Анализ путей построения AGI с точки зрения синергетики

Франсуа Шолле — ведущий эксперт в области искусственного интеллекта. Он создатель библиотеки Keras, автор книги «Глубокое обучение с Python» и основатель премии ARC Prize, международного конкурса, направленного на измерение истинного общего интеллекта.

Владимир Смолин

smolin@keldysh.ru

Институт прикладной математики имени М. В. Келдыша РАН, Москва, Миусская пл., 4, 125047, Россия

Семинар AGIRussia, 11 декабря 2025 г. (сетевой, модератор Колонин А.Г.)

Подходы, теории, определения, задачи, проекты

Живая экосистема научно-инженерной мысли:

•**Подходы** задают общее направление.

•**Теории и определения** — это инструменты мышления и коммуникации.

•**Задачи** — это вызовы, фокусирующие усилия.

•**Проекты** — это двигатели прогресса, где теория проверяется практикой, а практика ставит новые вопросы теории.

•**Примеры подходов в ИИ:**

- **Символический подход (GOFAI):** манипуляция символами и логическими правилами.
- **Коннекционистский подход (нейронные сети):** Интеллект возникает из обучения нейронов.
- **Задачный подход (Task-Oriented):** Системы для решения конкретных задач, а не общего ИИ.
- **AGI-подход:** Цель — создать универсальный интеллект.
- **Синергетический подход:** ***взгляд на мир (и на ИИ) как на иерархию сложных, открытых, нелинейных систем, способных к самоорганизации и рождению нового.***

Подход определяет, *какого рода задачи* мы считаем достойными решения (например, задачный подход фокусируется на узких, измеримых задачах).
Теория предлагает **качественные описания** или *математические формализации* для этих задач (например, задача классификации формализуется как поиск разделяющей гиперплоскости).
Определения задают *границы задачи* (что такое "точный перевод"?)
Задача - четко сформулированная проблема, которую нужно решить.
Проект — это точка, где все абстрактные уровни встречаются с реальностью.

Ученые рассказали о главном преимуществе мозга над ИИ

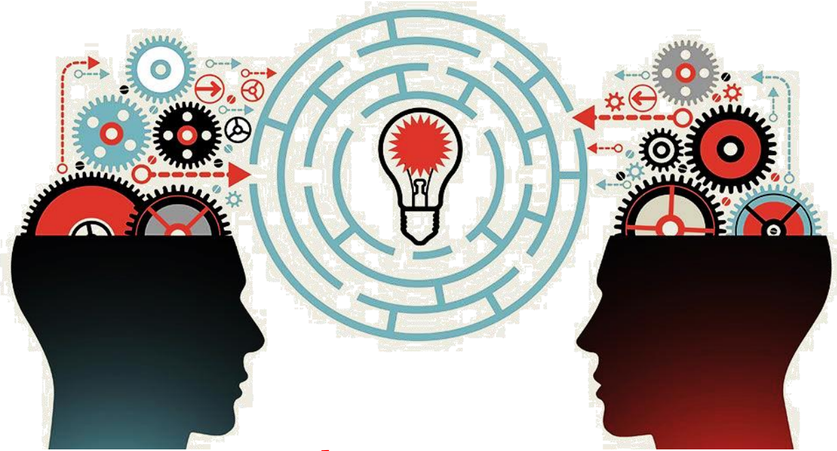
(и как его можно скопировать)

Человек не учится заново, а ***переиспользует*** уже знакомые навыки и комбинирует их с новыми. Мозг делает это автоматически, а вот для современных нейросетей это проблема — ***бенчмарки вроде ARC-AGI показывают***, что даже если модели показать несколько примеров решения задачи по одному и тому же правилу, она не всегда справляется с еще одной похожей задачей. Исследователи из Принстона выяснили, почему. Статья опубликована в Nature.

"Я думаю о когнитивном блоке как о функции в компьютерной программе, — объясняет Тим Бушман, старший автор исследования. — Один набор нейронов различает цвета, и его выход соединяется с другой функцией, которая управляет действием".

