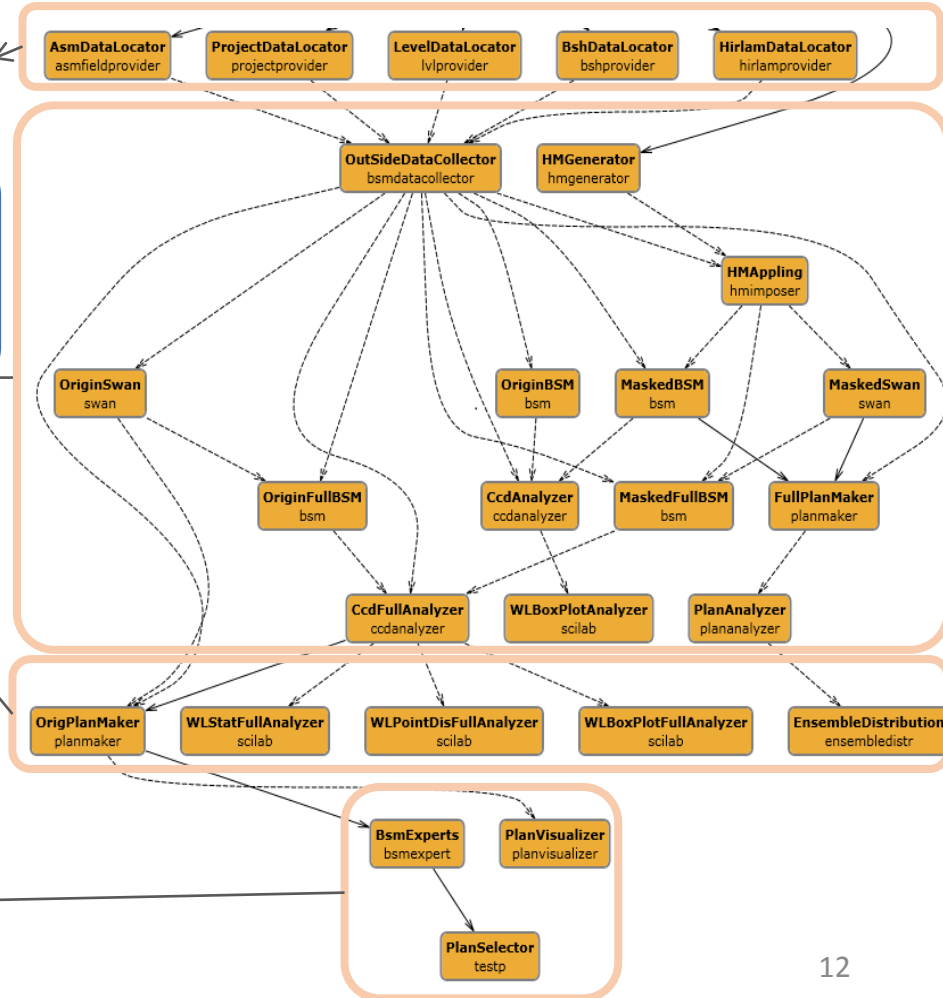
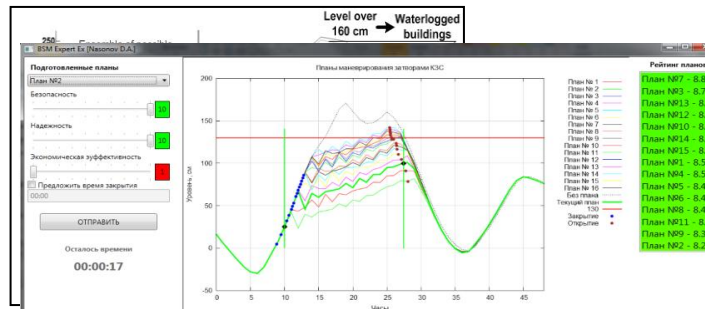
Благодаря обширной разработке механизмов планирования, проблема оптимизации ориентированных на работу приложений для ресурсоемких (по данным) вычислений также успешно решается.

Сенсоры данных

• **Маска с шумом** накладывается на скорость ветра, инициализируя ансамбль моделирования ветровых волн и уровней воды в Балт. море.

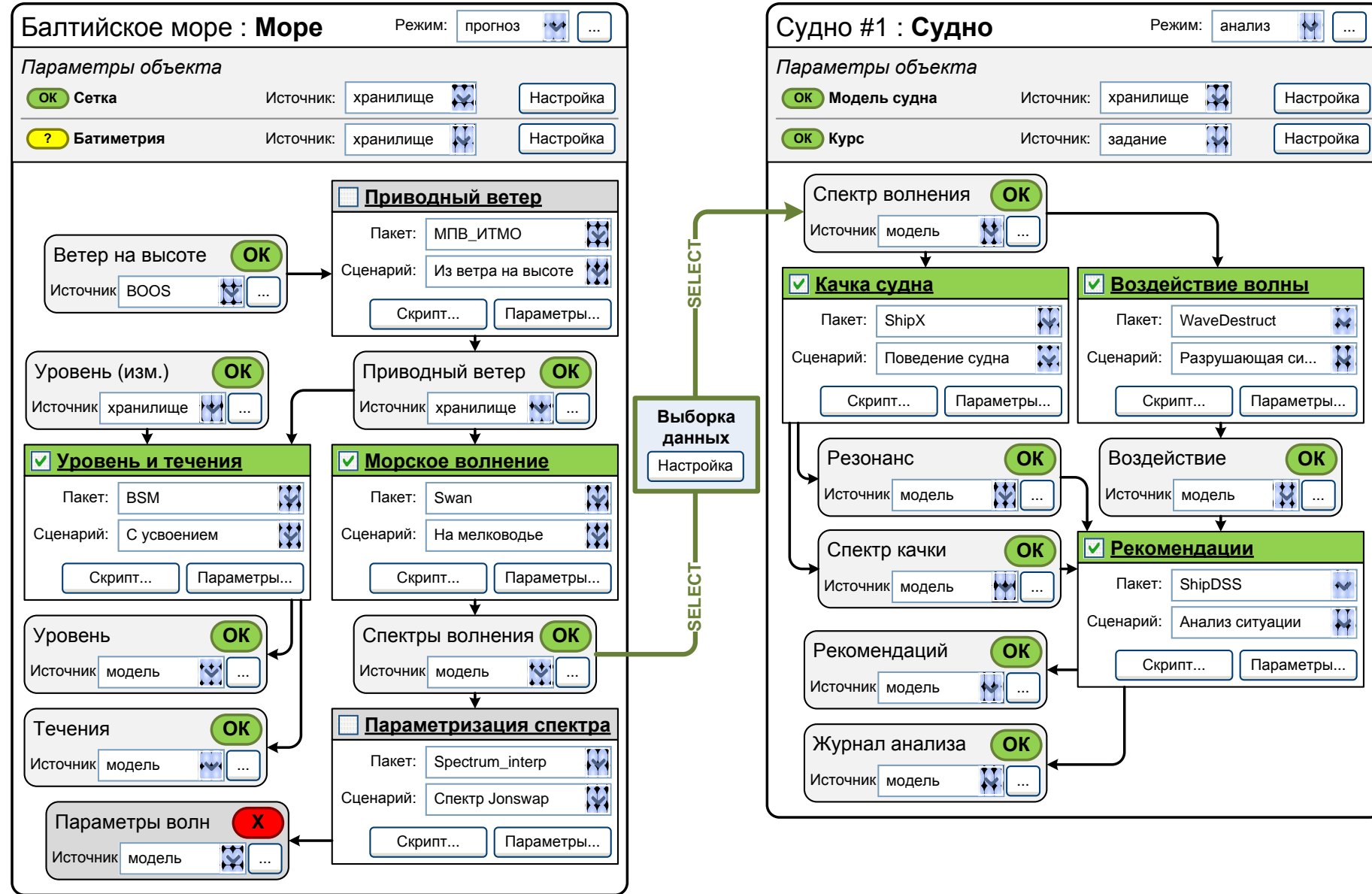
• **Первые результаты** агрегации и визуализации. Выполнение процедуры принятия решения по плану закрытия ворот.

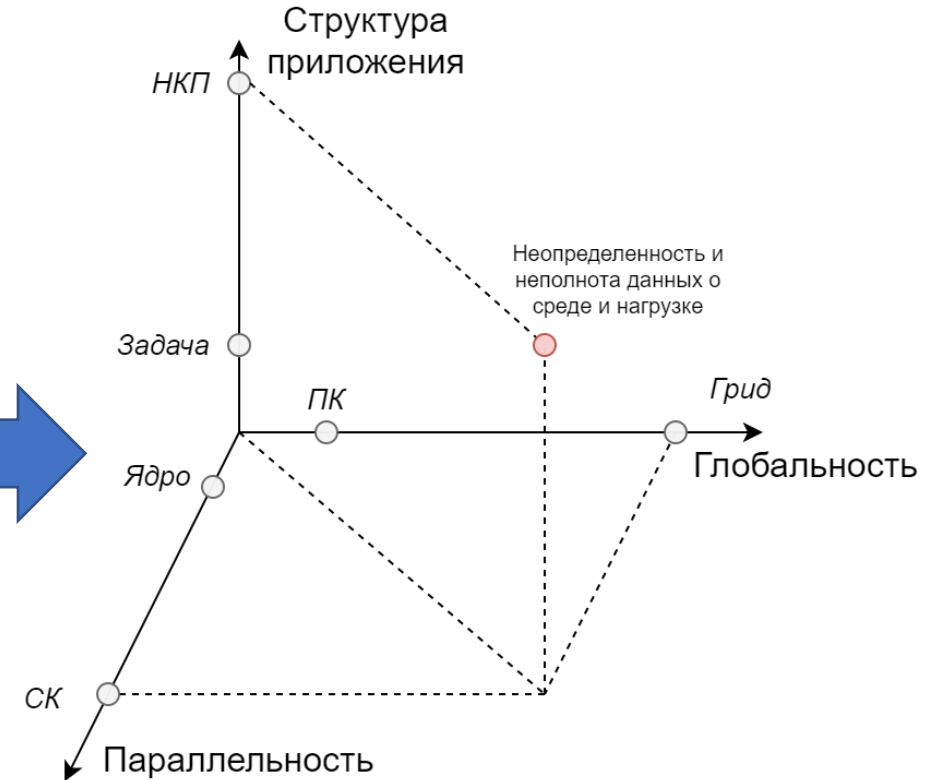
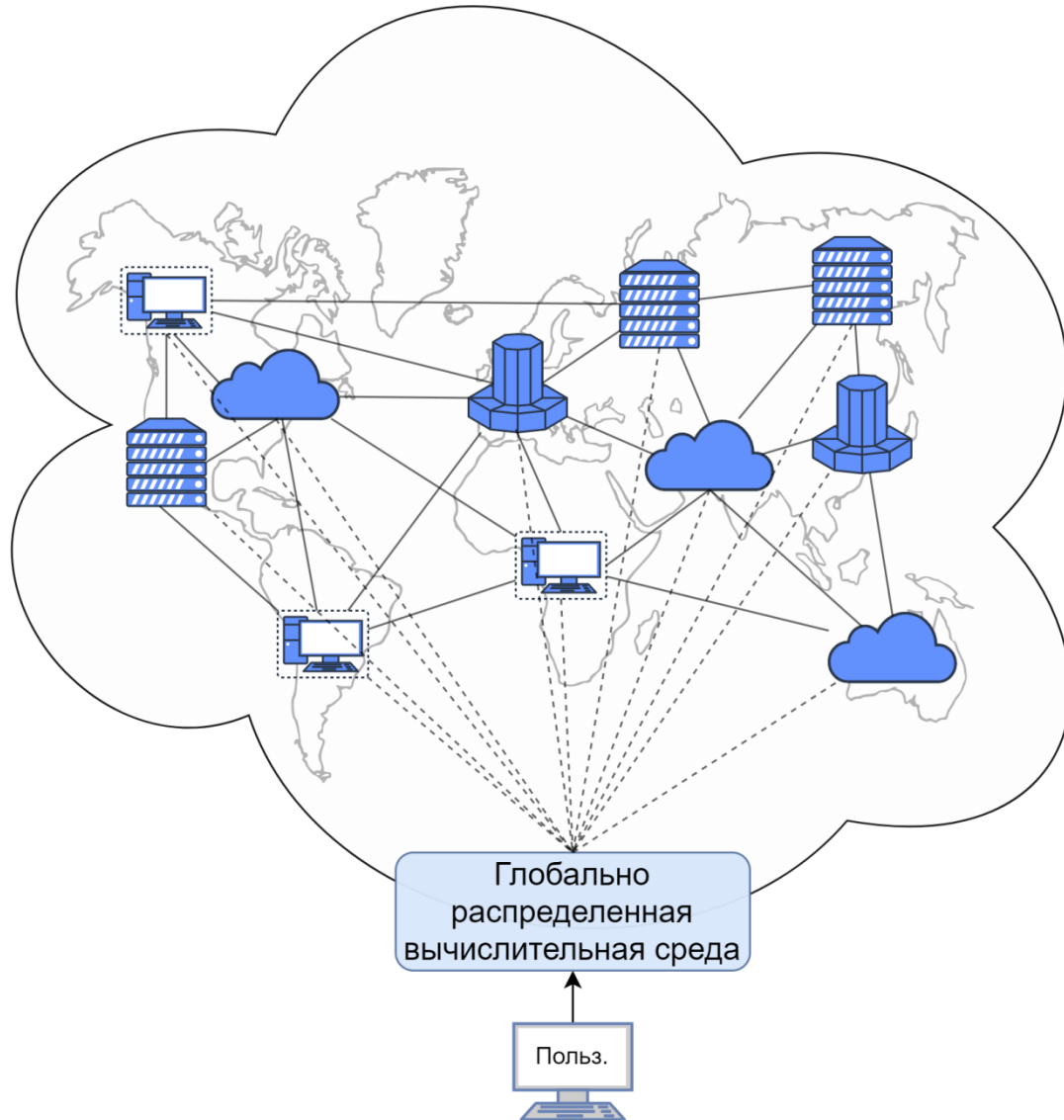
Генерация окончательного плана:



Виртуальные моделирующие объекты в платформе CLAVIRE:

- No-code-формирование цифровых объектов
- Автогенерация мета-WF
- Автогенерация конкретного WF
- Обеспечение процесса исполнения приложения в распределенной среде





Неопределенность и неполнота данных о работе облачной среды возникает при усложнении или увеличении:

- Глобальности;
- Параллельности;
- Масштаба приложения.

