ПОЛИТЕХ

ИНСТИТУТ КОМПЬЮТЕРНЫХ НАУК И ТЕХНОЛОГИЙ

ВШ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА

курс: Введение в профессиональную деятельность

ЛЕКЦИЯ 12 : ОТ ВЫЧИСЛЕНИЯ ЧИСЕЛ К ВЫЧИСЛЕНИЮ СЛОВ

27.04.2023

«2023 РЕВОЛЮЦИОННЫЙ ПЕРЕДЕЛ» КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ: ОТ ПРОГРАММНОГО АВТОМАТА К ТРАНСФОРМЕРУ-ИНТЕРПРЕТАТОРУ

ПОЛИТЕХ

«В начале было Слово»
Евангелия от Иоанна

вычисления **чисел** → вычисление **слов и смыслов**

эра
«интеллектуальных»
вычислительных платформ
«трансформер-интерпретатор»

эра
электронных автоматов,
вычисляющих числа с помощью
программ-алгоритмов

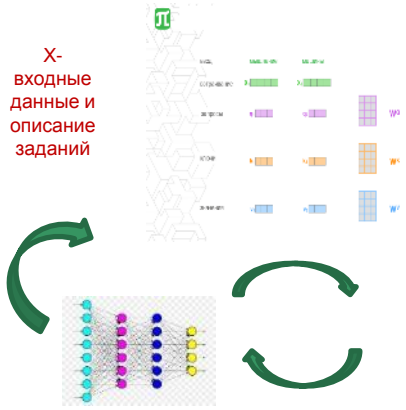
Эра
механических автоматов,
исполняющих **один алгоритм, вычисления**

Алгоритм записанный на естественном языке, понятном человеку

Алгоритм вычисления записанный **человеком** на языке «понятным» компьютерам

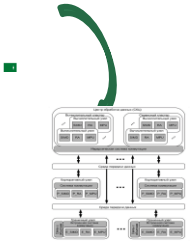
«Все есть число»
Пифагор
570-490 дон.э.

X- входные данные и описание заданий



описание процессов на «языке данных»

y –выходные данные - результаты



Описание процессов на «языке алгоритмов»

ПОЛИТЕХ

Построение языковой модели реальности

- Компьютер «понимает» только числа. Чтобы обработать с помощью компьютера текст вместо чисел, нужна модель, которая «разделяет» исходный текст на токены («символы») и кодирует их числами.
- Токеном в принципе может быть либо буква, слог или целое слово (обычно используют самые часто встречаемые слова, буквы и «значимые» слоги).
- То как закодированы токены текста, называют «эмбеддингом». Это название происходит от to embed — вставлять, встраивать, потому что токен как бы «укладывают» обрабатываемый текст в числовое пространство.
- Для получения «эмбеддинга» в настоящее время испольщуются специальные алгоритмы GLoVe, ELMo или word2vec, которые эффективно реализуются с помощью специальных SIMD процессоров, получивших «метафорическое» название GPU

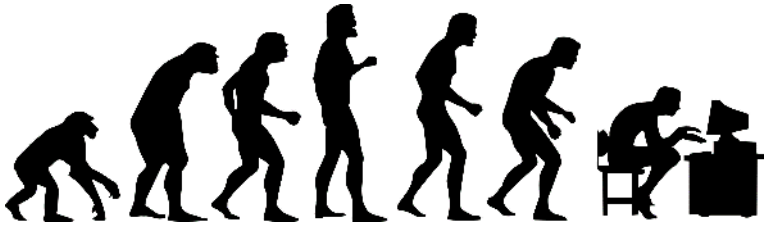
Другими словами, можно сказать, что эмбеддинг токена —это числовое обозначение для слова, слога или буквы.

Проблемы цифровой трансформации: Pro/Contra

Pro: *Само-улучшающийся ИИ спровоцирует взрыв технологий*

Ирвинг Гуд, 1956

Contra:

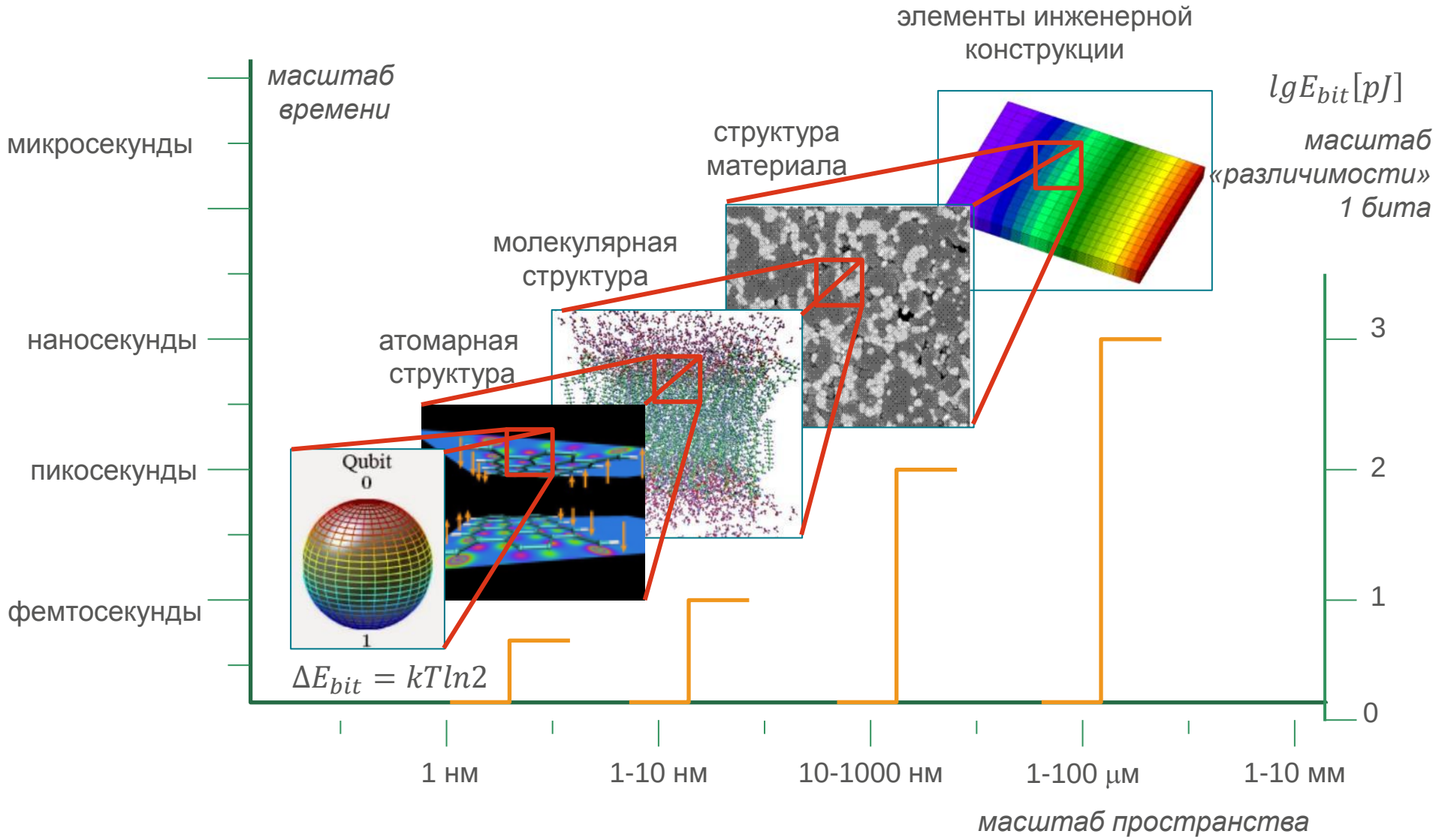


Угроза эволюция к «кликовому мышлению»

Фундаментальная проблема «цифровой трансформации»:

множество "содержательных" истин (признаков) всегда превосходит по объему множество истин (признаков), доказуемых с помощью любой формализованной (компьютерной) системы.

Содержательные «смыслы» физической реальности: различные формальные пространственно-временные сущности



ПОЛИТЕХ

Подходы к решению проблемы вычисления смысла текста

Суть проблемы: переход от алгоритмических (вычисление элементов числового поля) к когнитивным вычислениям (слов как элементов когнитивного пространства содержательных понятий).

Метод решения: представление свойств когнитивного пространства понятий как топологического многообразия (пространства), которое локально гомеоморфно p-адическому координатному пространству с заданной на нем дополнительной структурой – контекстными связями выбранной базы топологии.

Формализация: Построение гомологий и инвариантов топологического пространства понятий, с помощью которых описывается «суррогатная» модель смысла обрабатываемого текста

Технология: имитация когнитивных процессов человека с помощью цифровых конечных автоматов путем их целевой реконфигурации для генерации содержательной интерпретации результатов обработки различных данных, включая числа, таблицы, изображения и тексты, а также обработки ответной реакции и оценки со стороны человека.

История вопроса: значение слов и понимание смысла как мотивации к действию

- «Если значения слов не определены, то нет и смыслов. Если нет смыслов, то действия не происходят».

  (Конфуций).