Tafazoli, S., Bouchacourt, F.M., Ardalan, A. *et al*. Building compositional tasks with shared neural subspaces. *Nature* (2025). https://doi.org/10.1038/s41586-025-09805-2

Переход количества в качество: от шестерёнок и алгоритмов - к аппроксимация преобразований

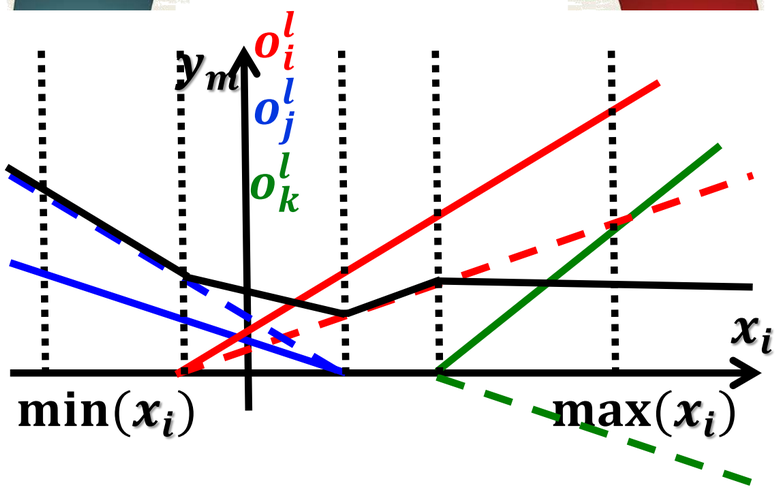


В эпоху торжества механики деятельность мозга было принято представлять как работу взаимодействующих шестерёнок.

В эпоху ЭВМ – в виде алгоритмов.

После нейросетевой революции в МО – появилось описание работы нейросетей как аппроксиматоров преобразований!

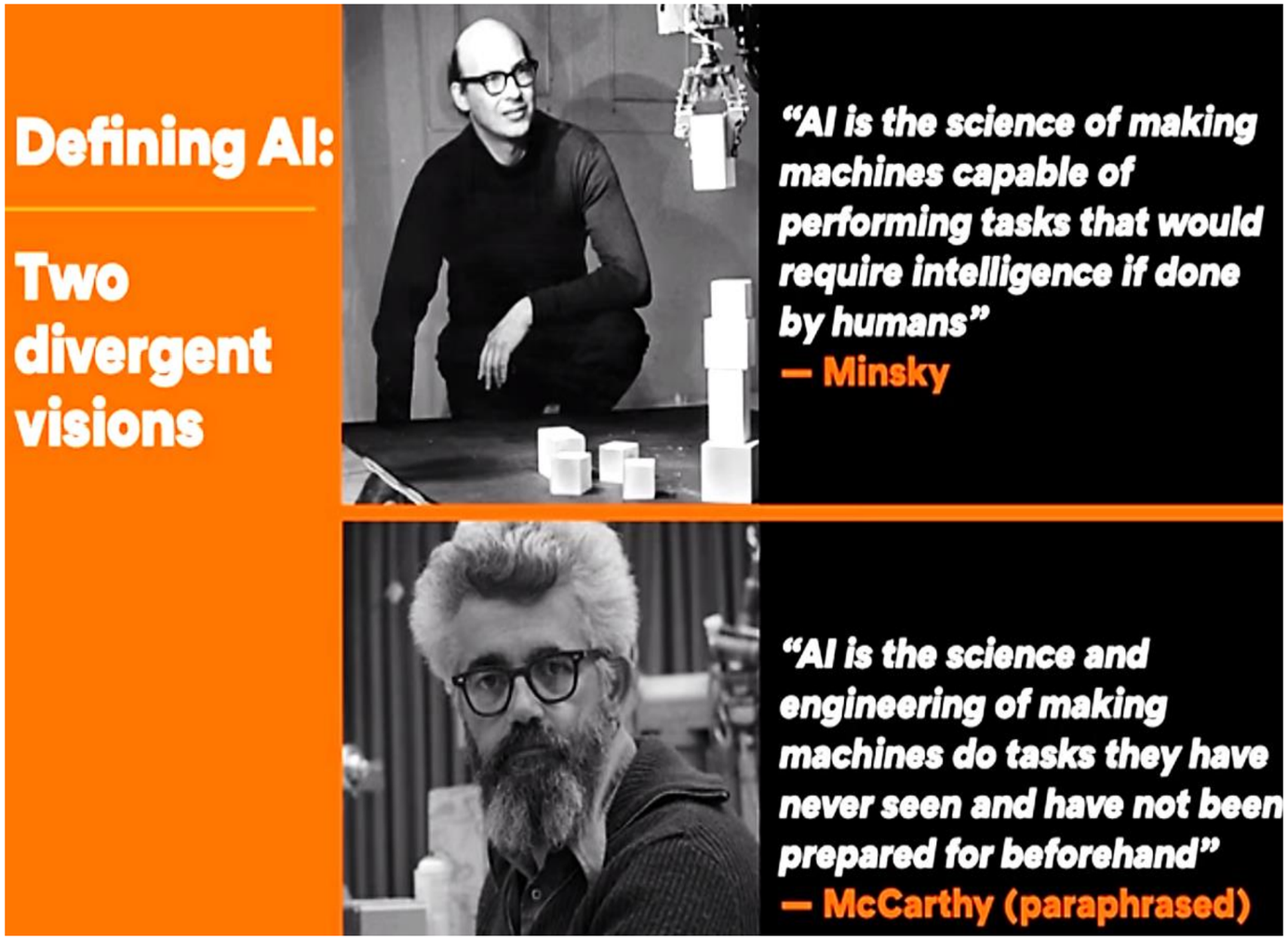
Но оно ещё только идёт в массы и шестерёнки и алгоритмы доминируют в массовом сознании.