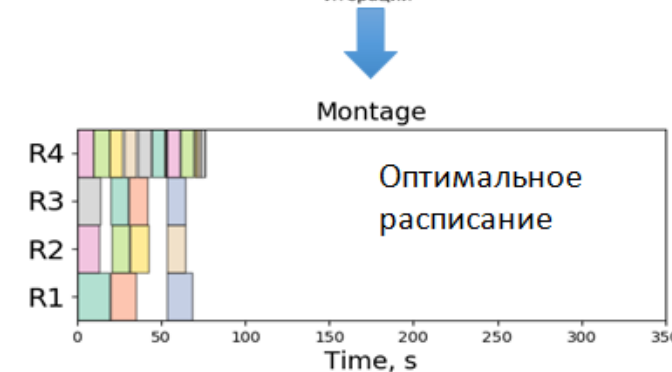
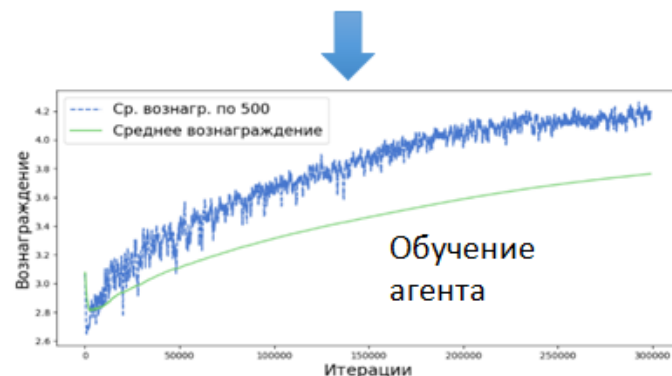
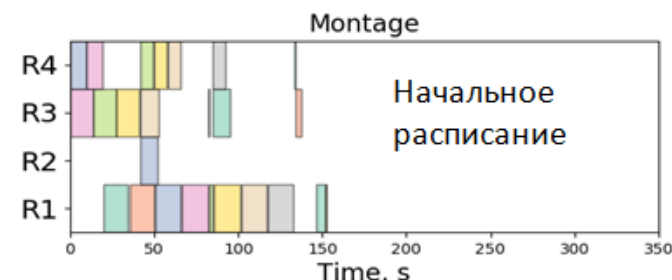
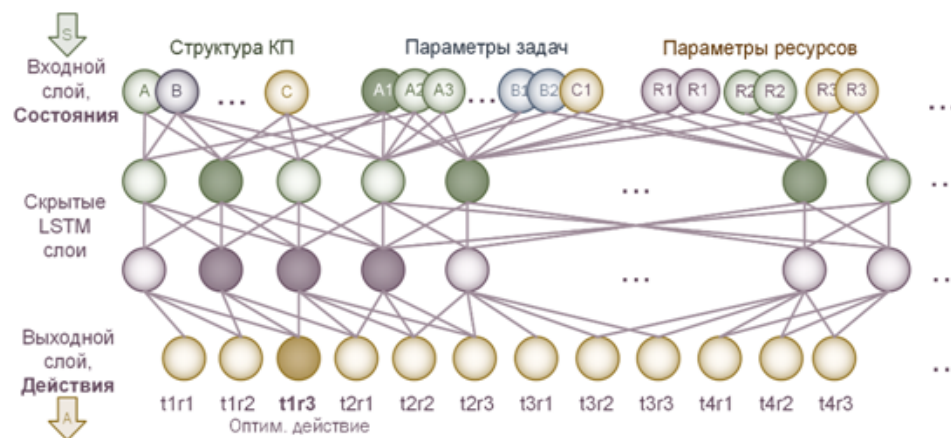
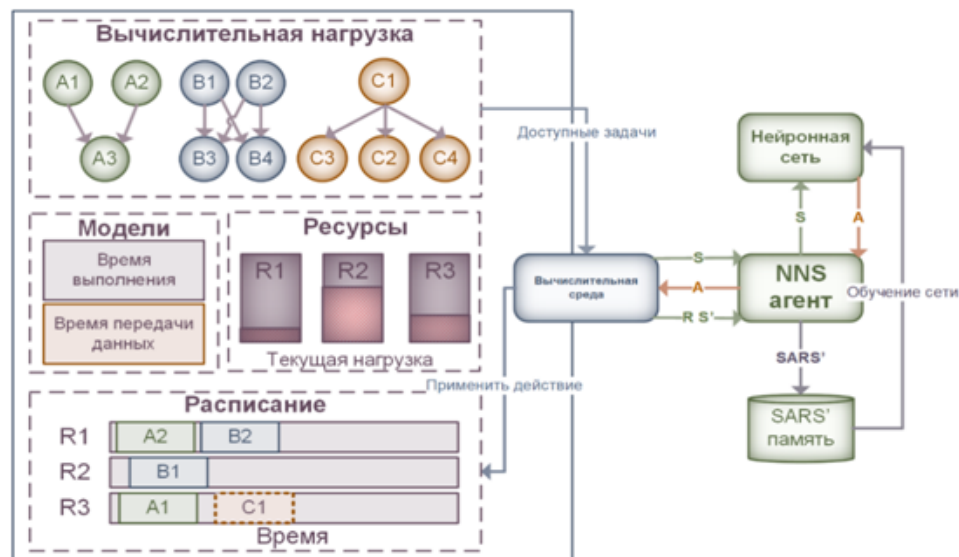
Задача: решение проблемы оптимального распределения вычислительной нагрузки по вычислительным ресурсам

Решение: создание «планировщика» композитных приложений

Алгоритм использует:

- Принцип обучения с подкреплением (SARSA агент, Q-Learning)
- Нейронные сети в качестве аппроксиматора функции полезности

Результат: Достижение производительности, эффективности, масштабируемости и адаптивности алгоритма планирования

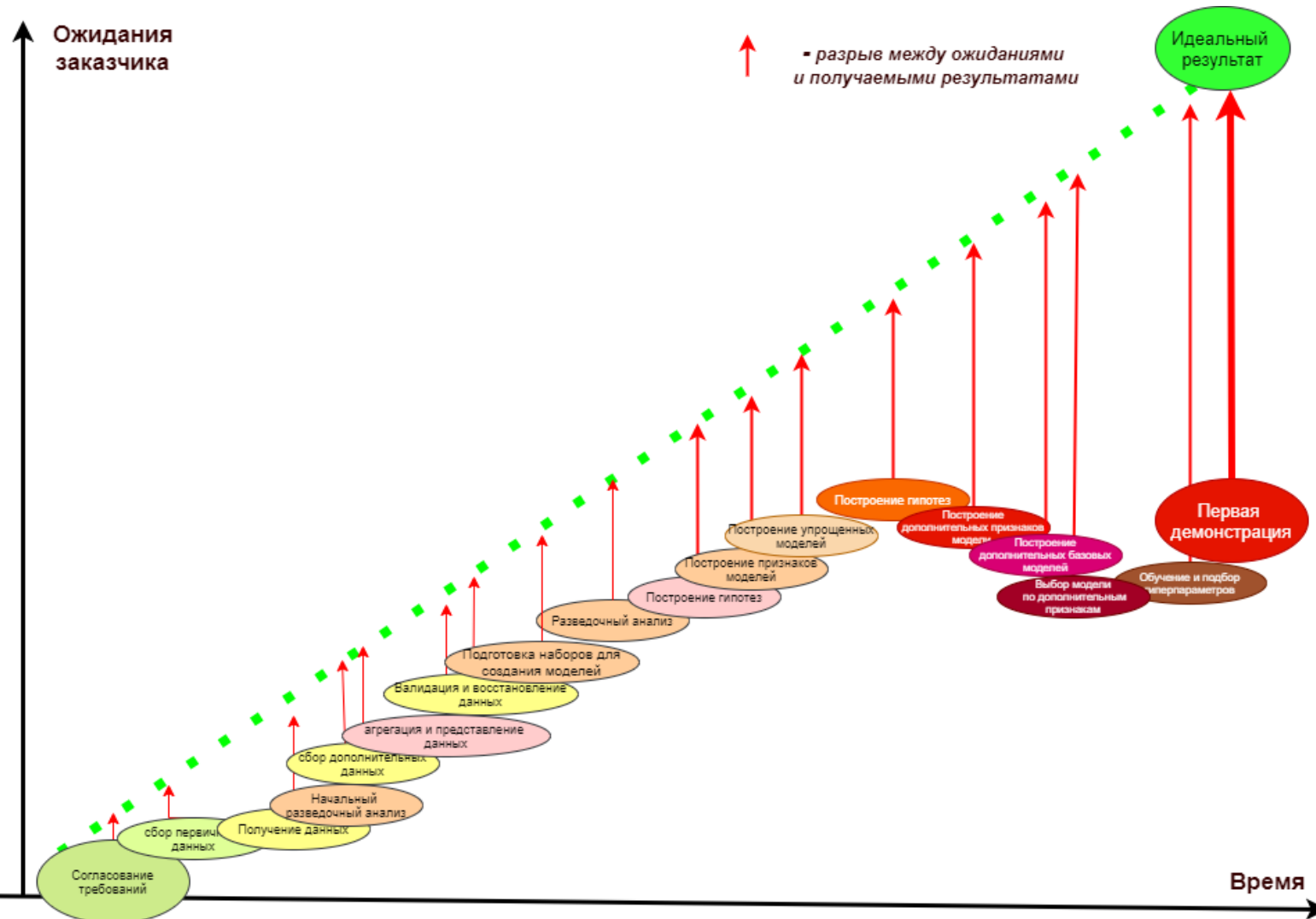




ТРАНСФОРМАЦИЯ ПРОЦЕССА КОЛЛЕКТИВНОЙ РАБОТЫ С ДАННЫМИ И МОДЕЛЯМИ

Риски в реализации проектов Data Science

1. Быстрые проекты - нет деления на разведочный и основной проекты.
2. Отсутствие предметной квалификации команды исполнителей.
3. Отсутствие данных или возможность создать корректные наборы данных.
4. Дорогие вычислительные ресурсы и ресурсы память.
5. Неопределенность в данных и закономерностях, что являются основой в построении моделей МО.
6. Неопределенность в качестве, как построения так и использования моделей.
7. Отсутствия единых практик процесса формирования, разработки и эксплуатации моделей МО.
8. Отсутствие коллабораций и механизмов переиспользования знаний, генерируемых артефактов.

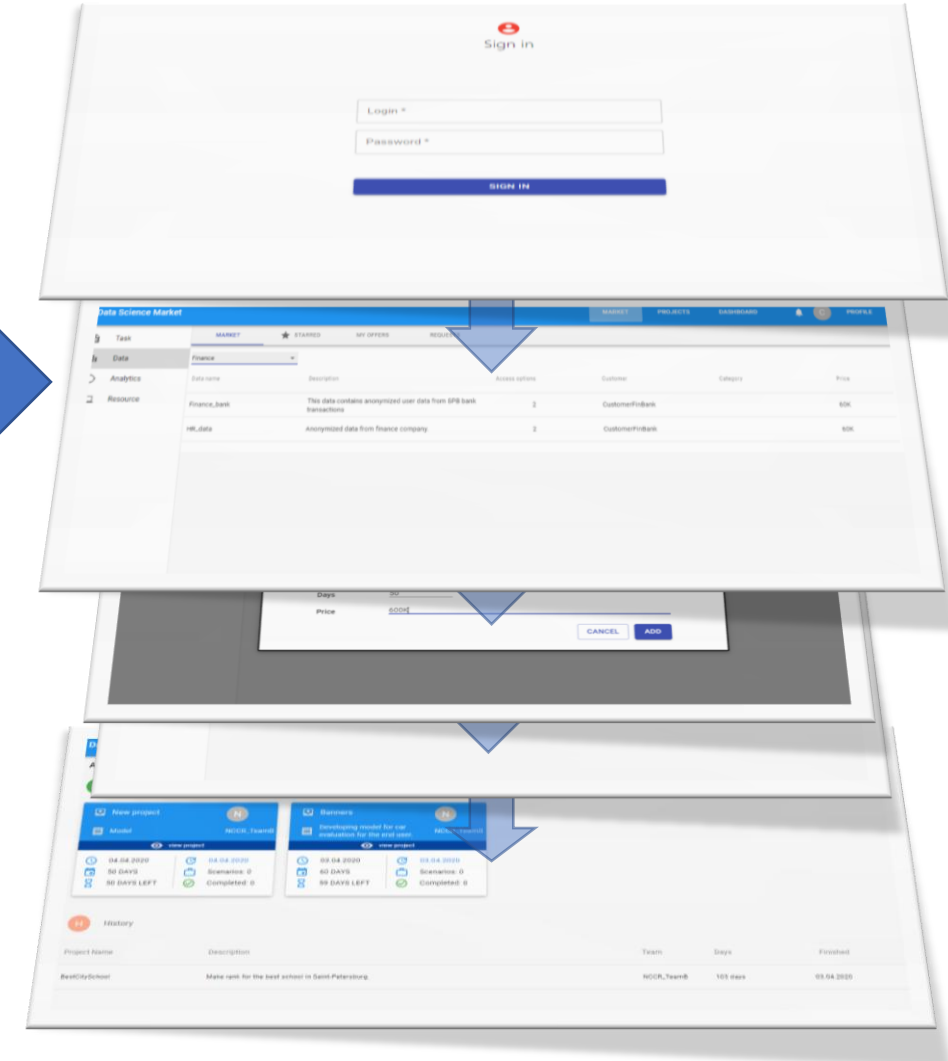
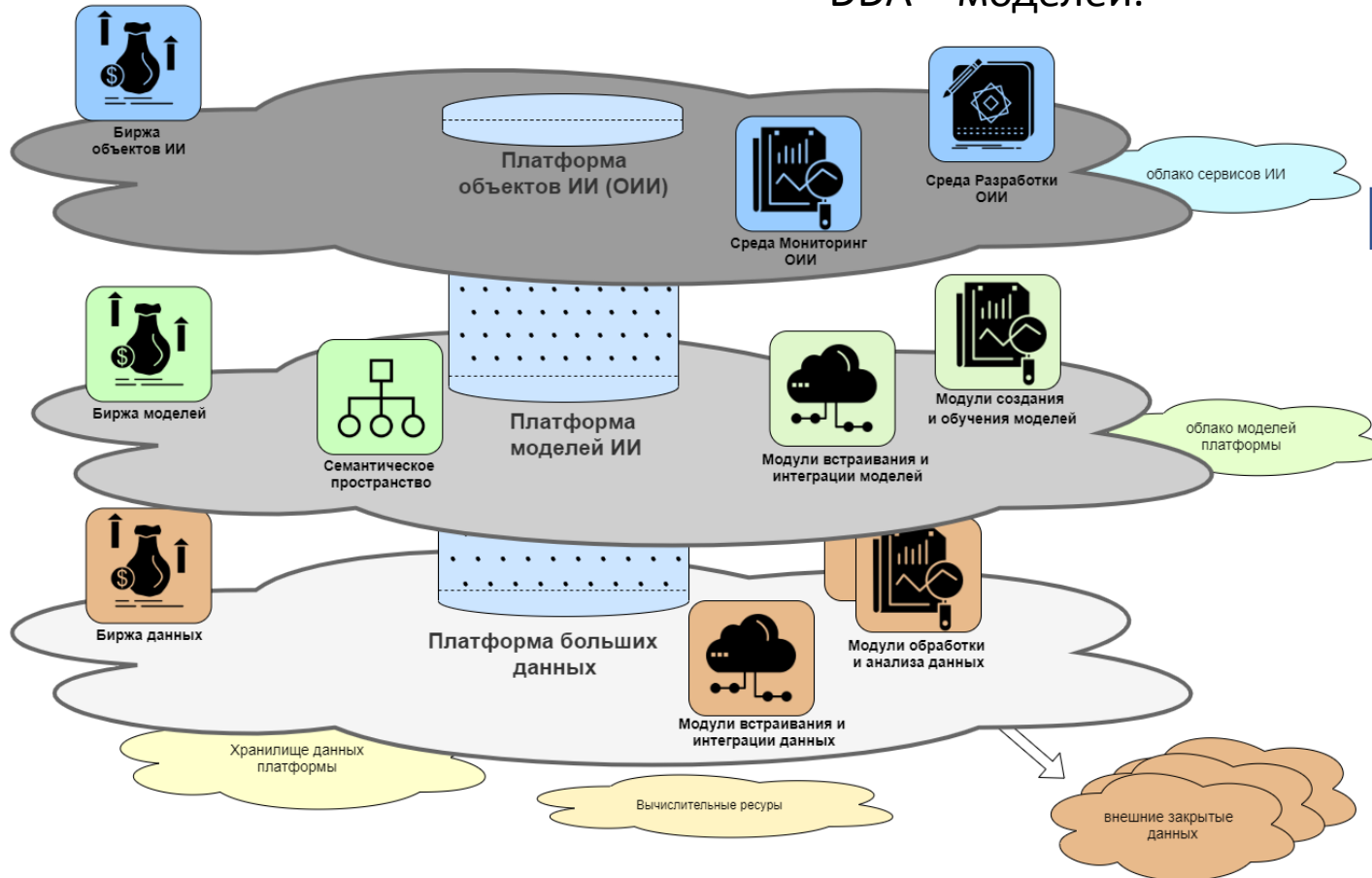