- «Определите значения слов, и вы избавите человечество от половины его заблуждений» (ошибочных действий).

  (Рене Декарт).

Содержательные проблемы современных Супер КТ: Проблема №1

«Углеродный след» компьютерных технологий :

год	число ядер	$R_{реак}$, ПФлопс	R_{max}, ПФлопс	эл. мощность, МВт
2022 Frontier	8,700,XXX	1680.XX	1100.XX	21
2020 Fugaku	7,300,XXX	513.XX	415.XX	28
2010 Tianhe-1	186,XXX	4.7X	2.6X	4
2000 ASCI Intel	9,6XX	0.03	0.02	-

- 1 кг угля -> 3 кВтч =0.003 МВтч
- 1 тонна угля -> 3 МВтч

21 МВт -> 21/3 =
7 тонн угля в час
168 тонн угля в день
60480 тонн угля в год

Причины:

- *Большая потребляемая мощность*
- *Сложность УПРАВЛЕНИЯ ПРОЦЕССАМИ планирования вычислений*
- *Низкая масштабируемость вычислительных ресурсов при решении прикладных задач*

Содержательные проблемы современных СКТ: Проблема №2

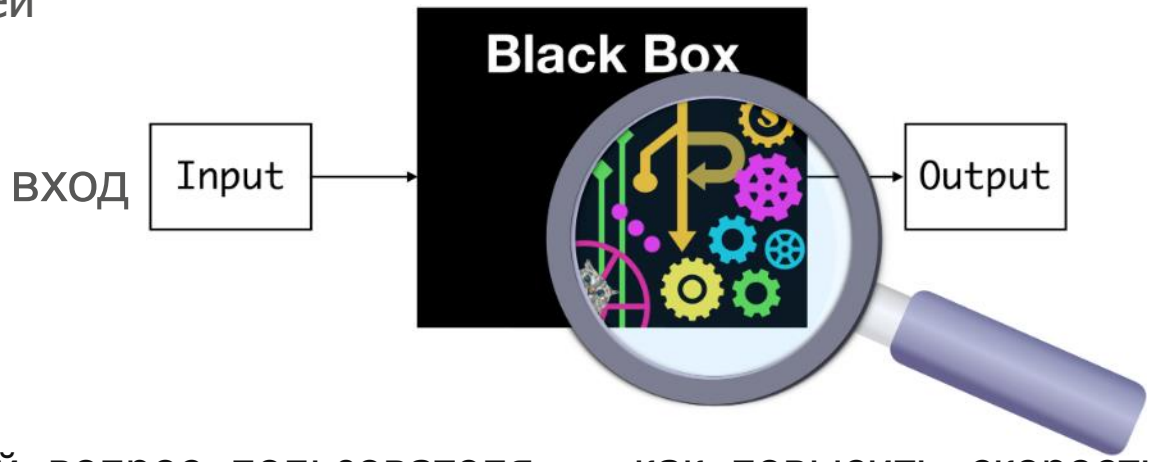
«Сложность использования СКТ»

Для пользователей почти всегда СК «черный ящик», а для «СК» – задачи пользователей - «каждый раз совершенно «новые».

Задачи пользователей

выход

Результаты вычислений



Типичный вопрос пользователя - как повысить скорость решения своих прикладных задач?

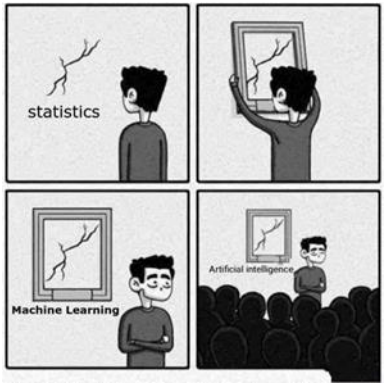
Решение проблемы: сделать так, чтобы на вопрос пользователей мог ответить «сам СК» ?

Содержательные проблемы современных СКТ Проблема №3:

«Вычисленные результаты надо объяснить»

- **Решение – от ИИ к экзо-интеллекту:** гибридной мкльтимодальной системы, интегрирующей возможности естественного и искусственного интеллекта

Скрытность смыслов



Ограниченность знаний

Мультимодальность решений

Дескрипторы «похожести»:

числовое равенство /с точностью до количеств 1+2=3

эквивалентность /с точностью до значения 1.0000=1

гомеоморфность /с точностью до инвариантов классов

Постановка задачи повышения эффективности вычислений интеллектуализации СКЦ «Политехнический»

- На основе гибридных платформ организации вычислений и машинного обучения повысить эффективность (сократить времени обработки и уменьшить ошибки, …рост производительности) процессов решения прикладных задач с использованием СК за счет:
  - Использования «умных» систем планирования заданий
  - Накопления «опыта»» решения различных классов задач
  - Построения «суррогатных» моделей пользователей
  - Интерпретации результатов вычислений и формирование рекомендаций пользователям по использованию СК ресурсов

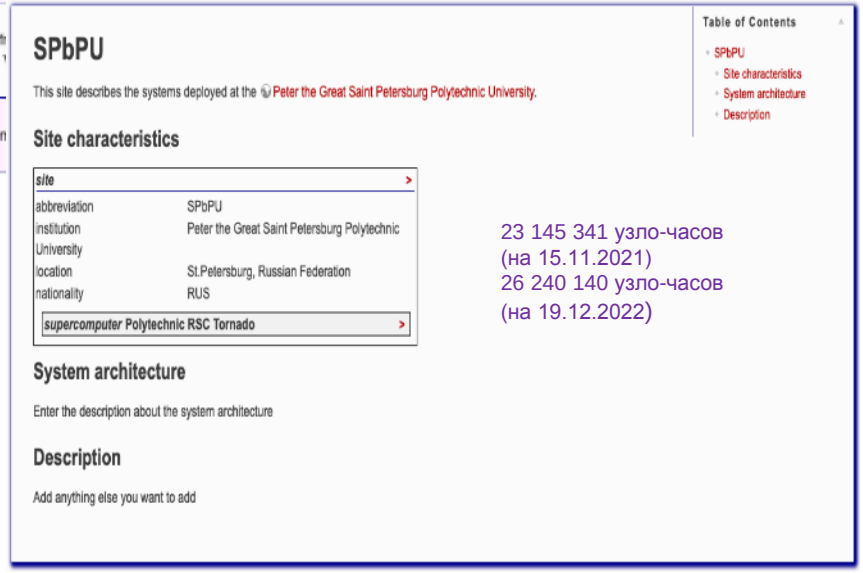
СКЦ «Политехнический» в цифрах и показателях

#	information								io500		
	list id	institution	system	storage vendor	filesystem type	client nodes	client total procs	data	score	bw GiB/s	md kIOP/s
21	isc20		Officinalis	Red Hat, Intel, QCT	CephFS	8	256	zip	66.88	28.58	156.48
22	isc20	SPbPU	Polytechnic RSC Tornado	RSC Group	Lustre	59	944	zip	64.29	21.56	191.73
23	sc19	DDN	AI400	DDN	Lustre	10	240	zip	63.88	19.65	207.63
24	isc20	Red Hat	EC2-10xl3en.metal	Red Hat	CephFS	10	320	zip	57.17	26.29	124.30
25	sc19	Google Cloud	EXA5-GCP-PD-STD	Google Cloud	Lustre	200	1600	zip	52.96	17.31	162.06
26	sc19	Janelia Research Campus, HHMI	Weka	WekaIO	wekaio	18	1368	zip	48.75	26.22	90.62
27	sc19	Oracle Cloud Infrastructure	Oracle Cloud Infrastructure with Block Volume Service running Spectrum Scale	Inf							
28	sc19	Penguin Computing Benchmarking and Innovation Lab	Penguin-ASG-NVBeeOne	Con							

Потребляемая мощность 1 МВТ

В рейтинге организаций Минобрнауки СКЦ «Политехнический» самый производительный гетерогенный кластер, пиковая производительность 3 Флопс, 26448 ядер CPU 445440 ядер GPU

за 2022 г: решено 1 970 442 задач



SPbPU

This site describes the systems deployed at the Peter the Great Saint Petersburg Polytechnic University.