Аппроксимация – это тоже алгоритм, но ***количество его параметров качественно отличает его*** от систем шестерёнок и блок-схем, пусть и с адаптивными параметрами.

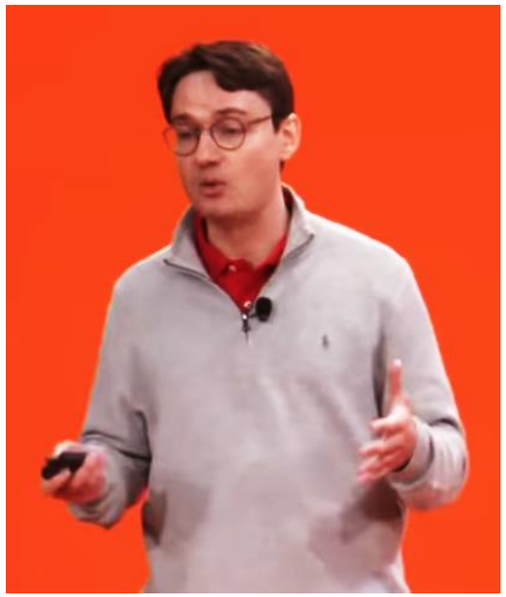
Как вода, которая обладает текучестью, но течёт в направлении уклона, который задан внешними условиям, так и алгоритмы аппроксимации обладают свойствами приближённо отображать значения функций. Но функции задаются внешними условиями.

Два расходящихся мнения



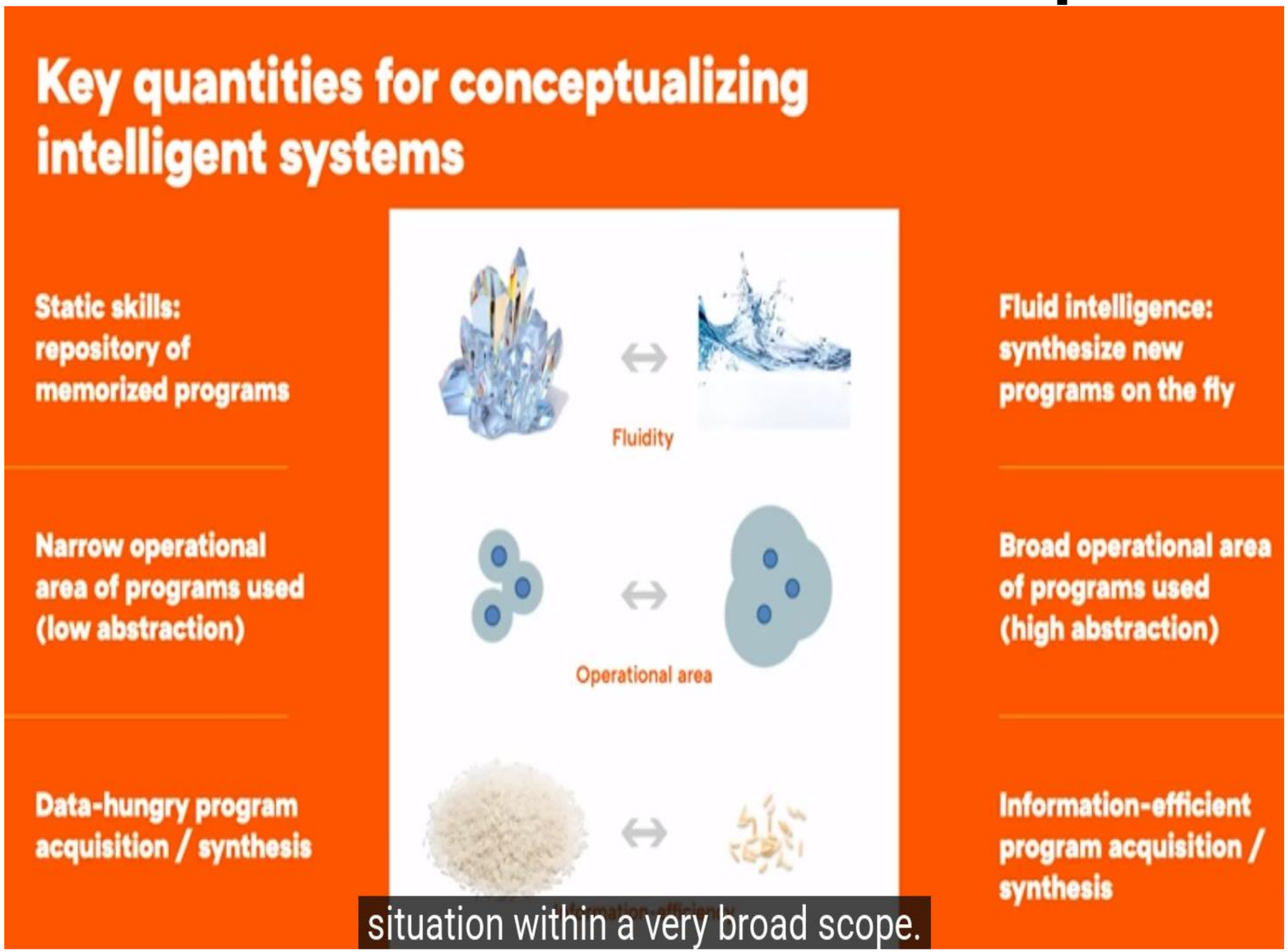
Франсуа Шолле:
Есть 2 различных подхода к ИИ: восходящий и нисходящий.
Именно нисходящий подход позволяет решать задачи, к которым машинные системы не были подготовлены заранее.
(Ниже объясню, почему я так не считаю).