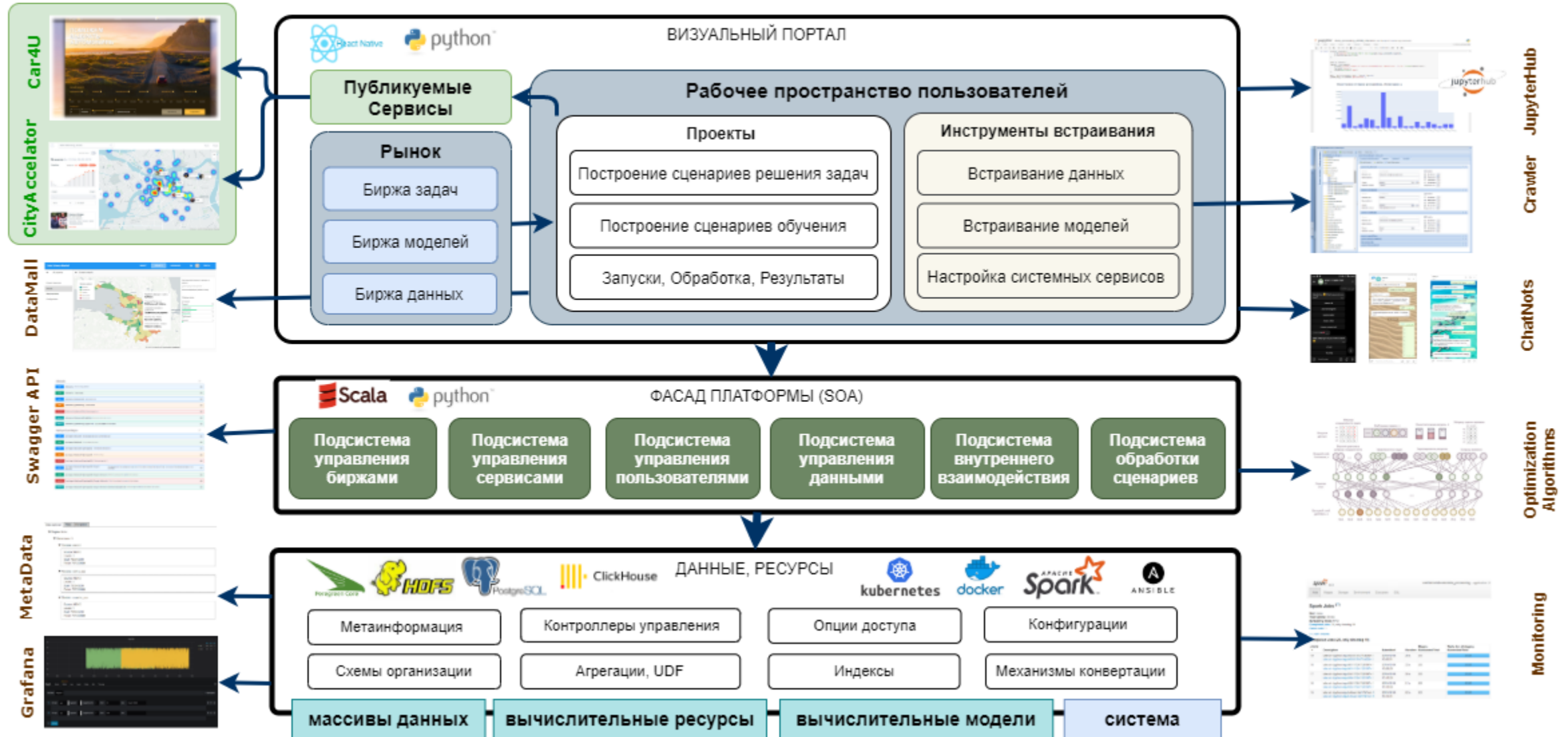
Уровни реализации платформы:

- уровень данных;
- уровень моделей;
- уровень готовых объектов.

Направления использования:

- платформа коллабораций;
- биржа данных, артефактов;
- вычислительная платформа DDA – моделей.

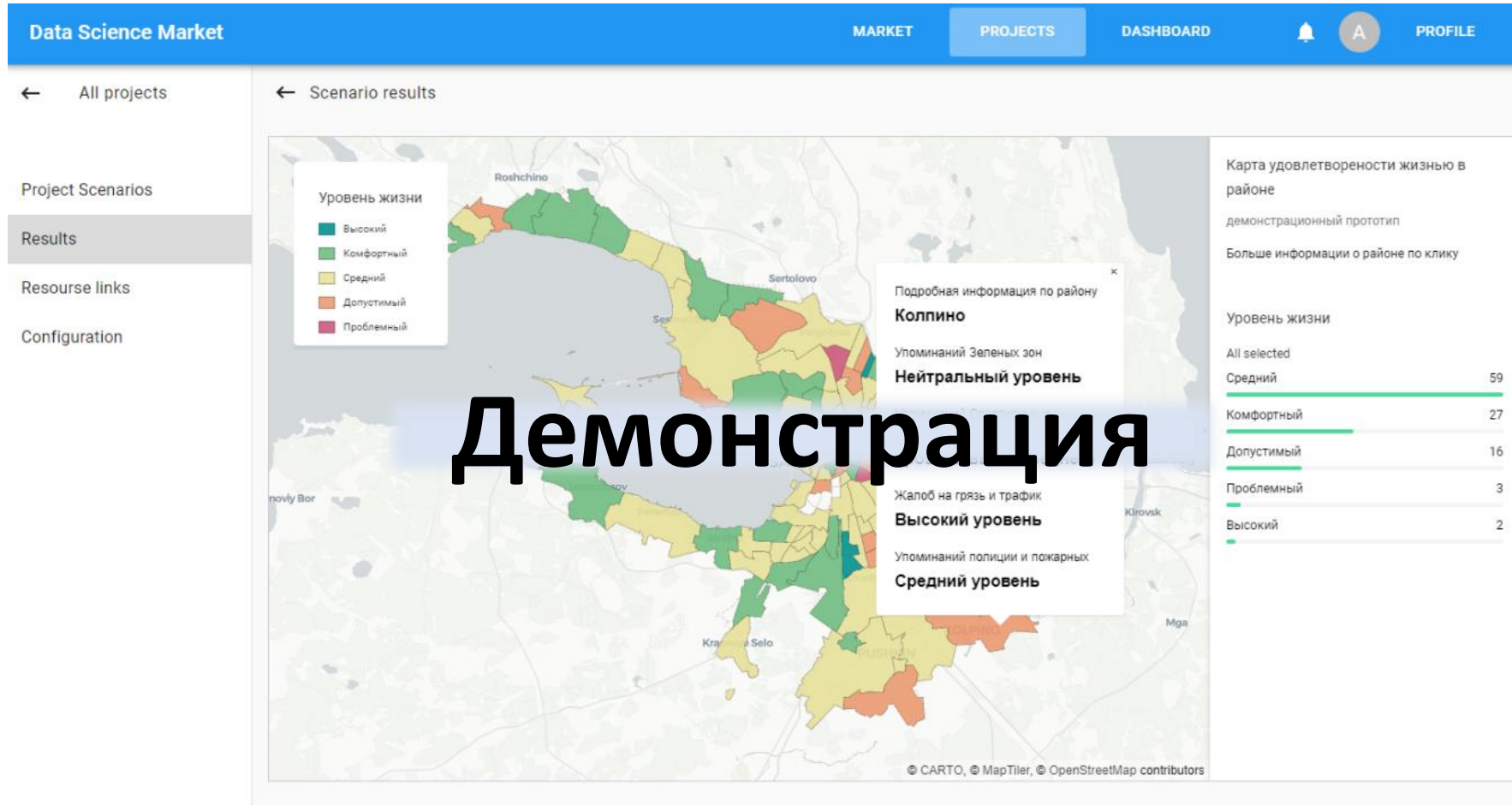




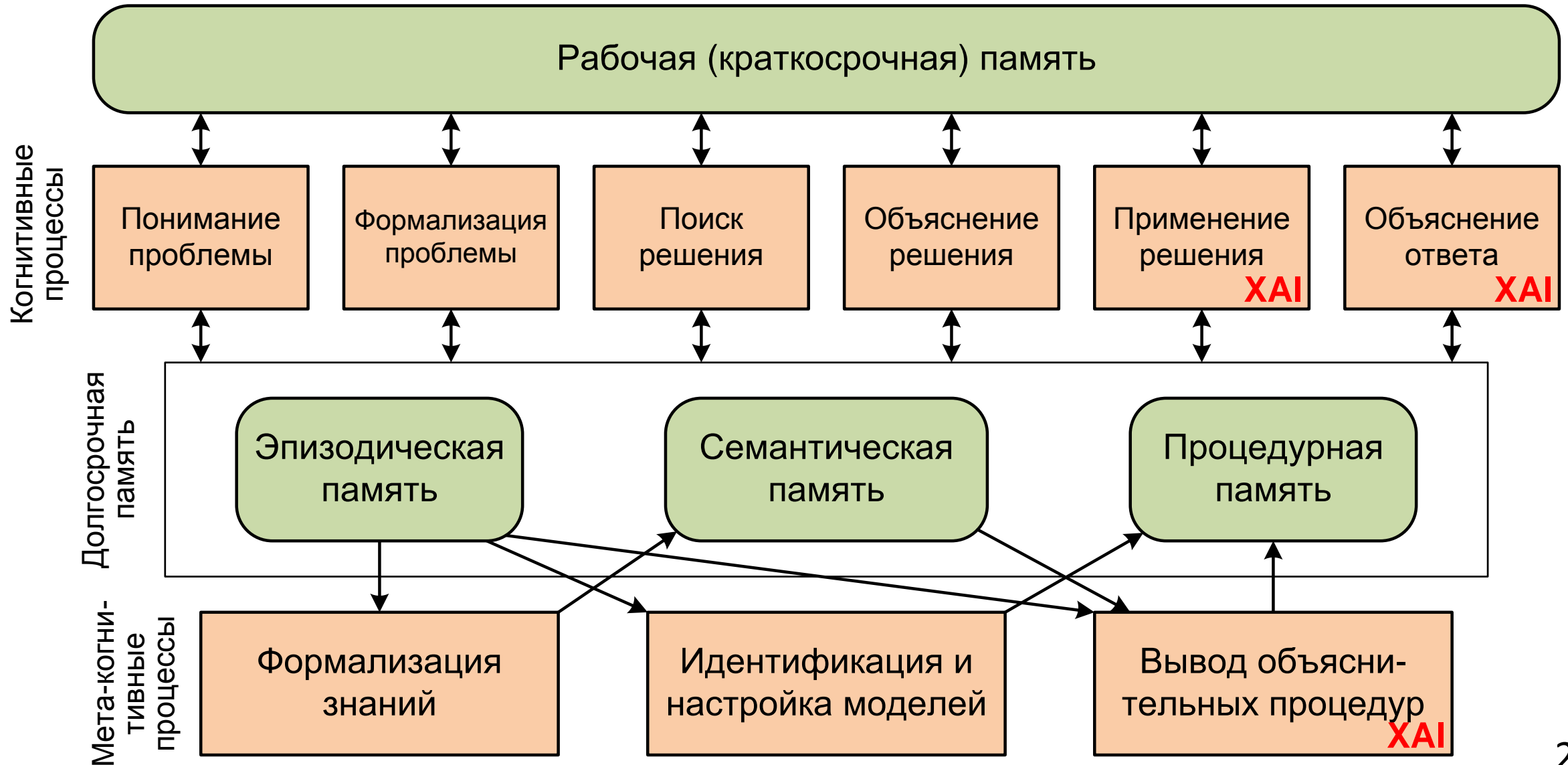
DataMall сейчас:

- механизм бирж;
- МО проекты на базе с JupyterHub;
- краулинг данных более 20 источников;
- вычислительная инфраструктура на базе Kubernetes, Spark, HDFS и ClickHouse;
- более 15 проектов*.