Table of Contents
- SPbPU
  - Site characteristics
  - System architecture
  - Description

Site characteristics

site

abbreviation	SPbPU
institution	Peter the Great Saint Petersburg Polytechnic University
location	St.Petersburg, Russian Federation
nationality	RUS

supercomputer Polytechnic RSC Tornado

System architecture

Enter the description about the system architecture

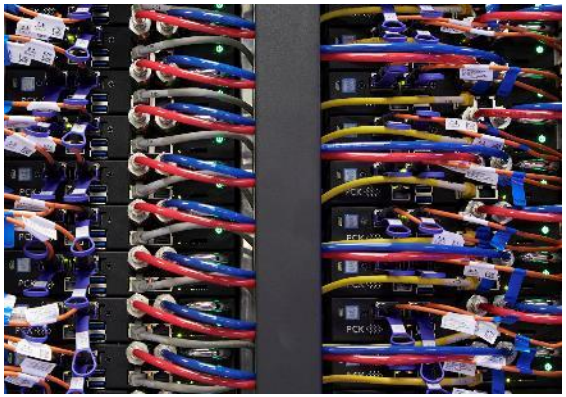
Description

Add anything else you want to add

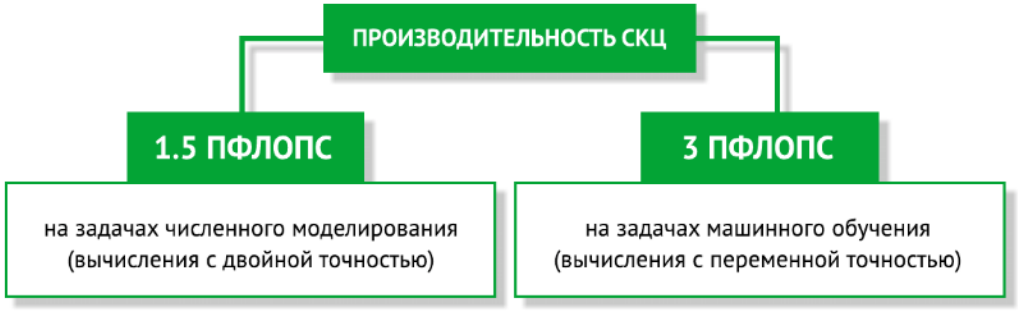
23 145 341 узло-часов (на 15.11.2021)
26 240 140 узло-часов (на 19.12.2022)

СКЦ «Политехнический» в цифрах

1. Первый по производительности гибридный (CPU + GPU + FPGA) суперкомпьютер в России среди организаций, подведомственных Министерству науки и высшего образования (согласно рейтингу top50.supercomputers.ru);
2. Девятый по производительности суперкомпьютер в России (согласно рейтингу top50.supercomputers.ru);



- Более 1000 пользователей
- Более 100 научных групп
- Более 30 промышленных организаций
- > 25 миллионов узло-часов за 5 лет работы
- > 2 млн выполненных расчетных задач



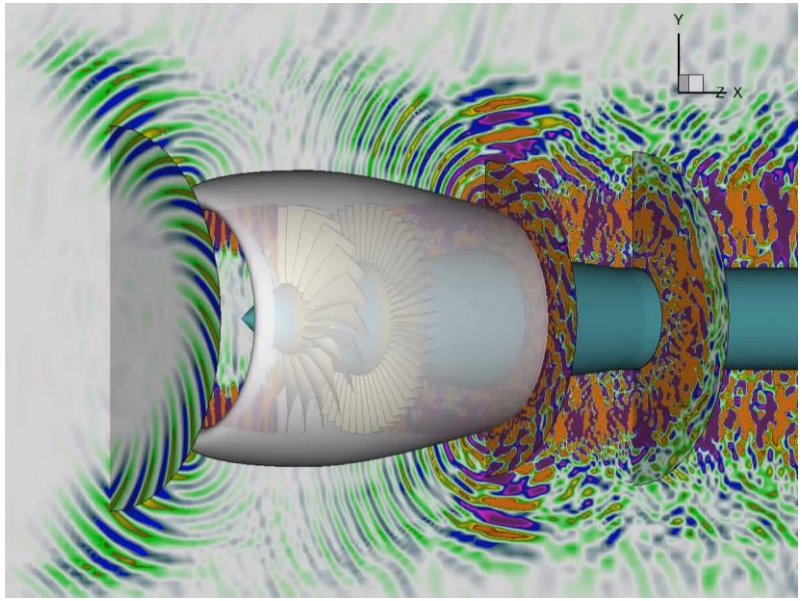
Понять нельзя вычислить: как правильно расставить запятые

прямой расчет и визуализация акустических характеристик турбореактивного двигателя

Физическая модель турбореактивного авиационного двигателя

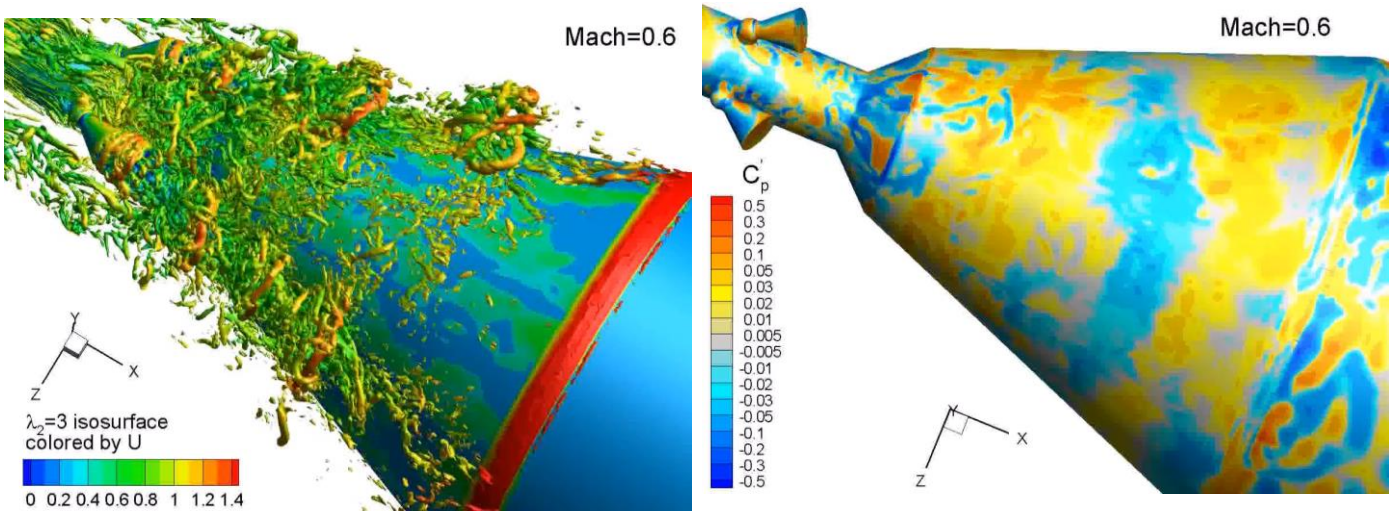
- Прямая задача: основанная на первых (физических) принципах аэродинамики модель
- Обратная задача: суррогатная модель визуализации протекающих процессов

Анимация – вариант суррогатной модели визуализации результатов расчета звуковых волн, излучаемых двигателем



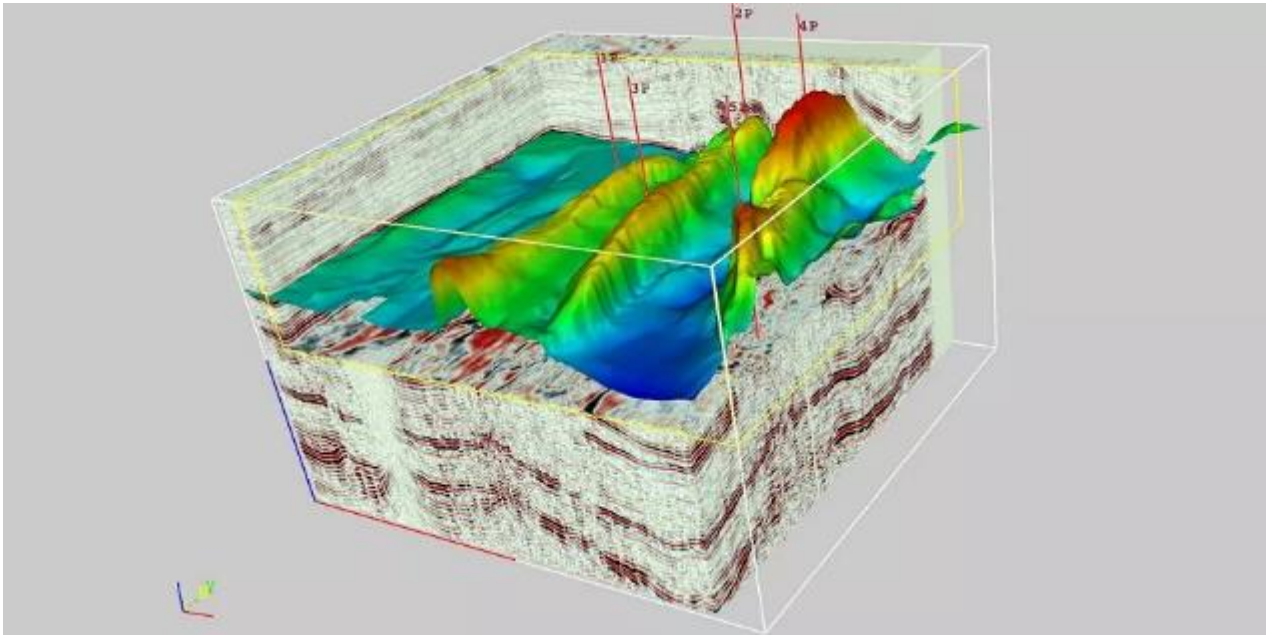
ПОЛИТЕХ

Пример использования мультимодаьных данных: решение «обратной задачи» инженерного «понимания»