Что такое интеллект?



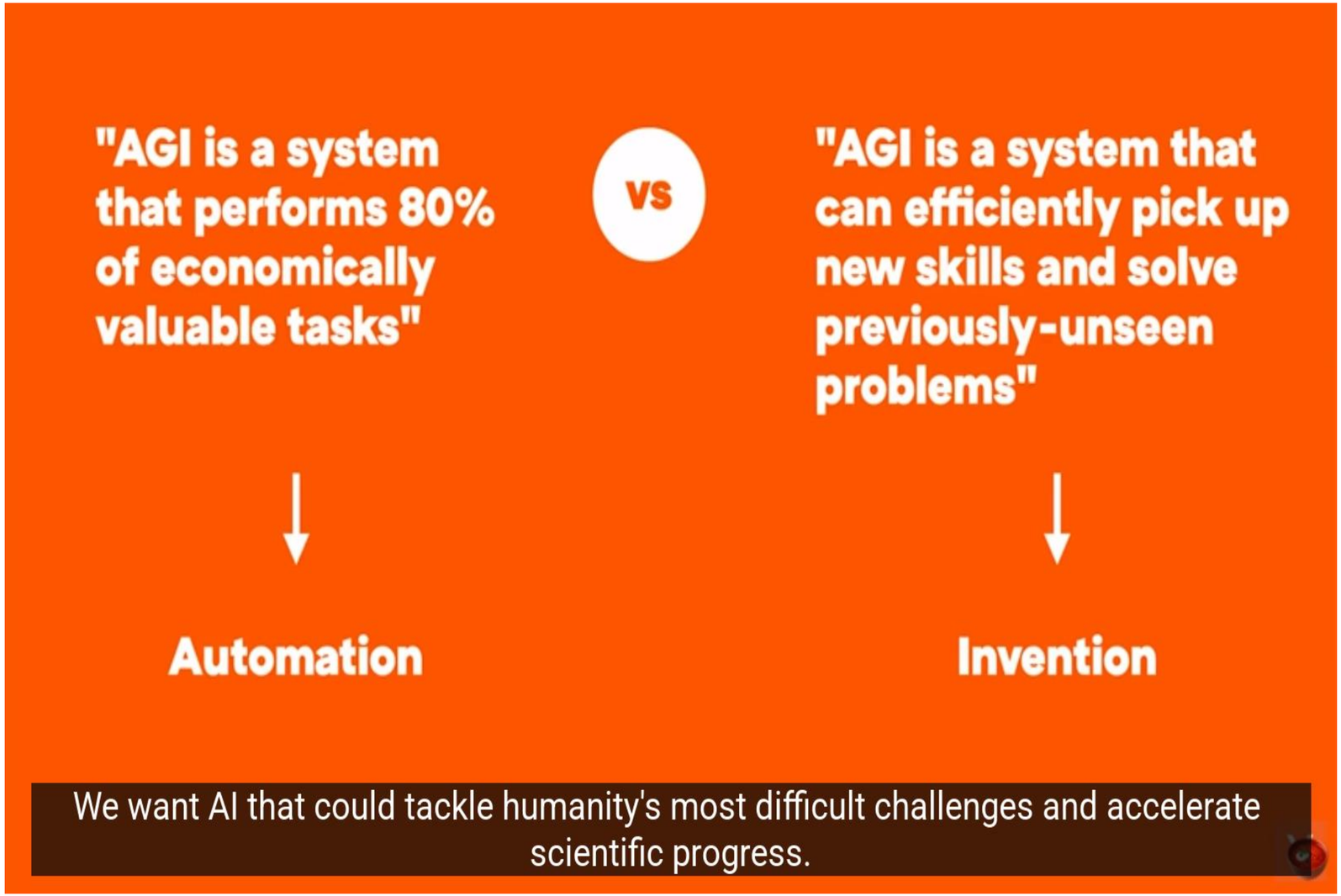
Франсуа Шолле:
Интеллект – это способность строить новые, более эффективные, способы решения задач.