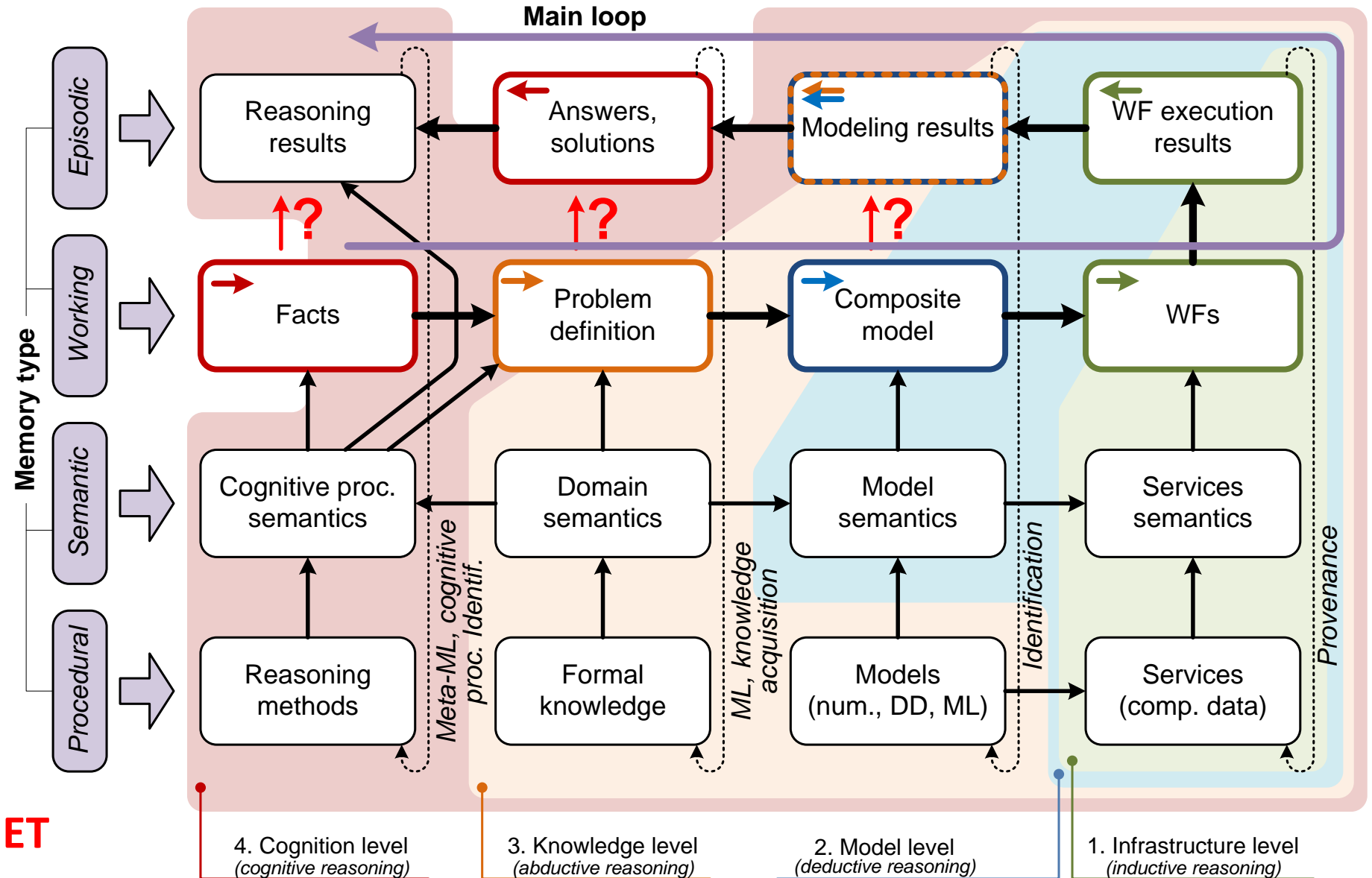




<https://actcognitive.org/platformy/datamall-platforma-podderzhki-zhiznennogo-tsikla-intellektualnykh-obektov-na-osnove-bolshikh-dannykh>



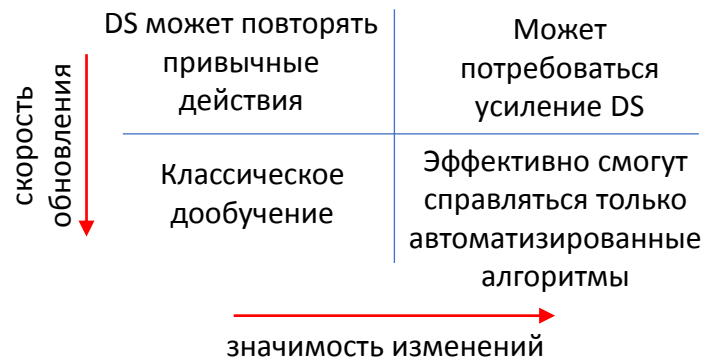
- Расширение принципов и подходов eScience на уровень когнитивной ОС
- Основной цикл – решение проблемы
- ИИ должен пройти путь от верхних уровней к нижним





ТРАНСФОРМАЦИЯ ПОДХОДА К СОЗДАНИЮ ЦИФРОВЫХ ОБЪЕКТОВ

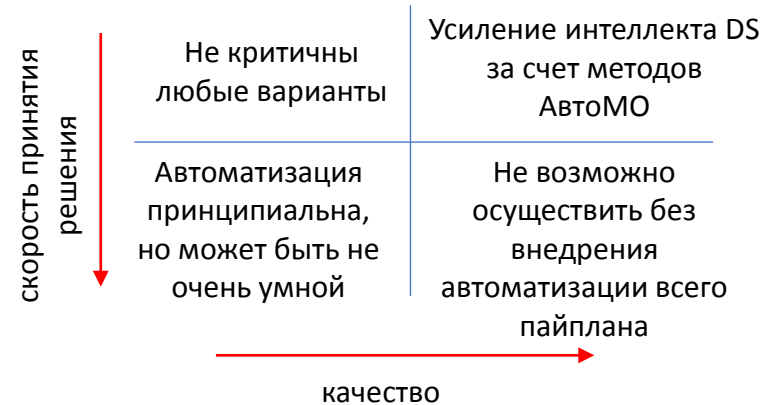
1. Данные обновляются быстро и значительно (по номенклатуре и/или вероятностным характеристикам).



В чем специфика?

- Если данные обновляются «пакетно» с большими интервалами, то их обработку может выполнять специалист.
- Если данные обновляются непрерывно или «мини-пакетно», однако основные вероятностные характеристики меняются плавно, то подойдет классическое дообучение.
- Если характеристики данных меняются значительно, то потребуются структурные изменения в модели или во всем пайплайне решения.**

2. Много мелких задач, решения для которых нужно делать быстро с фиксированным качеством.



В чем специфика?

- Степень автоматичности МО решений может быть различной, важно не обмануться.
- В треугольнике «время-деньги-ошибки» можно выйти за границы привычной эффективности только с применением инновационных решений, основывающихся на серьезной автоматизации процессов.

3. Задача не решается «стандартной» последовательностью шагов и нужен компьютерный интеллект для поиска вариантов.

Или проблема DS с низкой или средней квалификацией.

В чем специфика?

- Нужны мощные инструменты, которые позволяют покрывать действительно огромное количество возможных сценариев и эффективно искать наиболее подходящие варианты.
- Требуются мощные вычислительные ресурсы.
- Требуются DS с высокой квалификацией, которые могут эффективно управлять инструментами АвтоМО, но в значительно меньшем количестве.

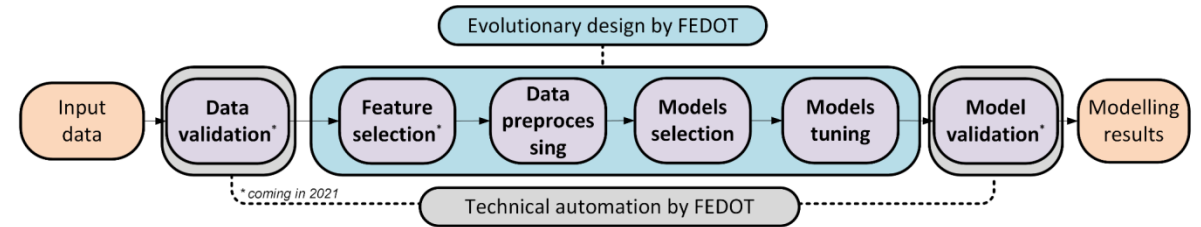
Особенности фреймворка:

- Variable Structure Pipelines
- Основа алгоритма - GP
- Поддержка Tabular, Timeseries данных
- Поддержка гибкой модификации и расширения модулей фреймворка

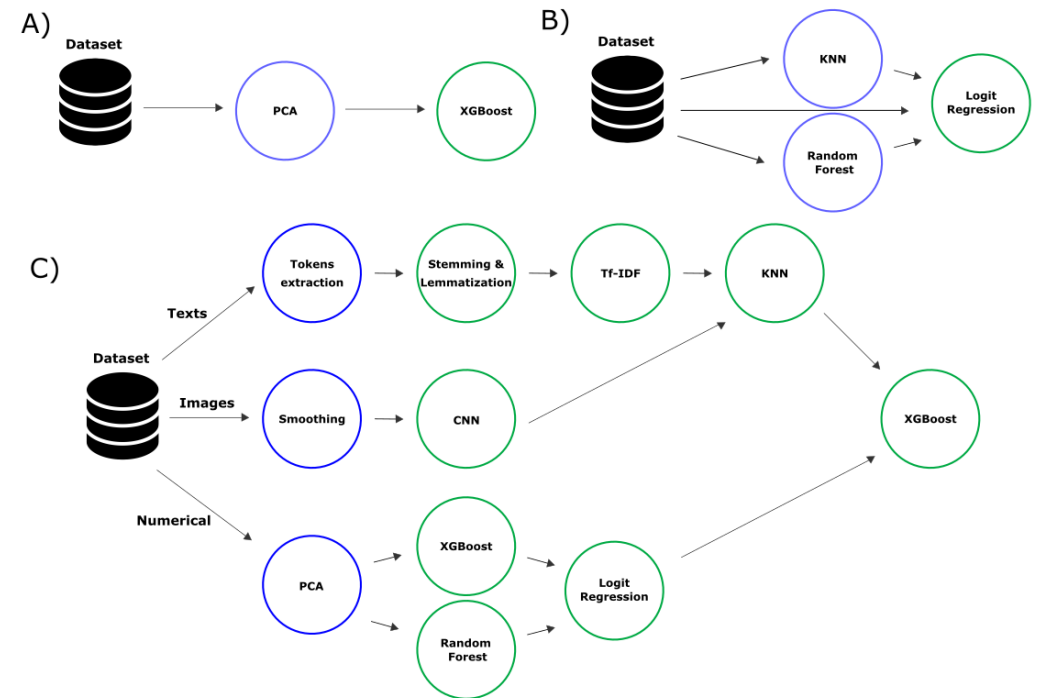
Основная идея: (а) поддержка широкого набора задач и типов данных и (б) возможность замены отдельных модулей

В разработке:

- Многокритериальная оптимизация
- Расширенная предобработка данных
- Поддержка NLP и CV задач
- Preset-ы для определенных типов задач
- WebGUI
- Масштабирование вычислений