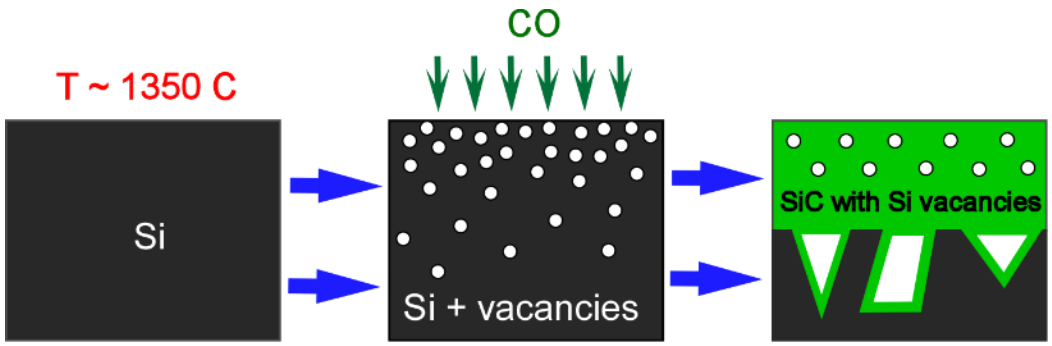
Анимация вихревых структур и пульсаций давления на поверхности возвращаемого космического аппарата в процессе выведения

Пример использования «больших» данных сейсморазведки

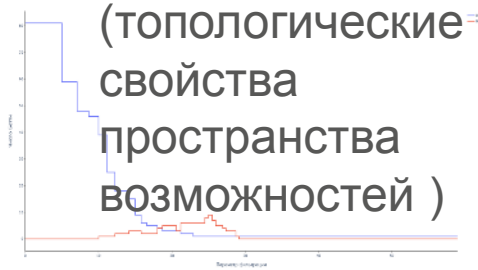


Что это дает: на практике : подавление комбинаторного взрыва - переход от анализа сейсмограмм «как искусства» к технологиям анализа на основе машинного обучения с использованием данных о структуре строения геологической среды и проектирования «виртуальных» скважин, на основе прогнозирования физических свойств и литологии горных пород.

Применение методов МО для задач синтеза новых материалов (Пример создания ЭКБ нового поколения на основе SiC)

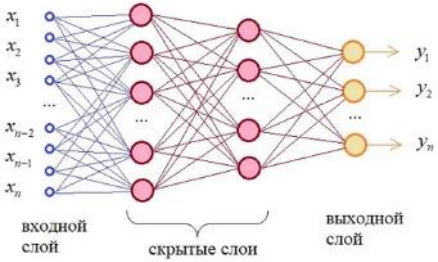


Кривая Бетти (топологические свойства пространства возможностей)



Накопление экспериментальных данных, и применение **топологических инвариантов** анализа данных для формирования «обучающей выборки» трансформера сетки МКЭ описание физических свойств образования образования кремниевых вакансий.

1. ОБУЧАЮЩАЯ ВЫБОРКА С ИЗВЕСТНЫМ СОСТОЯНИЕМ
2. ВХОДНЫЕ ДАННЫЙ, КОТОРЫЕ ГЕНЕРИРУЮТ НОВЫЕ ВАРИАНТЫ SIC



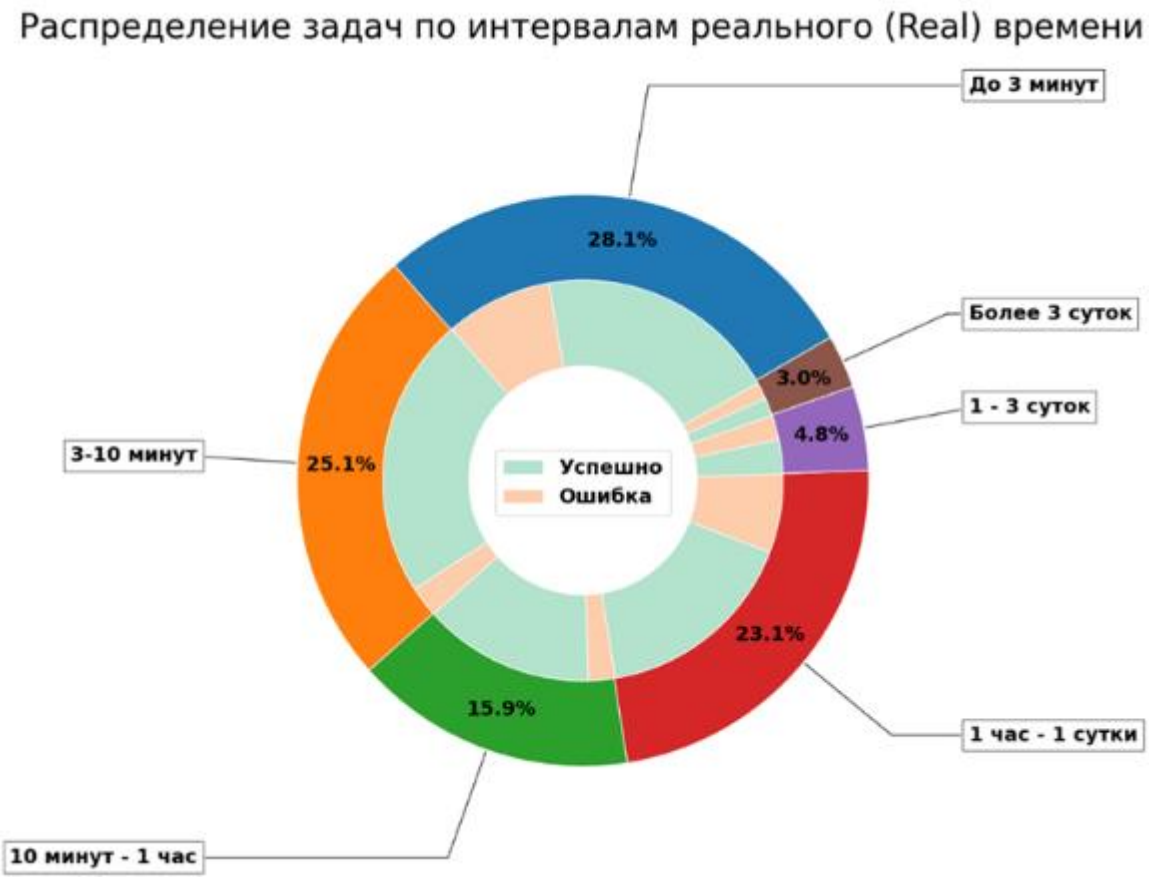
Результаты:

- ВАРИАНТЫ СТРУКТУРЫ СЛОЯ SIC С РАЗНОЙ КОНЦЕНТРАЦИЕЙ ВАКАНСИЙ
- «ТРАНСФОРМЕР»: ОБЪЯСНЕНИЯ ПОЛУЧЕННЫХ СВОЙСТВ (ВЫЧИСЛЕНИЕ КОЭФФИЦУИЕНТОВ ВЛИЯНИЯ РАЗЛИЧНЫХ ФАКТОРОВ НА КОНЦЕНТРАЦИЮ ВАКАНСИЙ)

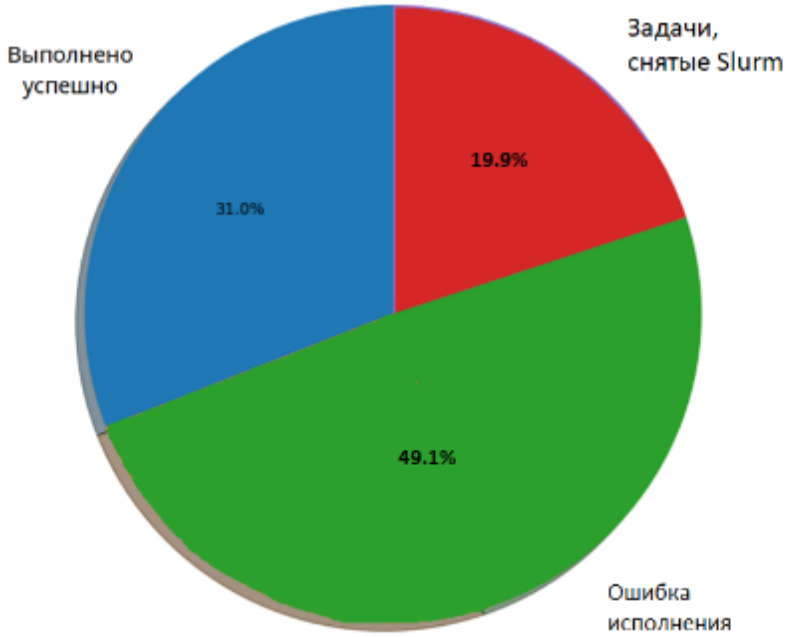
✓ результат «МАШИННОГО ОБУЧЕНИЯ» СК: УПРАВЛЯЕМЫЙ подбор параметров химической реакции взаимодействия, CO и Si (карбид кремния) которая приводит к образованию кремниевых вакансий в слове SiC 100-1000 нм и концентрации вакансий 10^{19}-10^{21} $см^{-3}$