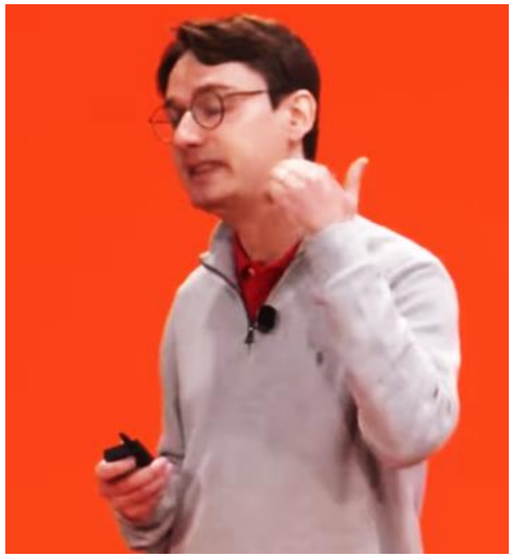
Зачем понимать и измерять интеллект?



Франсуа Шолле:
Следует оценивать информационную эффективность.
Инженеры могут сделать только то, что можно измерить.
То, ***как мы определяем и измеряем интеллект*** – это не техническая деталь, а ***отражение нашего понимания проблемы познания***.
Решая задачи вы достигаете цели, но упускаете суть.

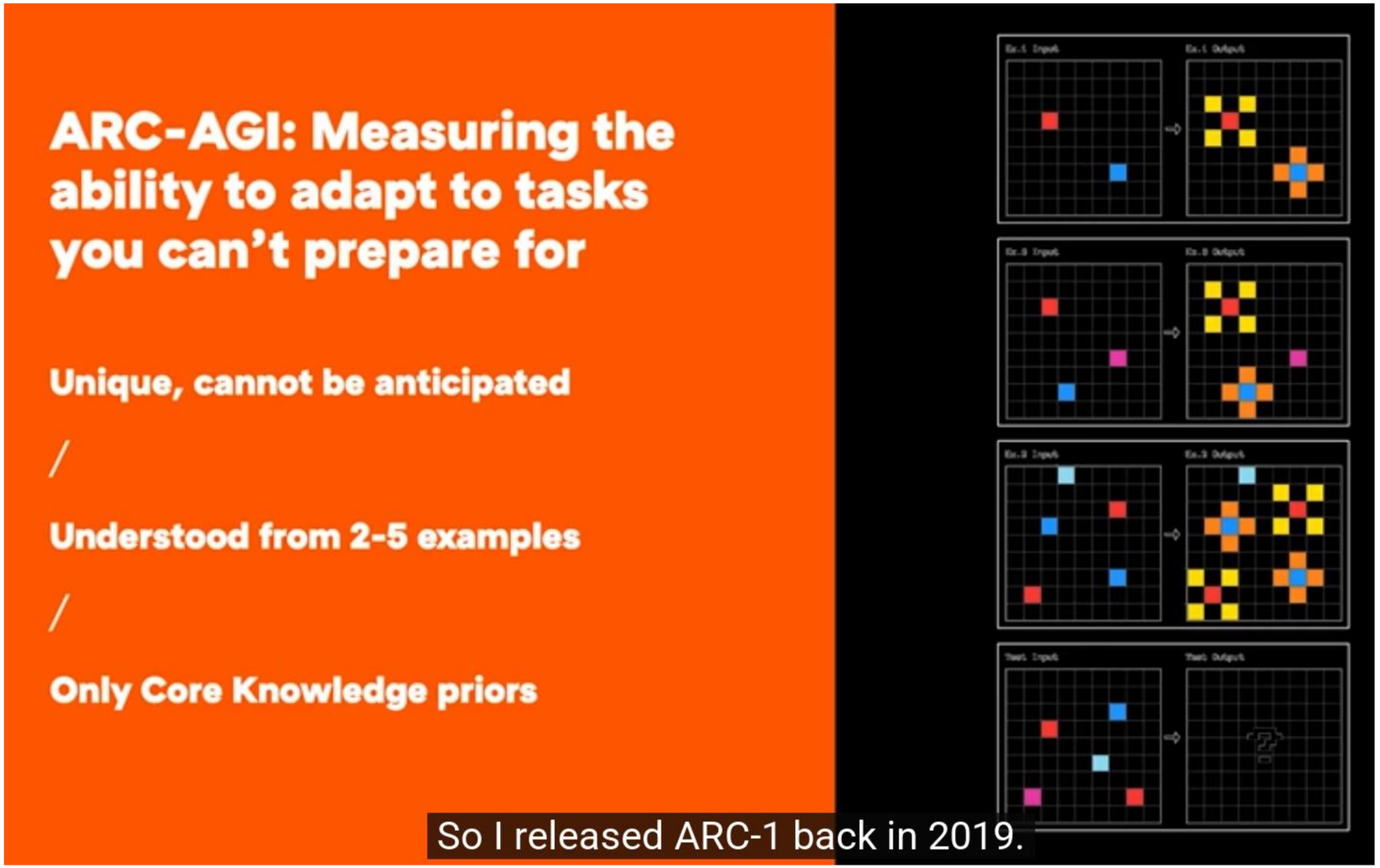
А что такое AGI?