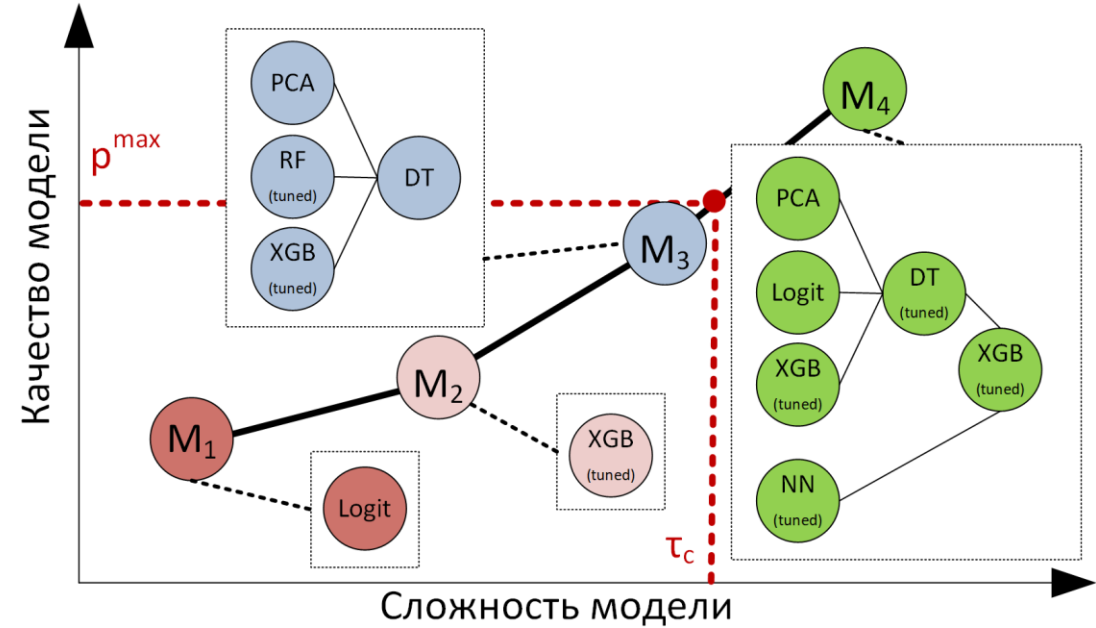
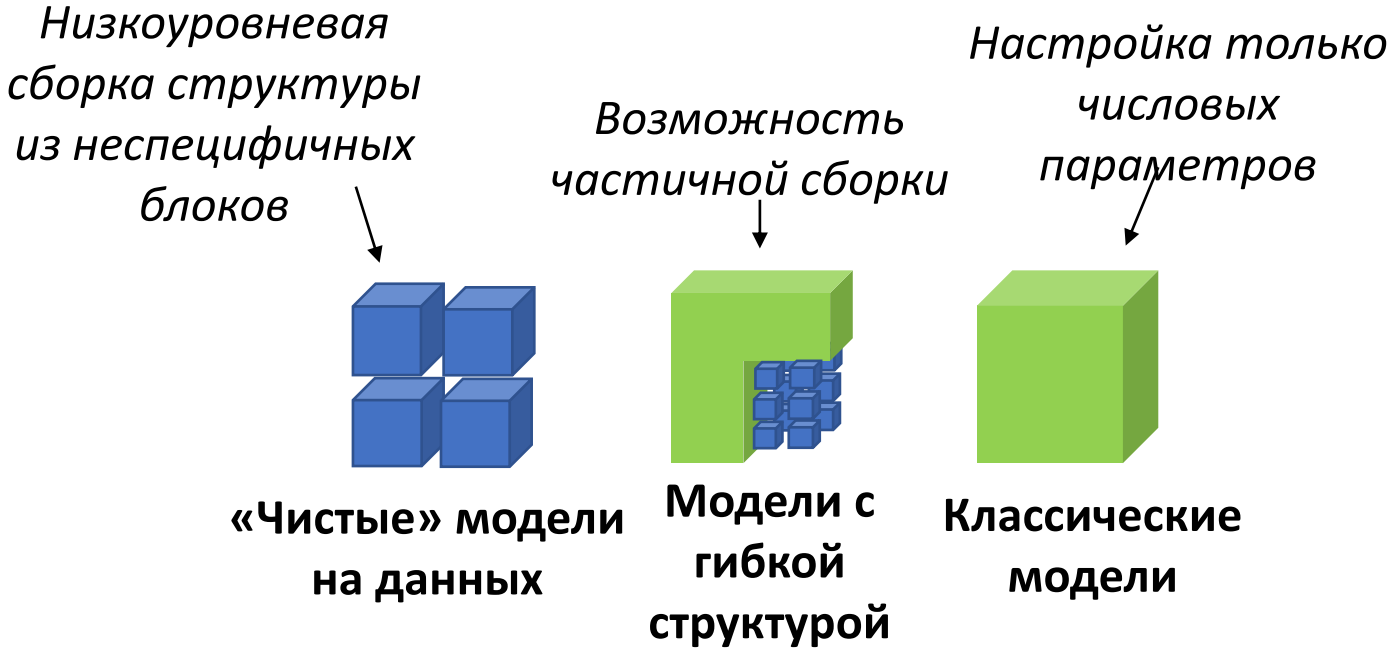


FEDOT в контексте classic ML-pipeline



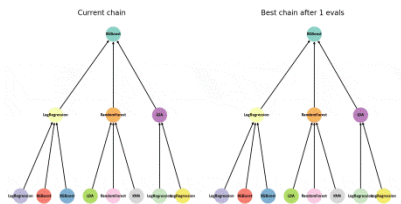
Типы пайплайнов, поддерживаемые FEDOT

Концепция гибкого управления сборкой модели

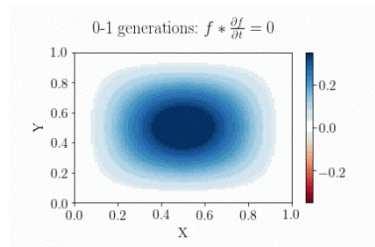


Управлением поиском компромисса между качеством и сложностью модели

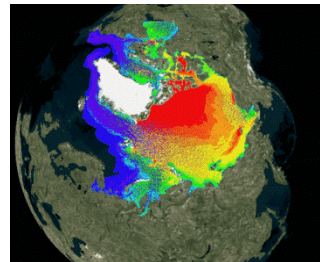
Основные возможности фреймворка:



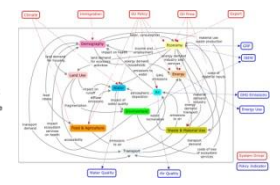
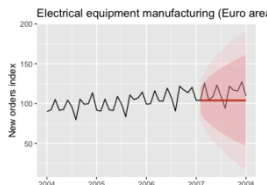
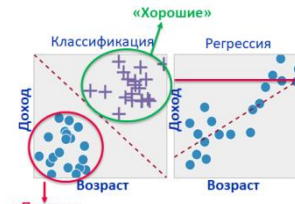
Сложная структура моделей



Построение интерпретируемых моделей

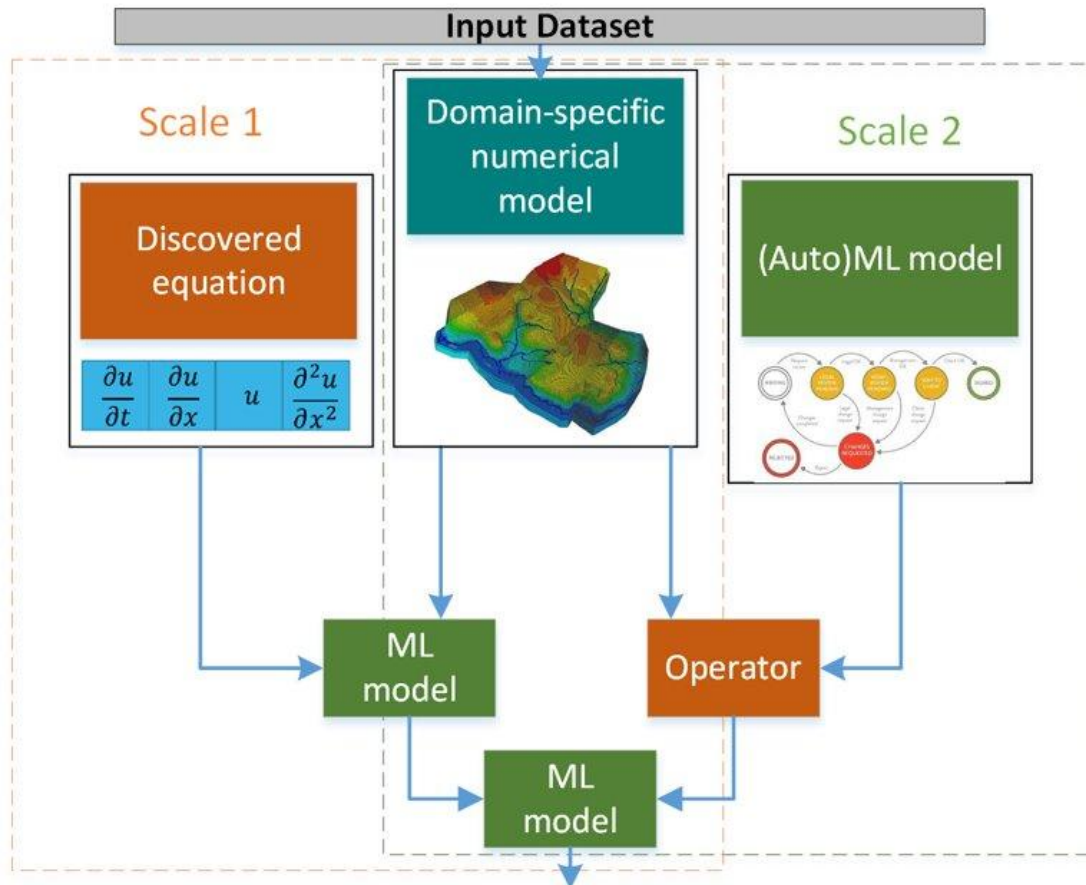


Встраивание предметных моделей

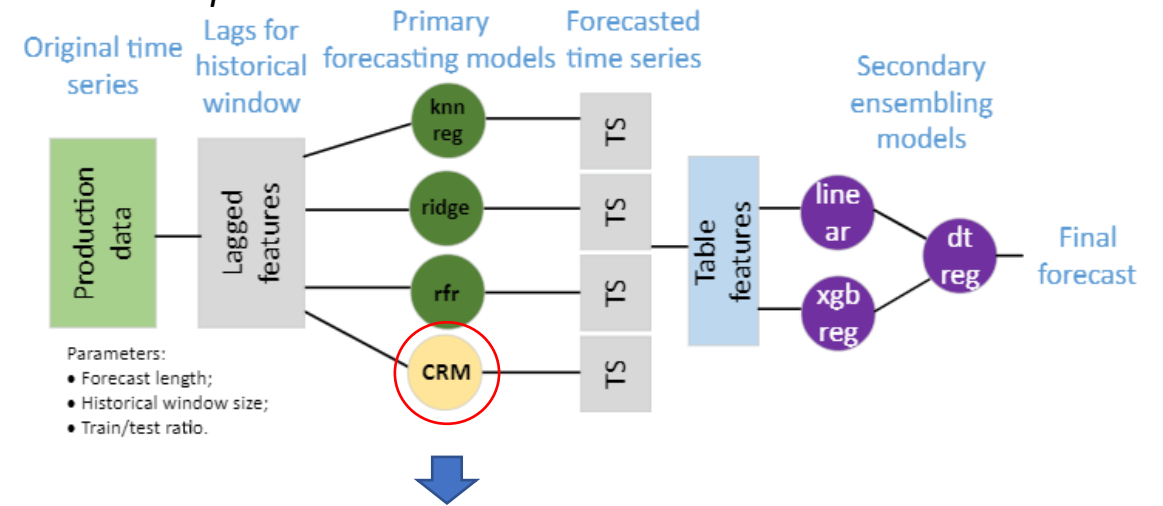


Применимость для различных задач

Можно ли применить ИИ для соединения всего самого лучшего, доставшегося от разных поколений?



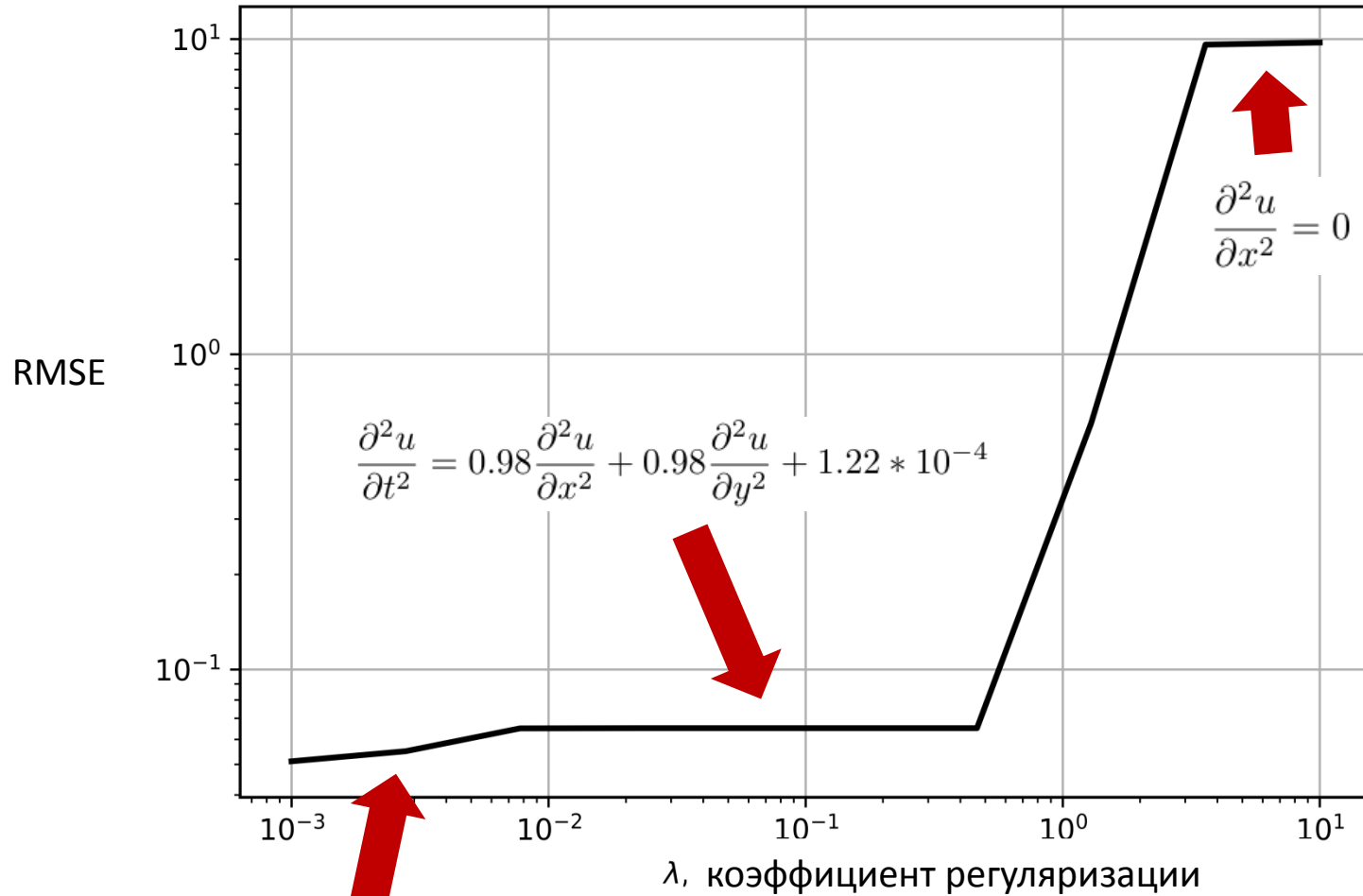
Структура гибридной модели для прогнозирования дебета нефти:



Формулировка CRM модели:

$$\begin{cases} c_t V_p \frac{d\bar{p}}{dt} = I(t) - q(t) & \text{уравнение материального баланса} \\ q(t) = J(\bar{p} - p_w) & \text{уравнение Дюпюи} \end{cases}$$

$$\frac{dq(t)}{dt} + \frac{q(t)}{\tau} = \frac{I(t)}{\tau} - J \frac{dp_w}{dt}, \quad \tau = \frac{c_t V_p}{J}$$



$$\frac{\partial^2 u}{\partial t^2} = 5.173 * 10^{-8} \frac{\partial u}{\partial x} + 0.9751 \frac{\partial^2 u}{\partial x^2} - 3.535 * 10^{-8} \frac{\partial u}{\partial y} + 0.9751 \frac{\partial^2 u}{\partial y^2} + 4.351 * 10^{-10} x + 1.218 * 10^{-4}$$

Пример применения алгоритма к данным, описывающим волновое уравнение:

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$$

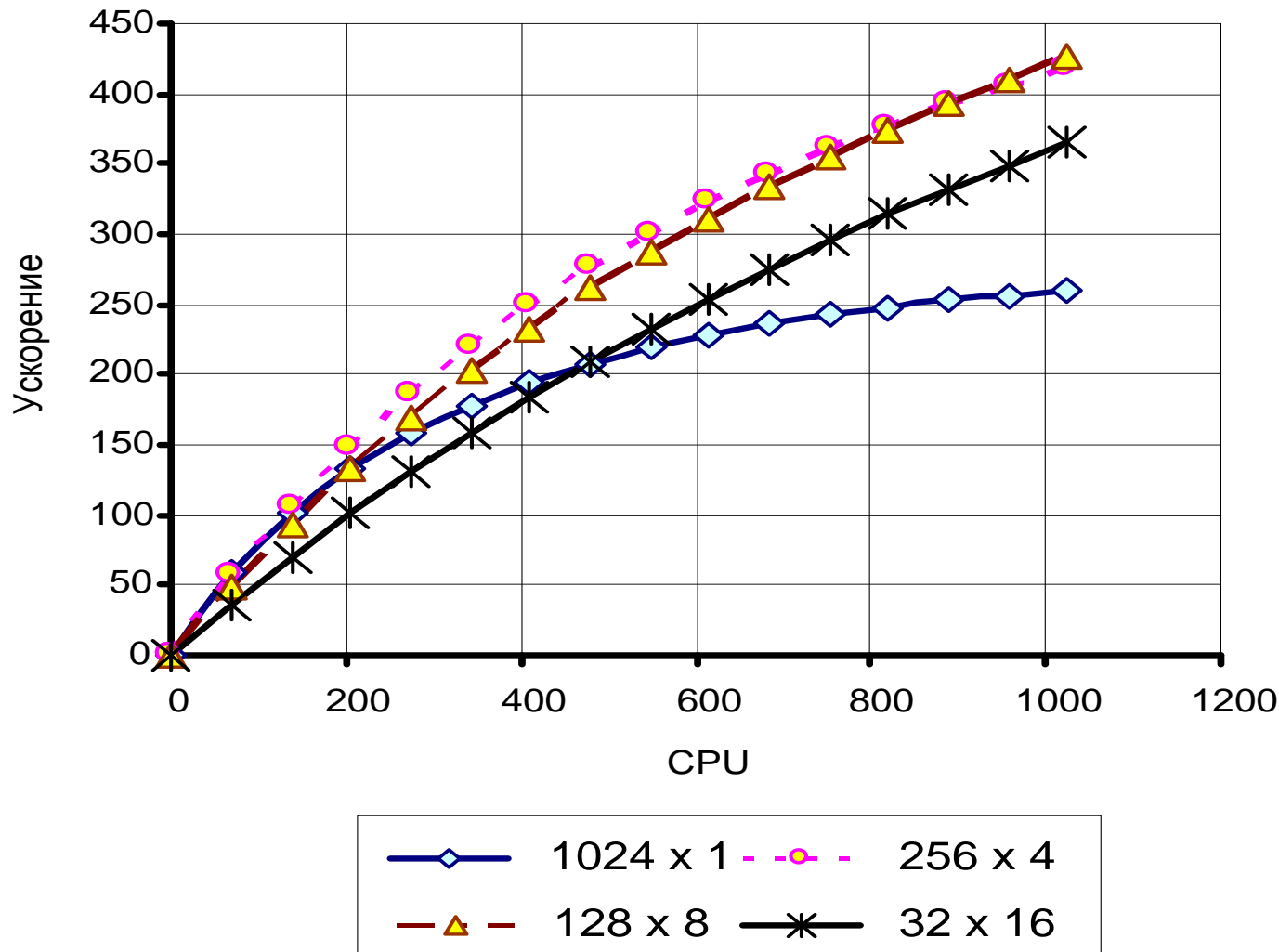
Коэффициент регуляризации, являющийся основным гиперпараметром алгоритма, определяет число слагаемых в полученном уравнении:

- при низких значениях мы получаем переобученную модель, воспроизводящую погрешности в данных: обычно - набор слагаемых из корректного уравнения + лишние;
- оптимальные значения - получаем уравнение, наиболее близкое, к реальности;
- завышенные значения - неполная структура уравнения.

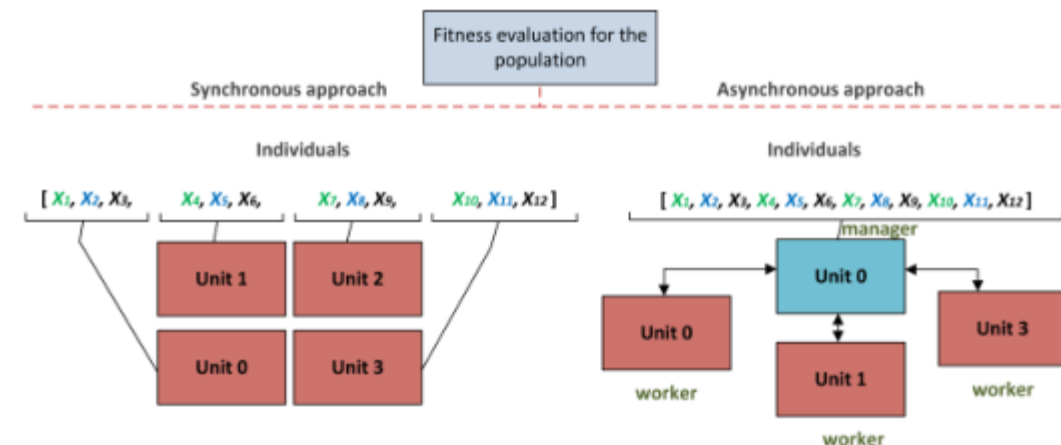


1. Что мы восстанавливаем: само уравнение или его дискретный аналог?
2. Чувствительность к способу численной реализации PDE: не всегда существуют критерии сходимости
3. Порядок производных: одно уравнение или системы нормальных уравнений?
4. Существование решения для всех особей из популяции
5. Единственность решения в динамике (бифуркации, перескоки между решениями)
6. А как быть с шумами и представлением производных?
7. А если мы разберемся с п. (1-6) – сможем ли мы получить количественный ответ **за разумное время?**

Моделирование полей морского волнения в
Каспийском море
5 лет, перекрытие 1 месяц



- 1) Дополнительные ограничения на структуру уравнений (например, запрет на перекрестные связи)
- 2) Варьирование параметрами распараллеливания (игра с моделями производительности)
- 3) «Запрещенный прием» – распараллеливание на уровне расчета целевой функции (функции приспособленности):

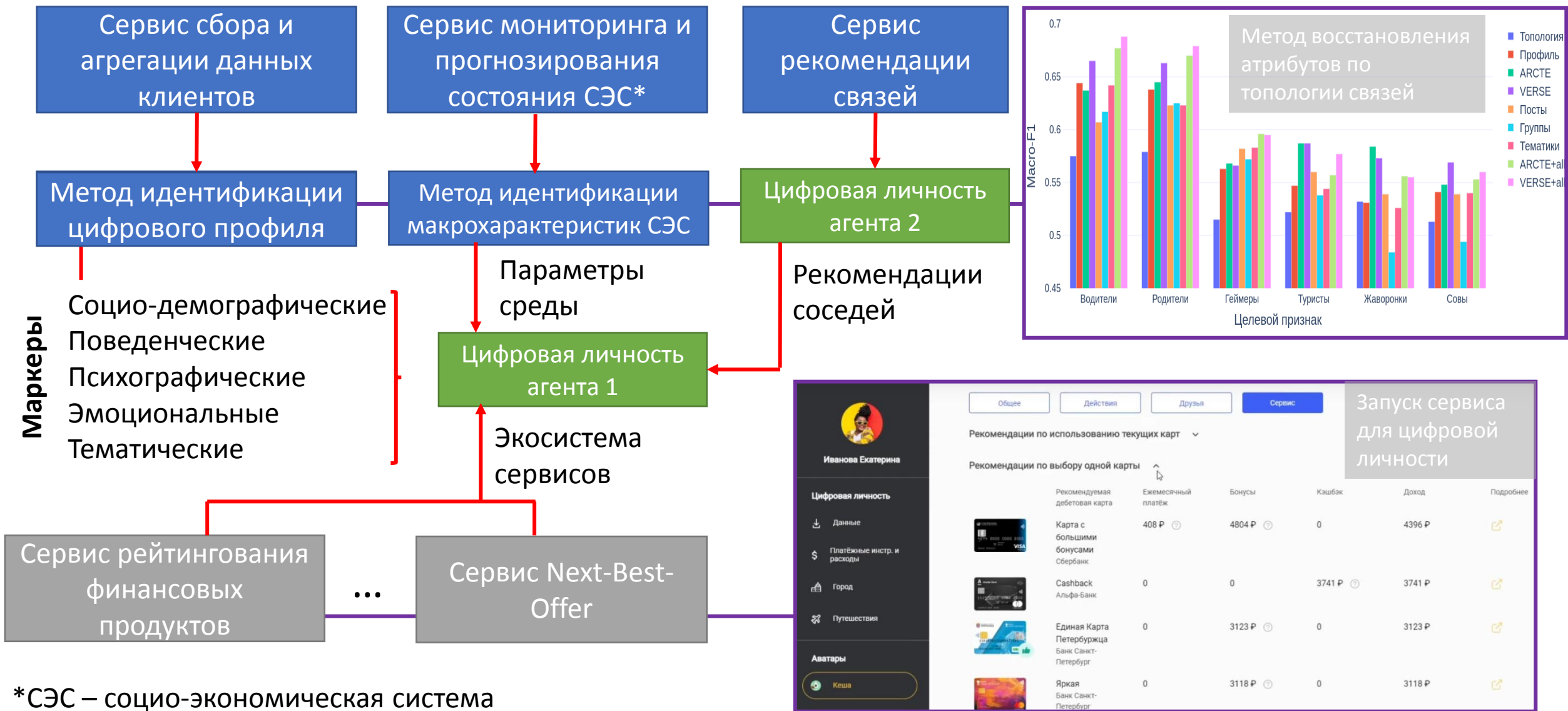




ТРАНСФОРМАЦИЯ МЕХАНИЗМОВ ОРГАНИЗАЦИИ ЭКСПЕРИМЕНТА

Экосистема цифровых личностей: аллегория



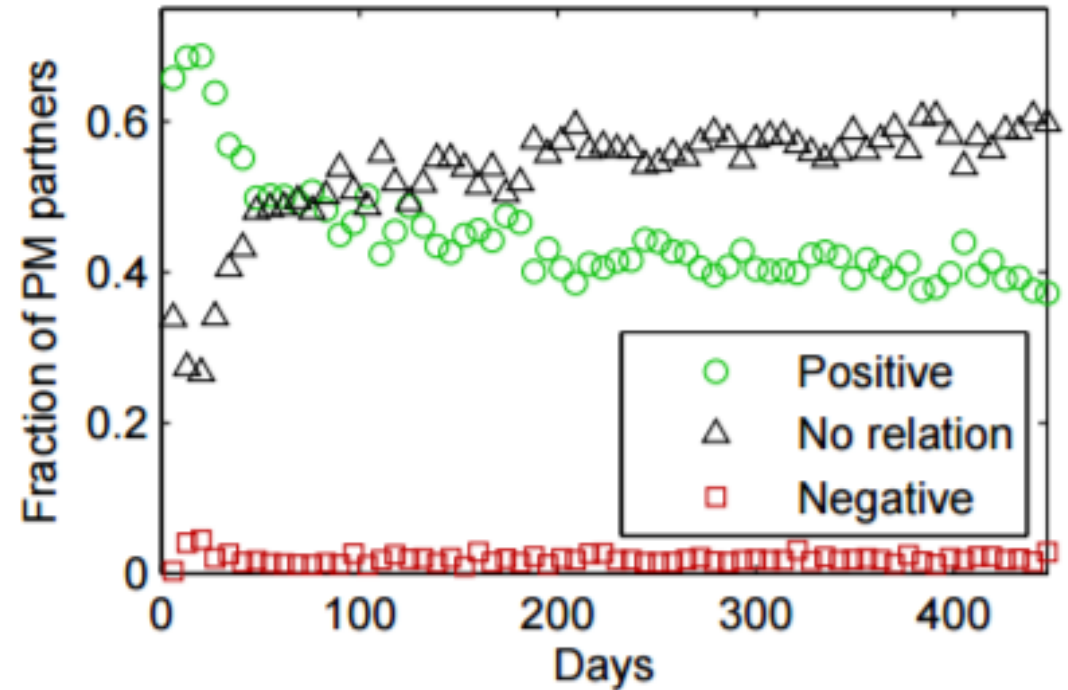
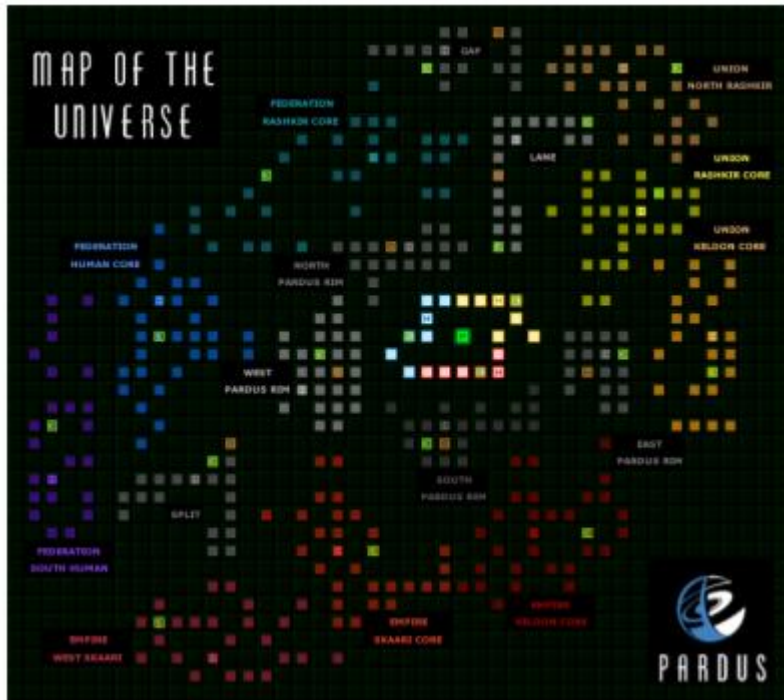


Изучение эффектов реального мира на игровых сообществах

Measuring social dynamics in a massive multiplayer online game.