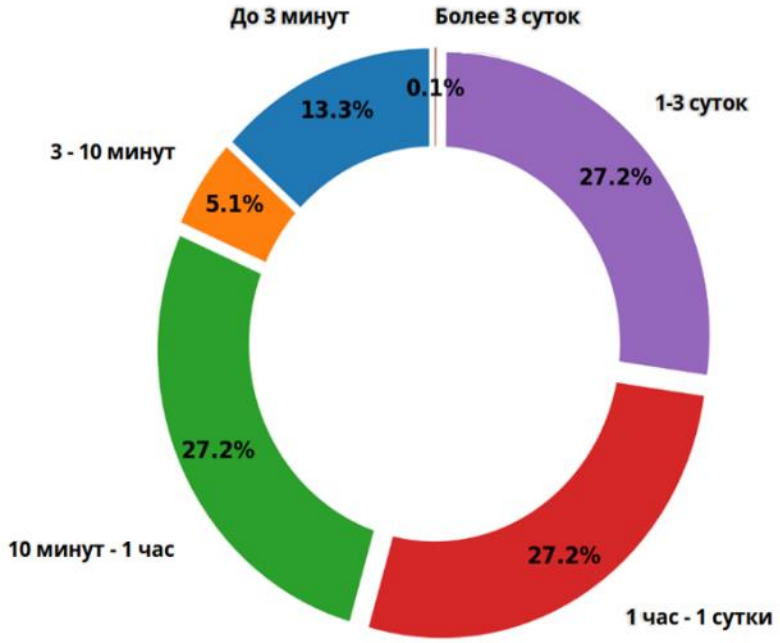
ПОЛИТЕХ

Временная сложность заданий пользователей



Проблема сложности «в цифрах»: статистика «выживания» заданий пользователей в среде СК

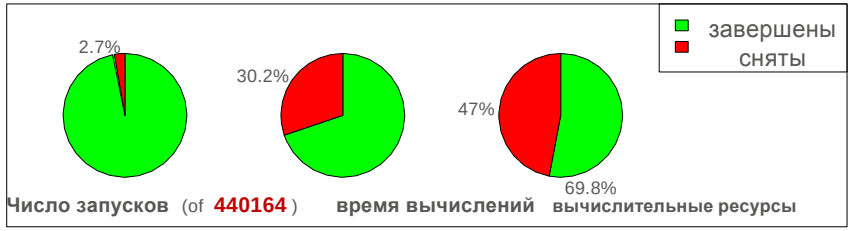




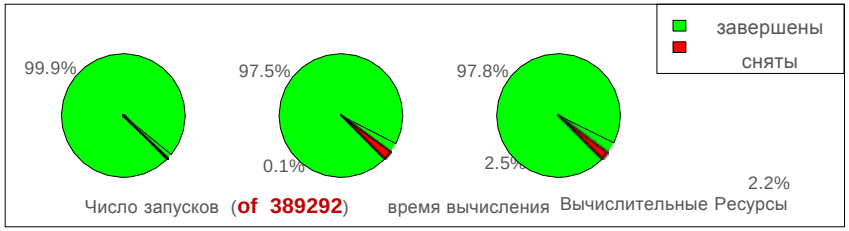
Проблема: Успешно выполненные задачи составляют **около 1/3** от общего числа заявок пользователей

Эффективность автоматизации «политехнических» вычислений

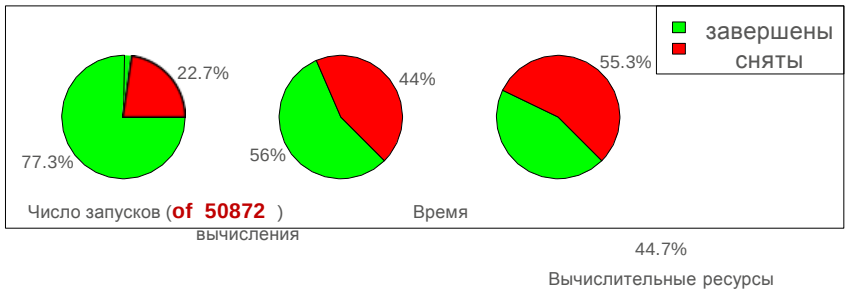
Общая **характеристика эффективности для всех видов запусков заданий**



Класс эквивалентности 1: Автоматический запуск известных заранее заданий



Класс эквивалентности 2: Запуск задания «вручном" режиме

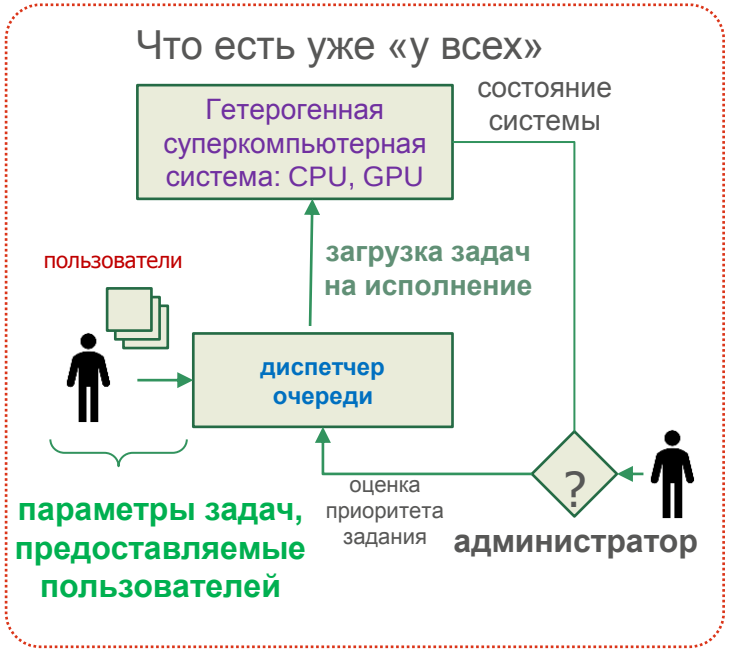


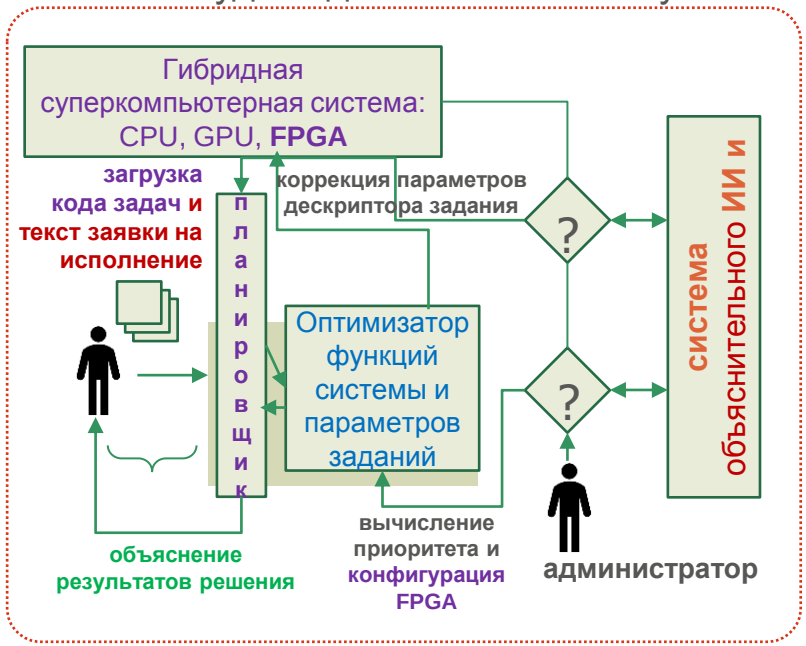
Вывод:

Если параметры заданий пользователей, хорошо (точно) известны, то настройки процессов вычислений можно автоматизировать и более эффективно использовать имеющие компьютерные ресурсы

Вопрос: «Кто виноват и Что делать» ?! : пространство возможностей для «машинного обучения» СК

Что будет сделано в 2023 г. «у нас»



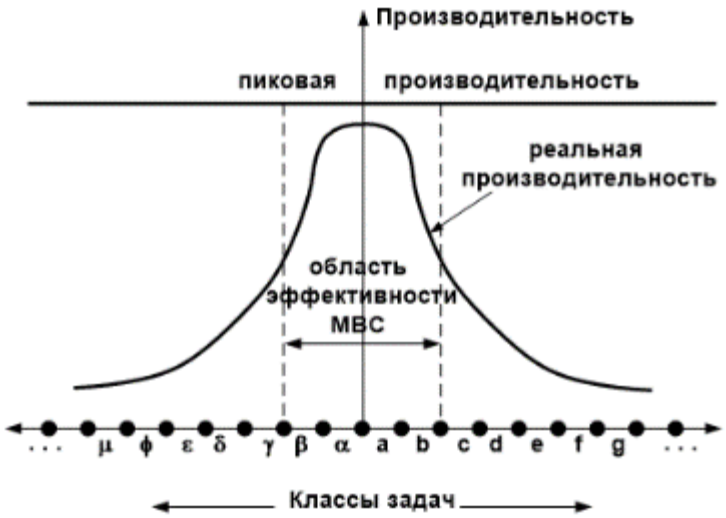


Проблемы:

- неточность оценок параметров заданий, которые формирует пользователей
- «ошибки» диспетчера в оценке времени исполнения заданий

Задачи: Создание системы ИИ, которая способна не только к «машинному обучению» планировщика заданий, но и «объяснению» результатов выполнения заданий, включая формирования сообщений планировщику заданий, администратору и пользователю

Возможность 1: «ручное» управление архитектурой вычислительной платформы …



- Цель: **реконфигурация аппаратного обеспечения ВС для** увеличения реальной производительности на узком классе прикладных задач.
- Задачи:
  - Использовать новые методы и метрики
  - Агрегировать опыт решения задач
  - Придания СК функции авто-рефлексии

Реализация версии «Машины Гёделя» - вычислителя, который знает как ***переписывать*** часть своего кода и реконфигурировать аппаратуру, если он находит доказательства того, что такие изменения позволяют повысить реальную производительность вычислений

Возможность 2: Адаптация гибридной СК платформы «под задачу»

Less Moore, More Brain

Меньше Мура, Больше Мозга

«Умнее» (управление архитектурой) , а не «Толще» (больше ядер и выше частота)

- «Умные» вычислительные платформы, должны уметь «вычислять» смысли решаемой задачи и адаптировать свою архитектуру к особенностям приложений (астрофизика, генетика, материаловедение)

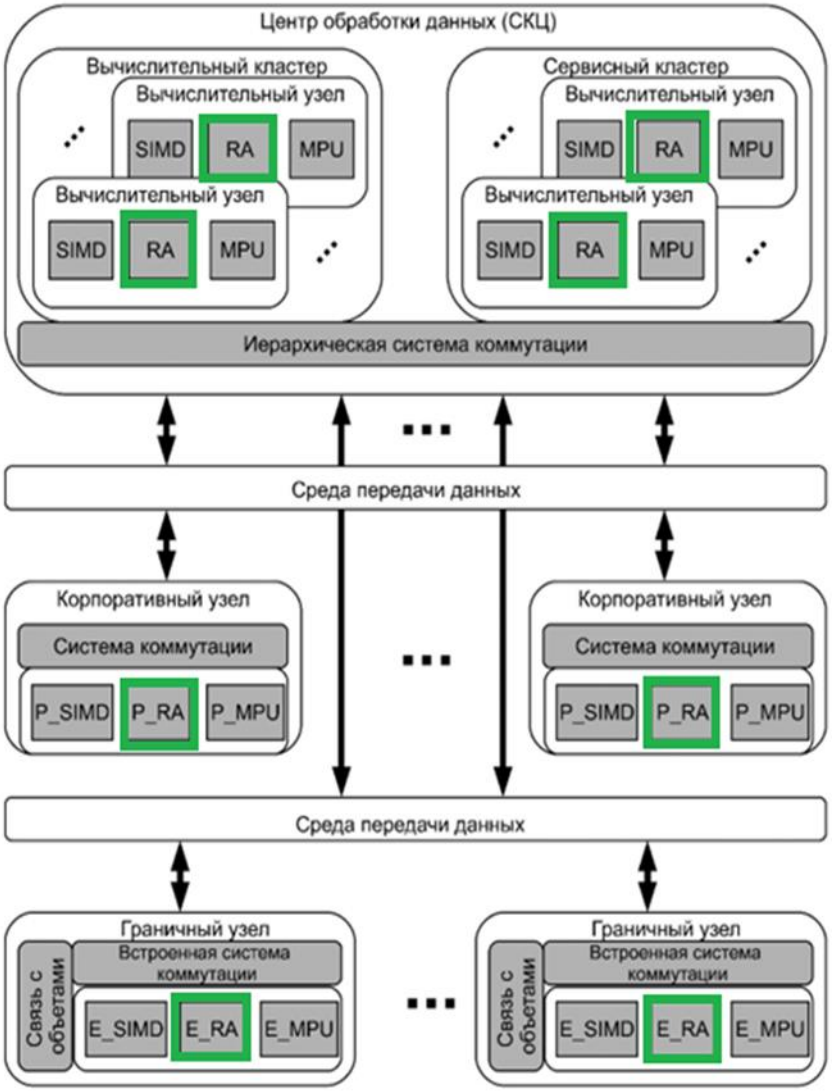
  - Управление на аппаратном уровне осуществляется
    - Сейчас : выбор «лучшей» из «возможных» (заранее собранных) настройки алгоритма решения прикладной задачи режим – «Слабый ИИ»
    - В будущем: Адаптивная реконфигурация «под алгоритм» архитектуры и программы – «Сильный экзоИИ»

Возможность 3: новая архитектура «интеллектуального» СКЦ

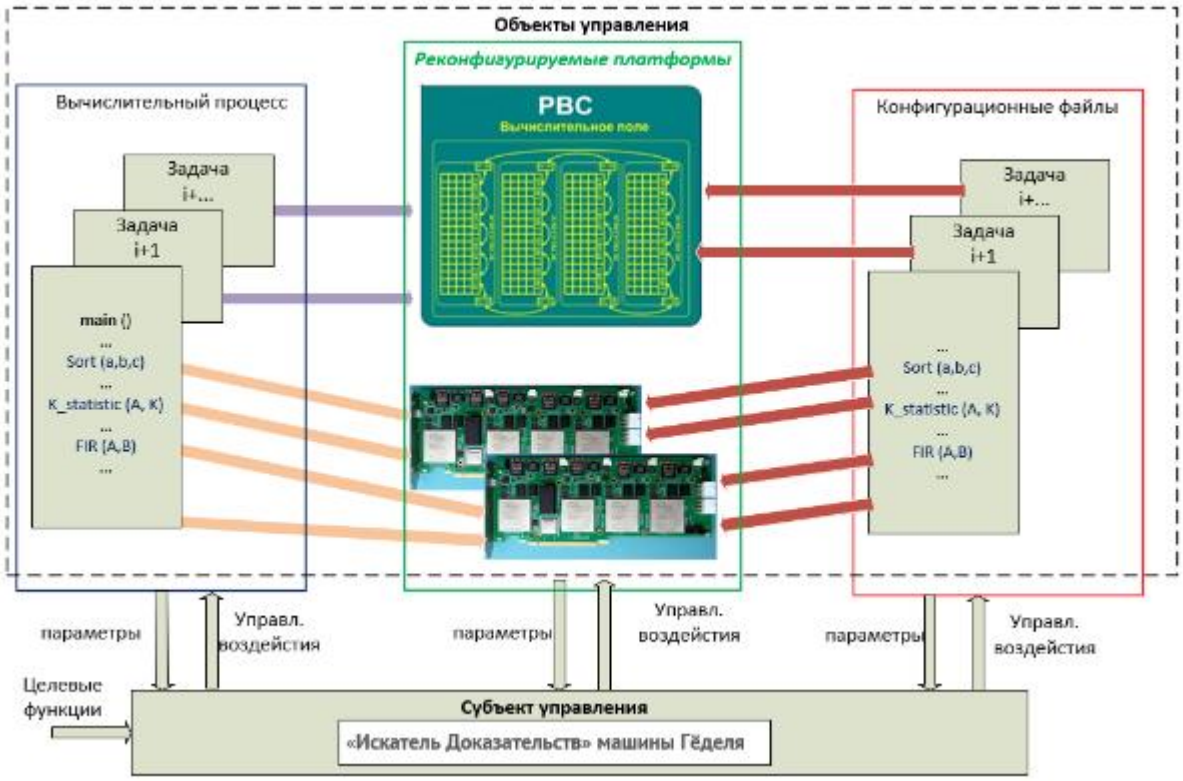
Уровень «объяснения» и доказательства правильности результатов моделирования
Энерго-вычислительная эффективность >4 Гфлопс/Вт

Уровень «агрегации» и машинного обучения путем разделения задач классы эквивалентности и фактор-множества вычислительной сложности
Энерго-вычислительная эффективность >10 Гфлопс/Вт

Уровень CPU/GPUгенерации (вычисления) «данных»
Энерго-вычислительная эффективность >20 Гфлопс/Вт



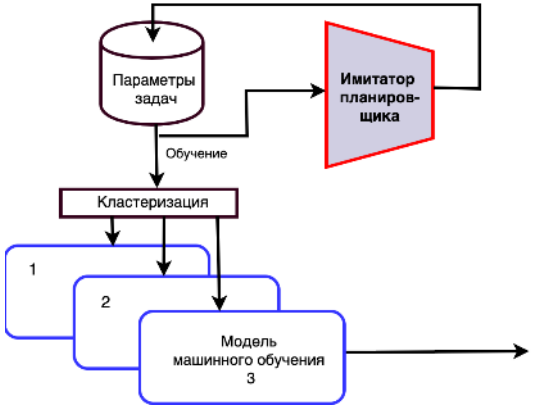
Возможность 4: реконфигурация вычислительных ресурсов «на лету»