Франсуа Шолле:
AGI - это система, которая позволяет эффективно приобретать новые навыки и решать невиданные ранее задачи.

Тесты ARC-AGI - 1, 2, 3…



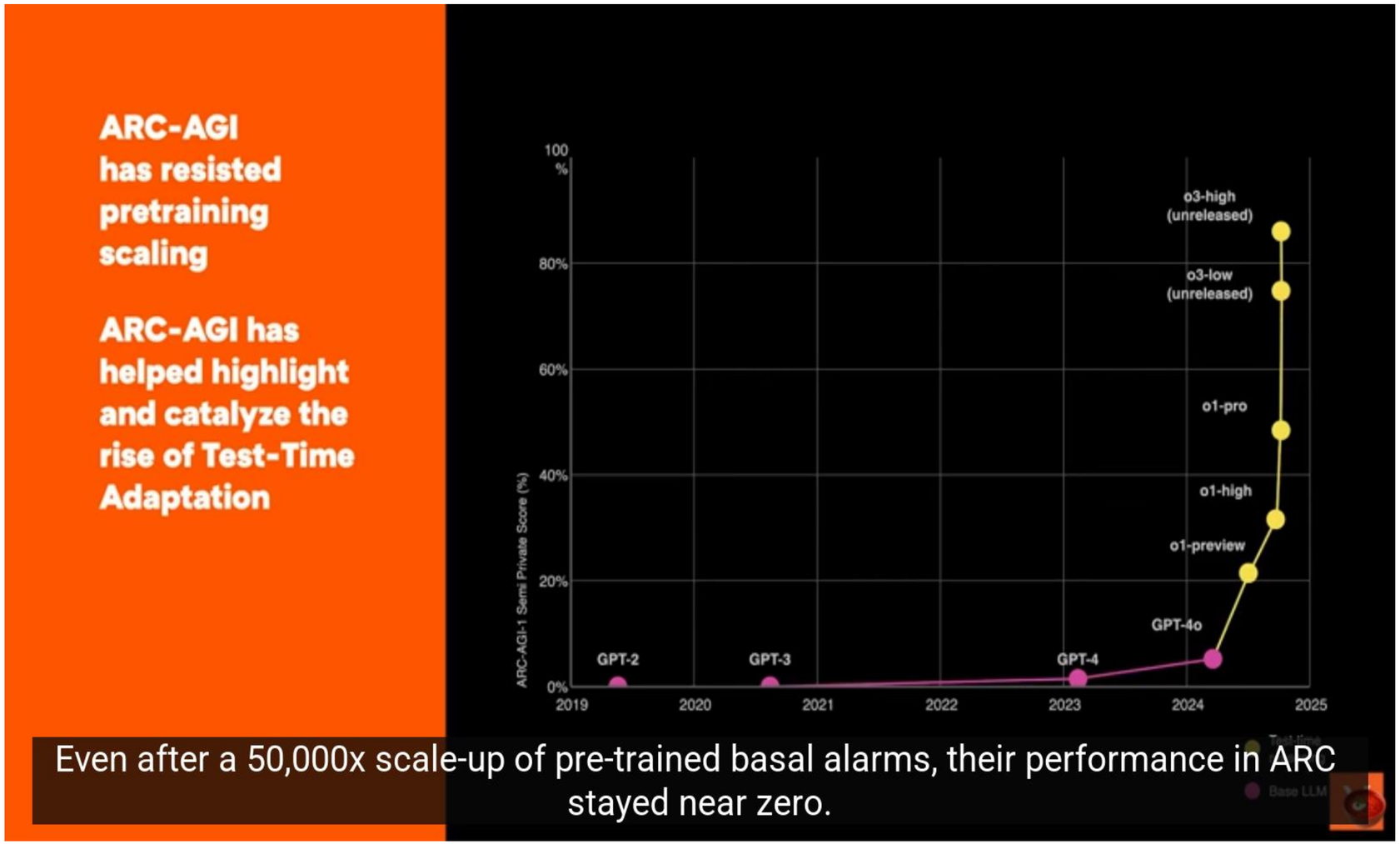
Abstraction and Reasoning Corpus (ARC) AGI Benchmark