Michael Szella , Stefan Thurner (2009)

Pardus (<http://www.pardus.at>), браузерный MMOG: 2005-2008



Все негодяи живут мало и дружить не умеют. У хороших людей это получается лучше 😊

Цифровой Аватар -- это программа-помощник на персональном устройстве пользователя, являющаяся его делегатом в публичном кибер-пространстве.

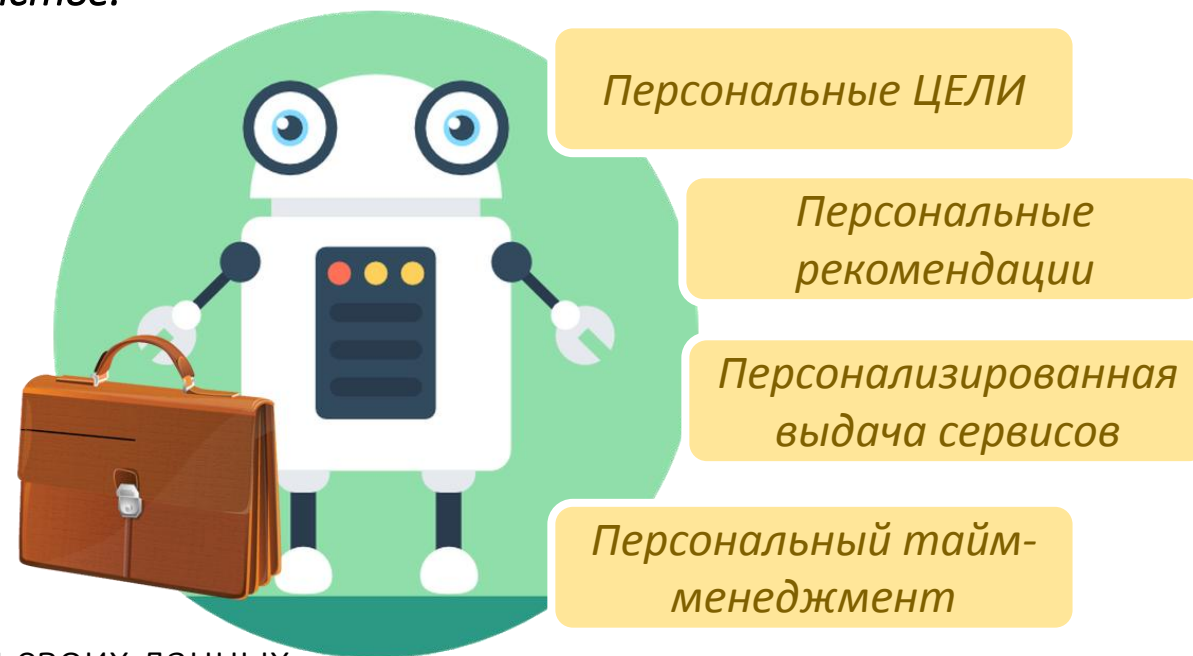
- Рефлексирует опыт пользователя
- Получает непрерывный фидбэк от хозяина
- Развивается вместе со своим хозяином
- Эффективно управляет временем
- Оценивает прогресс владельца в заданной системе ценностей
- Персонализированный подход к коммуникации, поддерживает взаимодействие на естественном языке

Владелец:

- разрешает Аватару пользоваться определенными видами своих данных
- формирует целеполагание и определяет ценности (стратегии действий аватара)

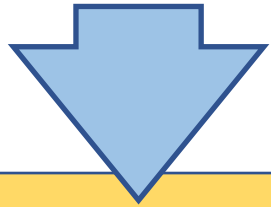
Аватар:

- помогает владельцу ориентироваться в потоках информации
- берёт на себя рутинные операции с этими потоками
- помогает принять решение в условиях неопределённости или недостатка времени



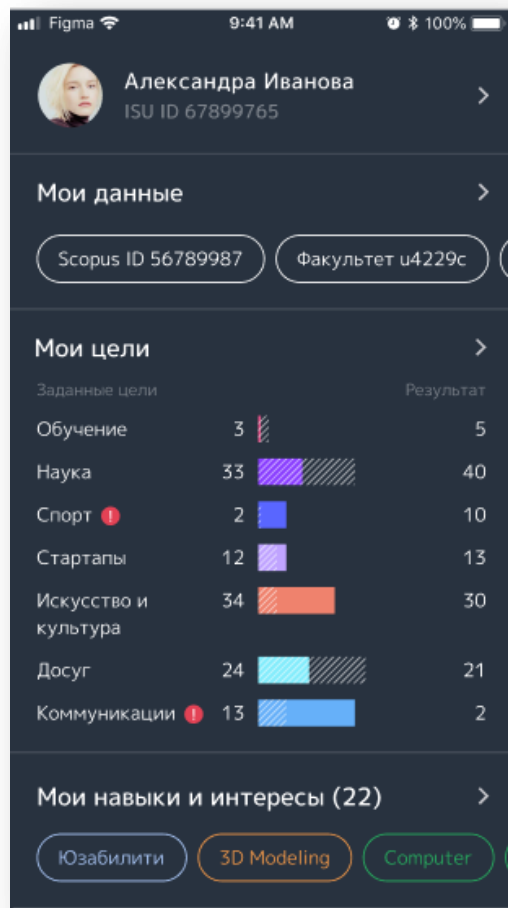
Экосистема Аватаров Университета ИТМО

- Упрощение взаимодействия студент-преподаватель
- Агрегирование и обобщение потребностей преподавателей, управление процессами
- Автоматизация администрат. деятельности

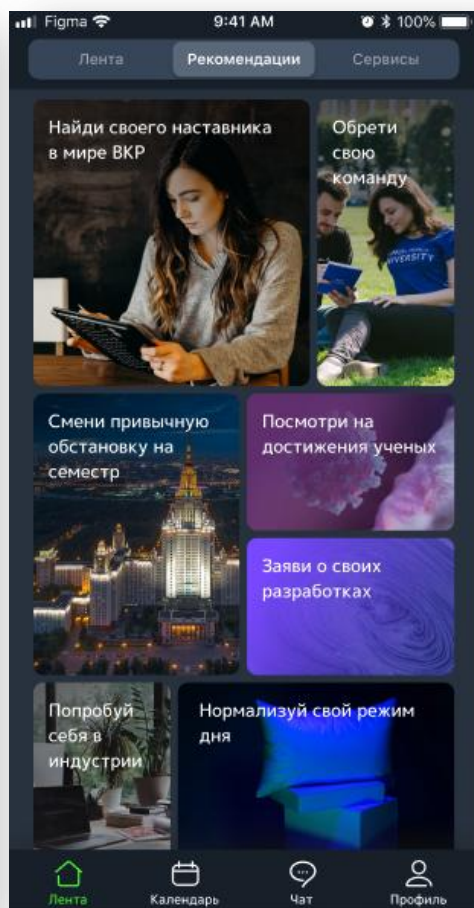


- 10 Увеличение эффективности сотрудников
- 10 Увеличение вовлеченности студентов
- 10 Персональная траектория развития в рамках ценностей университета