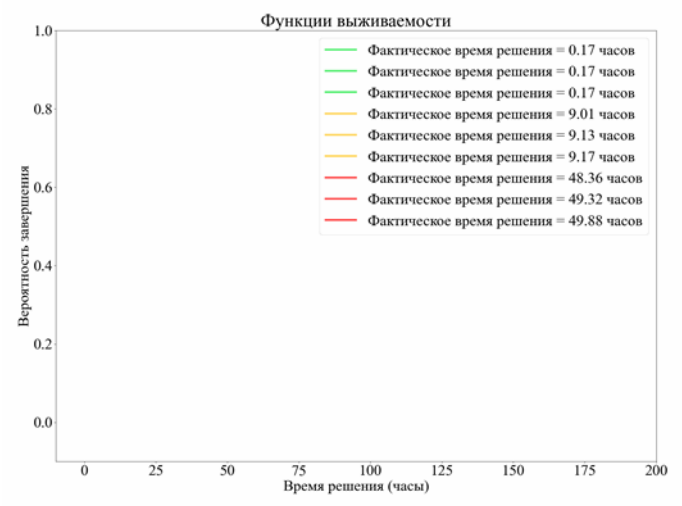


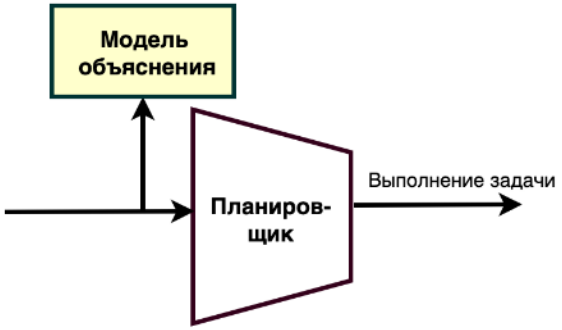
Управление процессом вычислений на основе реализации процедур «обучения» СК как выбирать различные конфигурации:

- на **аппаратном уровне –** реконфигурировать ресурсы используемой вычислительной платформы к особенностям алгоритма решения конкретной задачи
- на **программном уровне управлять** «траекторией» алгоритма решения задачи в пространстве аппаратных возможностей вычислительной системы.

Возможность 5: объяснение результатов вычислений с использованием методов ИИ







Результат вычислений **«число + объяснение»** того, что это число значит, как было получено и какие факторы наиболее значимы»

Что планируется сделать : Двухконтурная когнитивная модель процессов вычислений в ЦКП «Политехнический»

ПОЛИТЕХ

Выводы

Повышения эффективности СК возможно на основе не только роста производительности аппаратных ресурсов и новых программ, но и

- **построения** системы машинного обучения, которая прогнозирует время решения задачи и оценивает **компетентность** пользователей в использовании возможностей суперкомпьютерного моделирования
- сокращения времени решения прикладных заданий за счет накопления «опыта» по настройке параметров вычислительной системы
- **встраивания** в среду управления СК ресурсами «**системы интерпретации**», результатов расчетов
  - формирования вектора смысла результата (вектора внимания) с учетом использования различных параметров алгоритмов и характеристик вычислительных ресурсов
    - (число используемых узлов кластера, вид ускорителей вычислений (GPU/FPGA), объем оперативной памяти и пр.) на их успешное завершение задачи пользователя

ПОЛИТЕХ

Приложение
Тест на понимание

- Входом языковой модели являются «эмбеддинги» токенов анализируемого текста, а выходом - результат, который зависит от содержания задачи.
- Существует стандартный набор задач, который нужно выполнить на стандартном наборе данных, чтобы доказать, что трансформер справляется с задачей пониманием текста.
- Пример задачи —
  - выдать 1, если в двух разных вопросах спрашивают одно и то же,
  - выдать 0 — если нет.

ПОЛИТЕХ

Код формы

Форма объекта «кодирует» символьный дескриптор, который выступает как качестве не только описания объекта, но вариантов его взаимодействие с внешней средой.
Семантика дескриптора – динамическая сущность, которая меняктся в зависимости от контекста его использования.

Семантику дескриптора можно вычислить, анализируя корреляции между «словами» текста, где этот код используется

В тексте «Гамлета», очевидно, есть смысл, сюжет, персонажи и пр.
Но чтобы понять «смысл» надо совершить некоторую «работу», а именно, открыть книгу, различать буквы, читать текст и разбирать слова

ПОЛИТЕХ

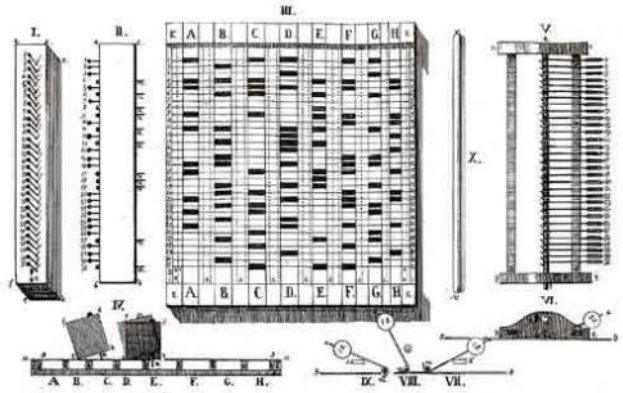
Исчисление смыслов

- Первая попытка дать исчисление смыслов, или того что образует множество «идеальных реальностей», принадлежит индийской эстетике IX-X веков

- В 18 веке И. Кант множество «идеальных реальностей» определил как совокупность умопостигаемых сущностей – ноуменов.
- Ноуменальное множество представимо в форме текста, поэтому текст это разновидность смысла.
- Смыслы могут быть представлены в речевом потоке слов, но в этом случае одновременно присутствуют весьма многочисленные смысл как компоненты культурной реальности (понятия вне текущей потока слов)

Начало технологии
идеоскоп С. Н. Корсакова

- Изобретенные С. Н. Корсакова (1787-1853). Механическая машина позволяют находить, сравнивать и классифицировать множества информационных записей (идей) по набору многочисленных признаков (деталей), позволяя находить:



- 1) все соответствия, которые есть у сравниваемых **идей** при их соприкосновении; 2) все то, что находится в заданной идее, но отсутствует в той **идее**, с которой ее сравнивают, в сей момент; 3) все то, что отсутствует в заданной **идее**, но есть в той идее, с которой ее сравнивают; 4) все то, чего нет ни у одной, ни у другой **идеи**, но есть у других идей из той же таблицы

Эволюция концепция идеоскопа от идей «редукции» к идеи «генерации»

- Логическая схема редукционизма проста – целое состоит из частей. А для того, чтобы понять все целое – достаточно понять все его части.
- Это не позволяет описать свойство «эмерджентности» : целое «реальнее» своих частей:
  - квантовые частицы не имеют определенных траекторий (принцип неопределенности Гейзенберга);
  - «волновая» пси-функция квантовой частицы принципиально не может быть измерена никаким прибором, то есть является невещественной сущностью,
  - только умозрительное, т.е. доступное исключительно разуму человека, позволяет понять свойства того, что чувственно воспринимается или измеряется приборами.
- Чтобы описать эмерджентность нужны механизмы описания процессов синтеза и свойства «генеративности» , такие как, например, Generative Pre-Training Transformer (GPT) - языковое моделирование — как генеративное предсказание следующего слова (или фразы из слов) с учётом предыдущего контекста.

Проблема множественности смыслов слов

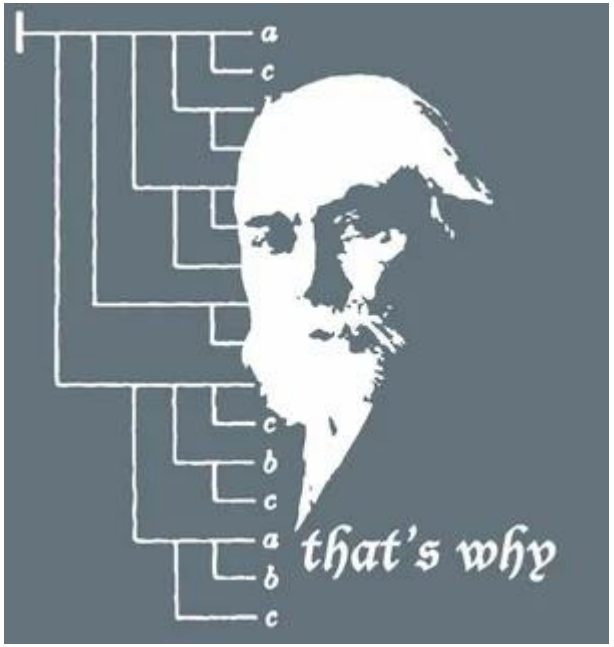
Было: Перечислимость множеств,
Вычислимость функций,
разрешимость множеств…

Требуется: Объяснимость результатов вычислений

Актуальная бесконечность – реализуется тогда, когда часть может быть равна целому

Смысл и значение - семантический треугольник

Готлоб Фреге 1892 определил понимание семантики знаковых выражений.



Фреге вводит третью семантическую составляющую , которую называет «смысл»: теперь знак относится к означаемому не напрямую, а через составляющую смысла.