Франсуа Шолле:
Тест ARC-AGI: оценка способности адаптироваться к задачам, к которым вы не можете подготовиться

Почему тесты ARC-AGI важны?

Франсуа Шолле:
ARC-AGI требует разработки новых правил на лету путем рекомбинации того, что вы знаете.
Это очень просто для людей, но сложно для современного ИИ.

Метод «грубой силы» не решает ARC-AGI

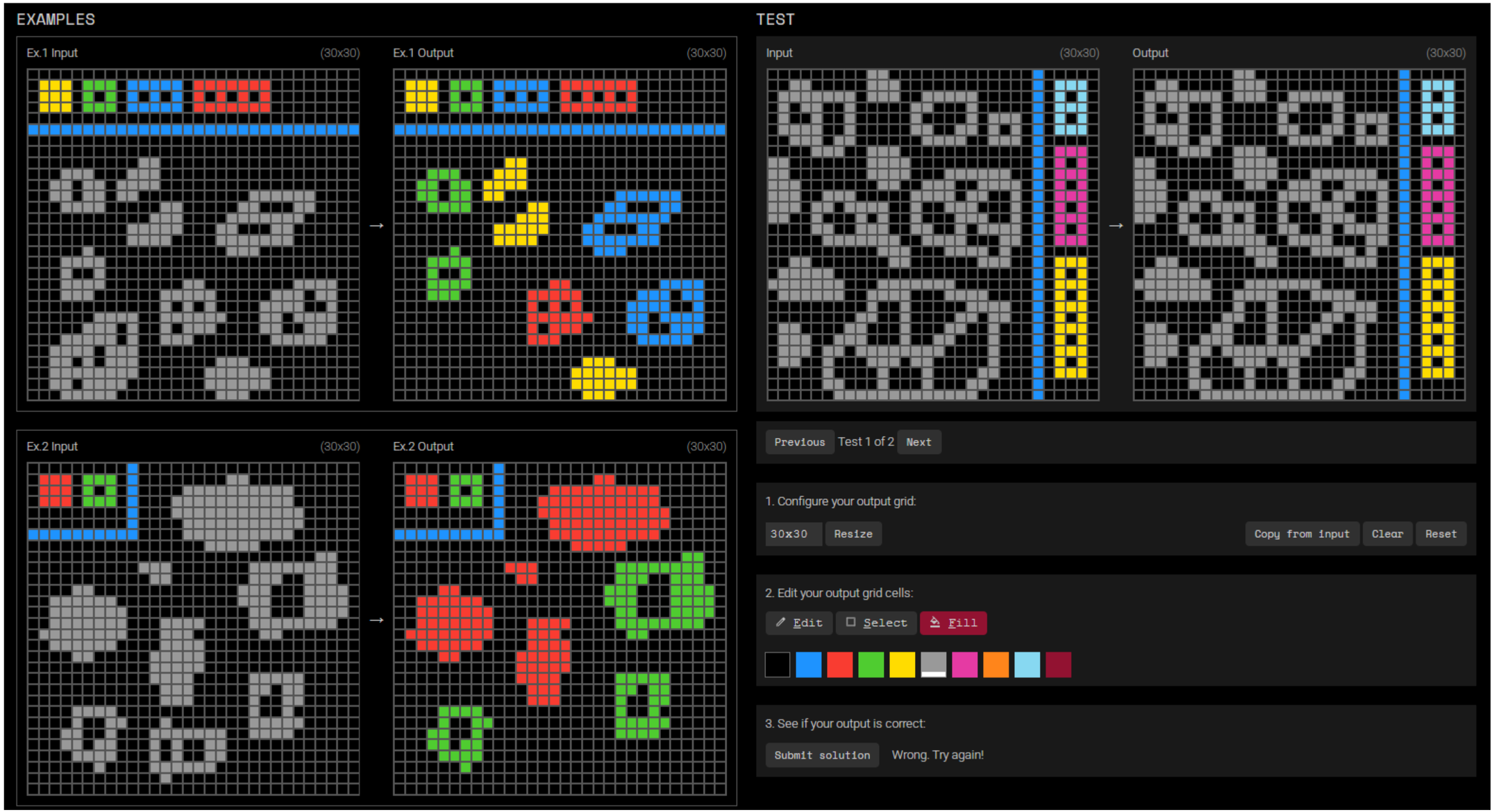


Франсуа Шолле:
Даже после масштабирования в 50000 раз результативность ИИ оставалась около нуля.