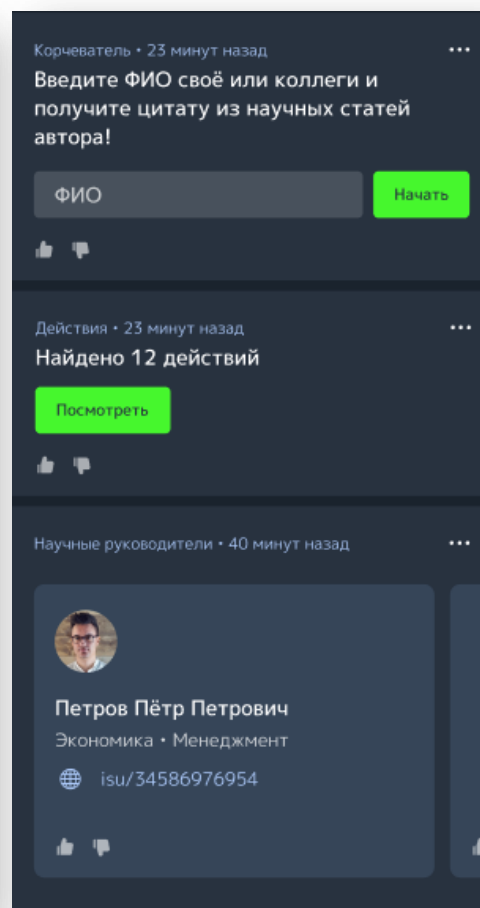




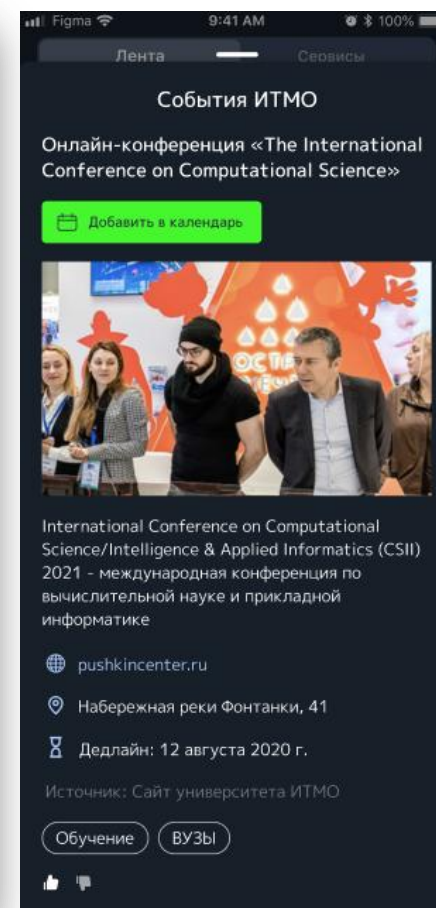
Задать цели и
отследить
движение к ним



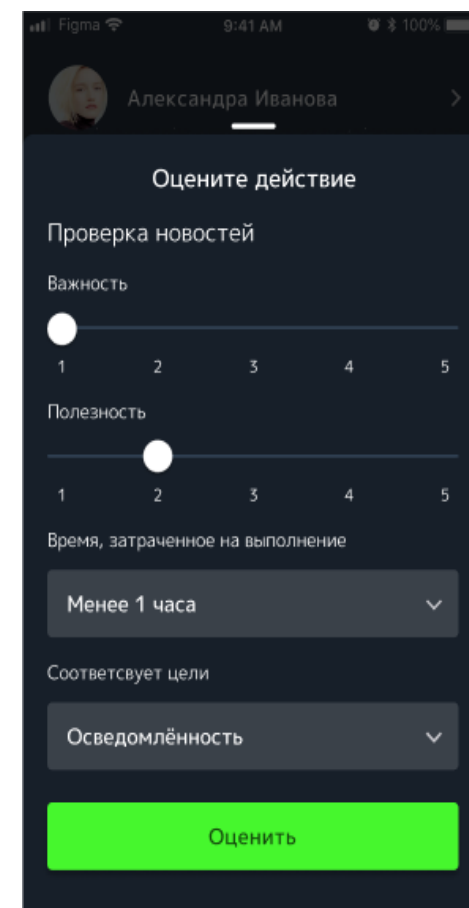
Выбрать текущий
фокус развития



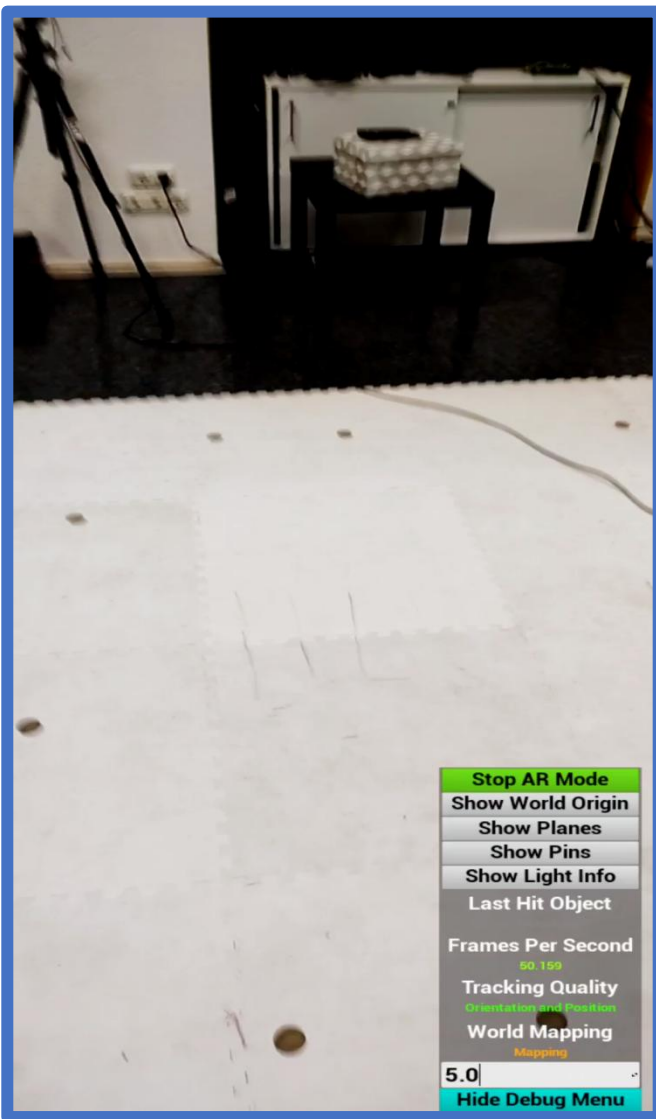
Получить
рекомендации от
сервисов



Заполнить карту
событий



Оценить
полезность
действий



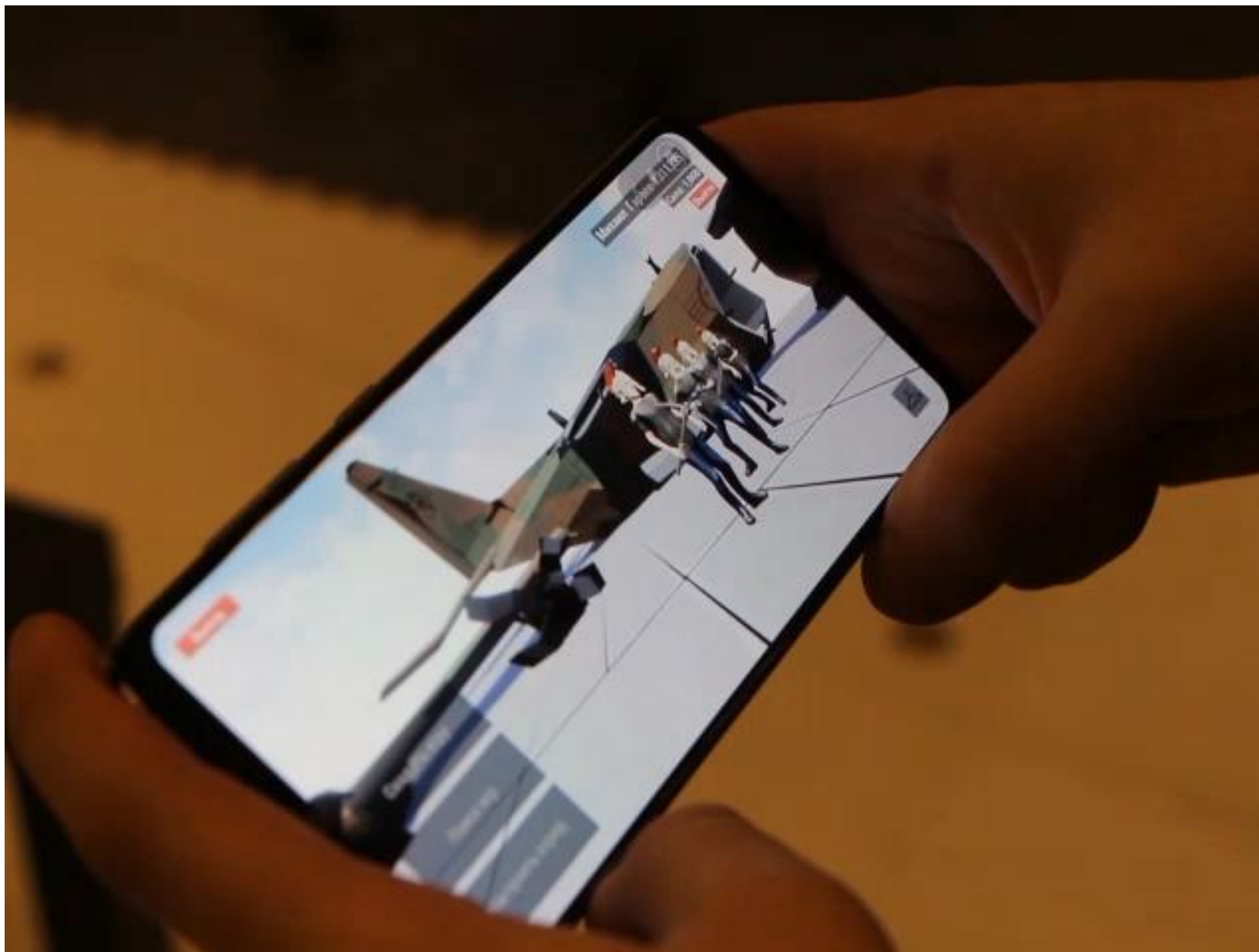
Отображение качества работы студента с аватаром в его игровые характеристики.

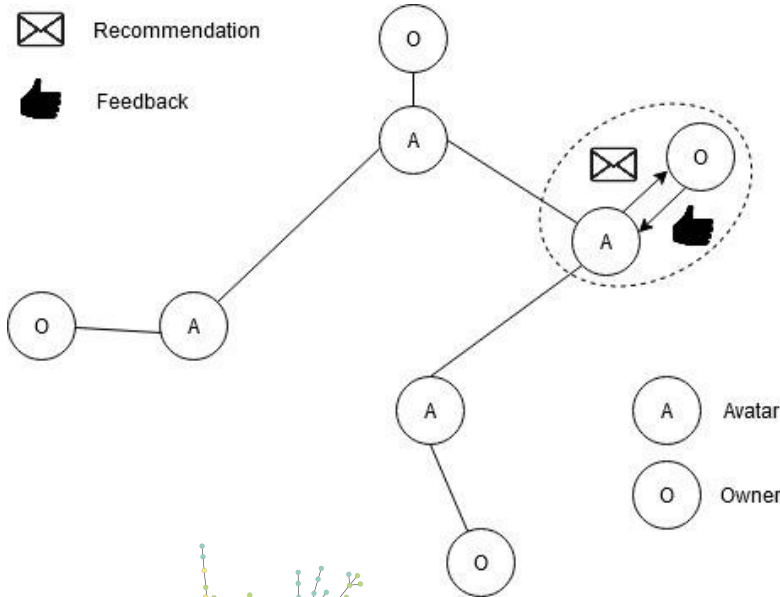
- успешный студент будет иметь более усовершенствованного персонажа в игровом мире
- Положительное подкрепление и мотивация к дальнейшему развитию.

Аватар транслирует фактические данные о студенте в игровые характеристики и систему гибкой двусторонней мотивации, прививая студентам на примере положительных и отрицательных сценариев ценности важные для университета.

Unique Points:

- **Единое виртуальное пространство** для множества пользователей;
- Масштабирование идей под различные бизнес-задачи (например совещание с презентациями);
- **Множество игровых сценариев** – от головоломок до командных турниров.
- **Персональные игровые аватары** – отсканированные пользователи;
- Игра как внутри персонального девайса так и в дополненной реальности;
- **Взаимодействие с реальным пространством** и объектами окружения.





Аватар выполняет рекомендации на основании баланса между мнениями владельца и соседей

$$Q_i(a) = (1 - \alpha) \cdot Q_i^{neighb}(a) + \alpha \cdot Q_i(a)$$

α – коэффициент смешивания

Оценивается:

- средняя удовлетворенность владельца по t итерациям
- скорость достижения порогового значения удовлетворенности

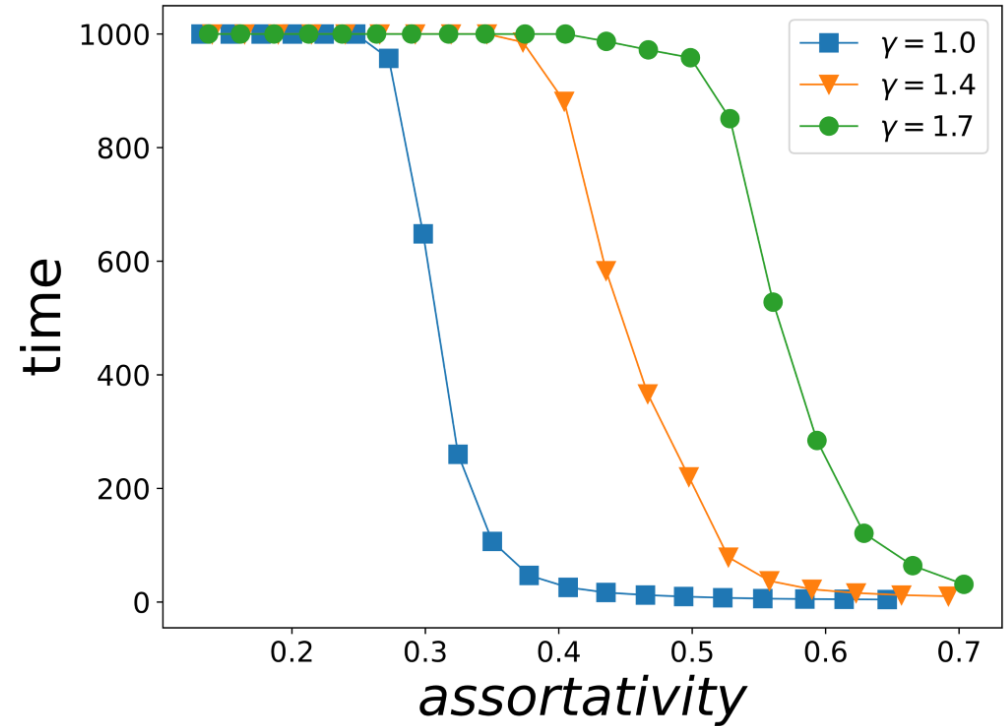
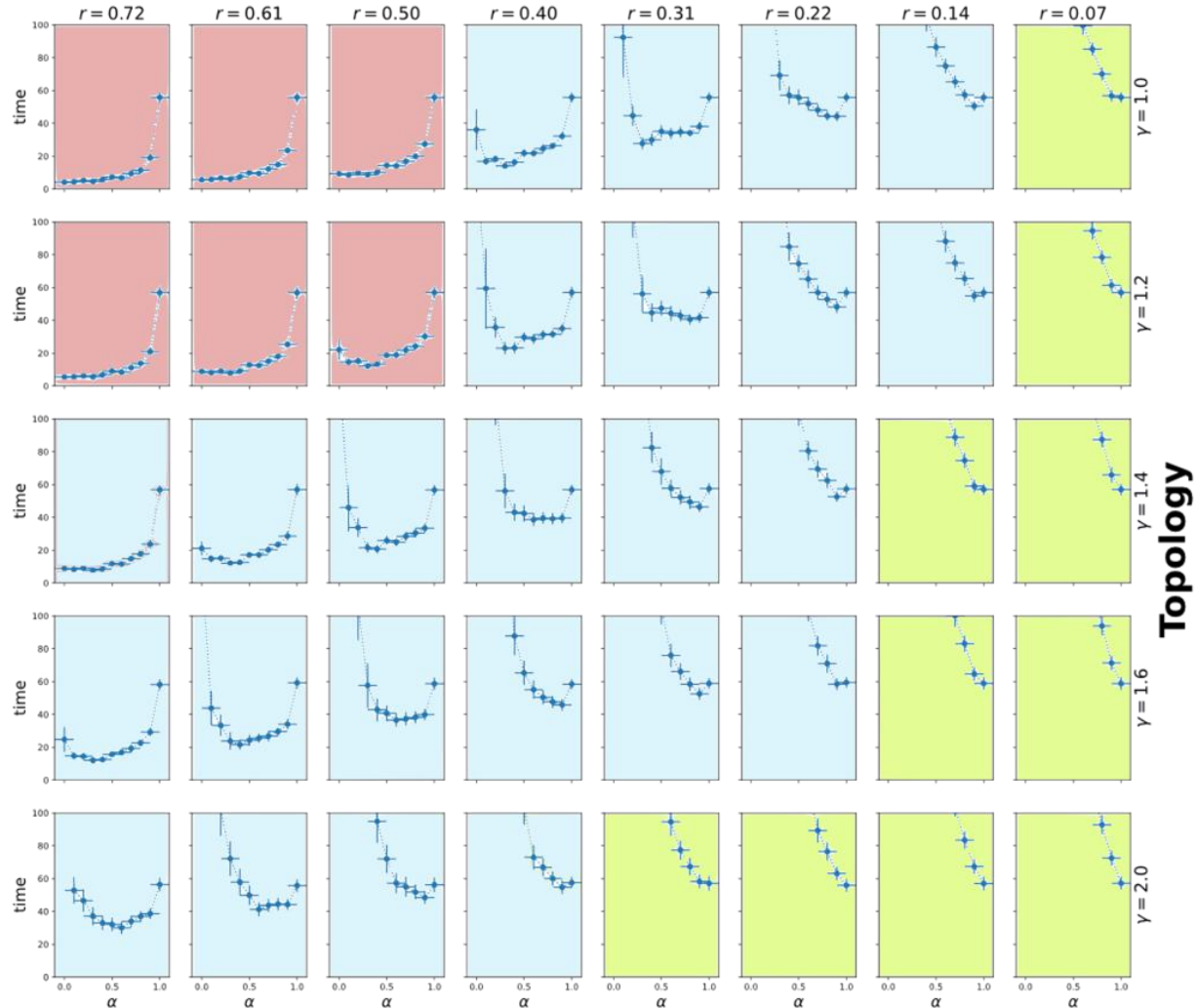
assortativity = 0.35
 $\gamma = 2.5$

assortativity = 0.84
 $\gamma = 1.5$

Для разных сетей и распределений атрибутов узлов варьируется скорость сходимости для фиксированного α

Три режима: (1) выгодно использовать информацию только от соседей; (2) выгодно смешивать информацию владельца и соседей; (3) выгодно использовать информацию только от владельца

Assortativity



В ситуации абсолютного доверия соседям существует критическое значение (дис)ассортативности, после которого резко замедляется сходимость



ВМЕСТО ЗАКЛЮЧЕНИЯ



Цифровая трансформация – не панацея, а хайп. Потому eScience «полетит» не здесь, и не сейчас. Однако ученым-предметникам от этого будет жить ощутимо легче... и веселее 😊