ПОЛИТЕХ

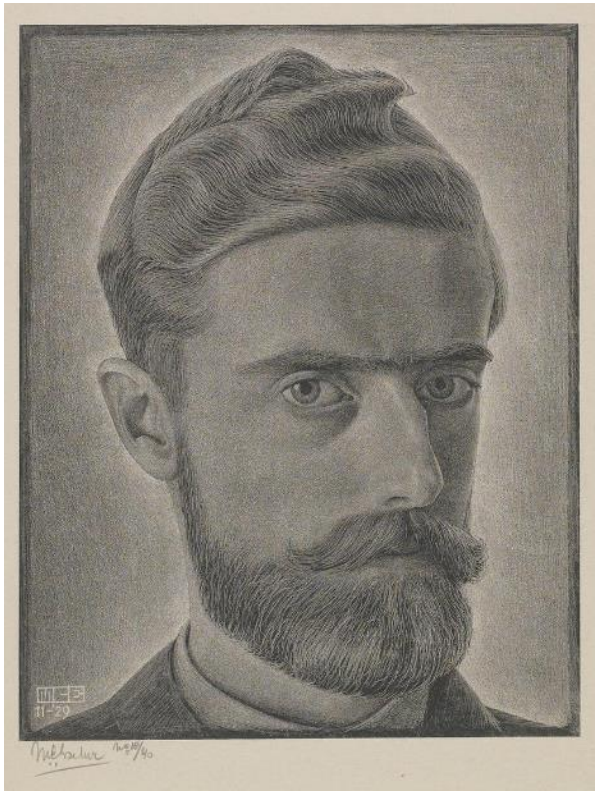
СЕМАНТИЧЕСКАЯ МОДЕЛЬ ФРЕГЕ

Эта модель решение множества трудных проблем в различных областях науки.
Например, из нее следует, что знак может иметь смысл, но не иметь значения (или предмета, который он обозначает).

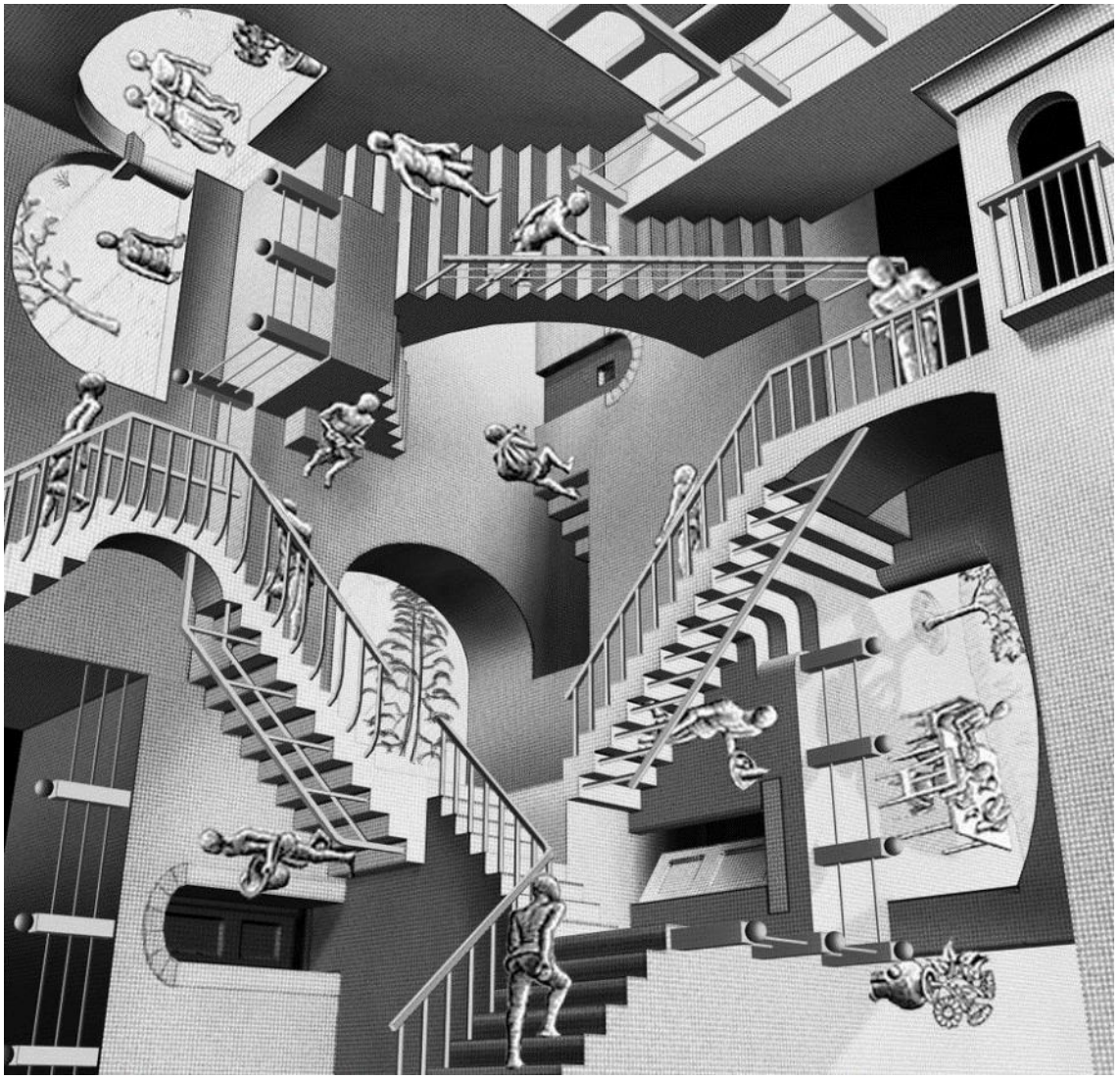
Не имея предмета, к которому он относится, знак тем не менее имеет смысл (если только это не бессмысленная конструкция, нарушающая правила семантики).
Например, такими знаками в математике является число π или число $\sqrt{2}$.
Итак, знак не является бессмысленным, даже если он ничего не обозначает. Несуществование предметов, которые обозначают такие выражения, **как «круглый квадрат»,** не мешает нам понимать и считать осмысленными высказывания вроде: «Круглых квадратов не существует».

ПОЛИТЕХ

Пример того, что физически не реализуемо, но… имеет смысл



Мауриц Корнелис Эшер



Что из этого следует

выражения синтетическое выражение типа a = b принципиально отличающимися от выражений типа тавтологии a = a.
Итак, синтетические суждения дают нам прирост знания о предмете. Но … выражение «вода — это H_2O», сформулированное в 18 веке Лавуазье ничем синтаксическки не отличается от выражения «вода — это xyz»

Из семантической модели Фреге следует, что мы достигаем существенного прироста знания, если связываем с одним и тем же предметом выражения (знаки), имеющие разный смысл, причем связь смыслов этих знаков с одним и тем же предметом **не является очевидной.**

Пример: название планеты Венера

«Утренняя звезда»

«Вечерняя звезда»

Планета Венера

Что из этого следует

- Смысл есть у знаков, не обозначающих никаких предметов.
- С одним предметом может быть связано множество смыслов.
- Зная смысл знака или выражения, мы не всегда можем установить предмет, который этот знак обозначает.
- Установление соответствия составляет сущность, например, научного открытия. Смыл объективен и интерсубъективен,поэтому доступен для понимания разным участникам коммуникации.

Топологический смысл – объект один, а форм его представления много

- Топология изучает свойства метрических пространств, которые остаются неизменными при непрерывных деформациях
- Топологи́ческое простра́нство — множество с дополнительной структурой определённого типа (топологией)

ПОЛИТЕХ

Комментарий:

- Итак, наблюдатель объективно не создает никакой новой информации, но придает полученным данным смысл.
- Это равносильно утверждениям - информация существует и передается в разных формах:
  - элементарные частицы фермионы с ненулевой массой покоя хранят информацию о себе, формируя информационное содержание материи, хранящейся в Вселенной,
  - Элементарные частицы бозоны - носители взаимодействия могут передавать информацию только в форме сигнала
  - Слова в тексте образуют смысл фразы с учетом контекста и места, где этот текст воспринимается
  - ……

Смыслы и контексты

Смыл слова зависит от контекста. Информация есть и «причина» и «смысл» возникновения событий

- Формализация возможна и основана на априорной гипотезе (априорном смысле), доказательства которой есть наблюдаемые факты. Для совершения действий (появления события) надо вычислить, вероятность того, что принятая гипотеза верна с учетом «смысла» новых фактов, а именно:
    - Используем вероятность P (доказательство|гипотеза), чтобы ответить на вопрос: «Какова вероятность наступления событий-доказательств в том случае, если принятая гипотеза верна?»
- Заметим, что вероятность P (доказательство|гипотеза) оценить легче, чем вероятность P(гипотеза|доказательство), так как у вероятности P (доказательство|гипотеза) –гораздо более ограниченная область суждений о «мире» - сужая область, можно упростить задачу.

(аналогия: «огонь» –гипотеза, а наблюдение дыма – событие, доказывающее наличие огня. Вероятность P(огонь|дым) оценить сложнее, поскольку вызвать дым могут различные события, например, выхлопные газы). P(дым|огонь) оценить проще, где есть огонь, наверняка будет и дым