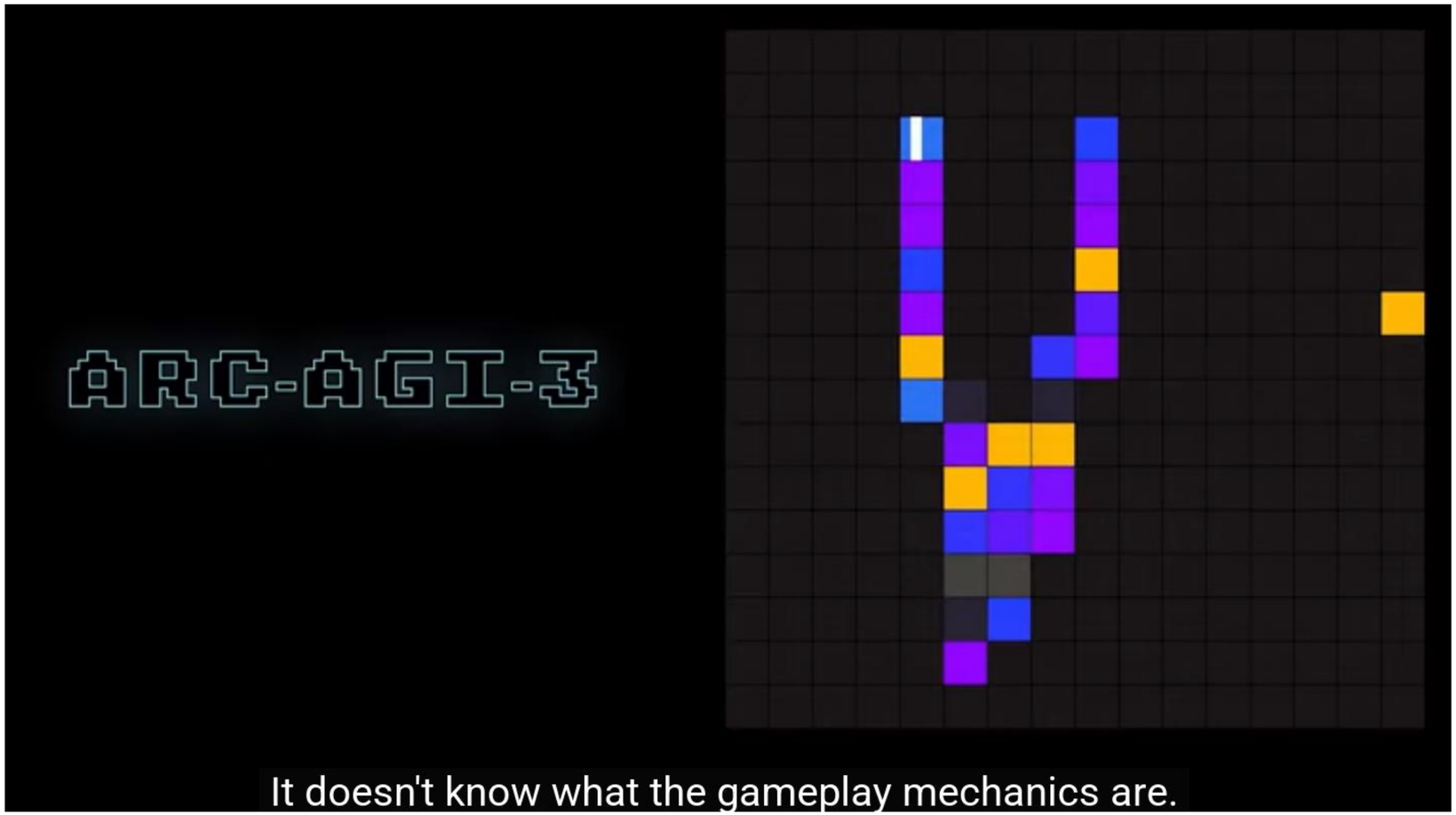
Справились с ARC-AGI-1? Вот ARC-AGI-2, будет и 3…

Франсуа Шолле:
Следует оценивать информационную эффективность. Инженеры могут сделать только то, что можно измерить.
То, ***как мы определяем и измеряем интеллект*** – это не техническая деталь, а ***отражение нашего понимания проблемы познания***.
Решая задачи вы достигаете цели, но упускаете суть.

ARC-AGI-2 - более сложные правила

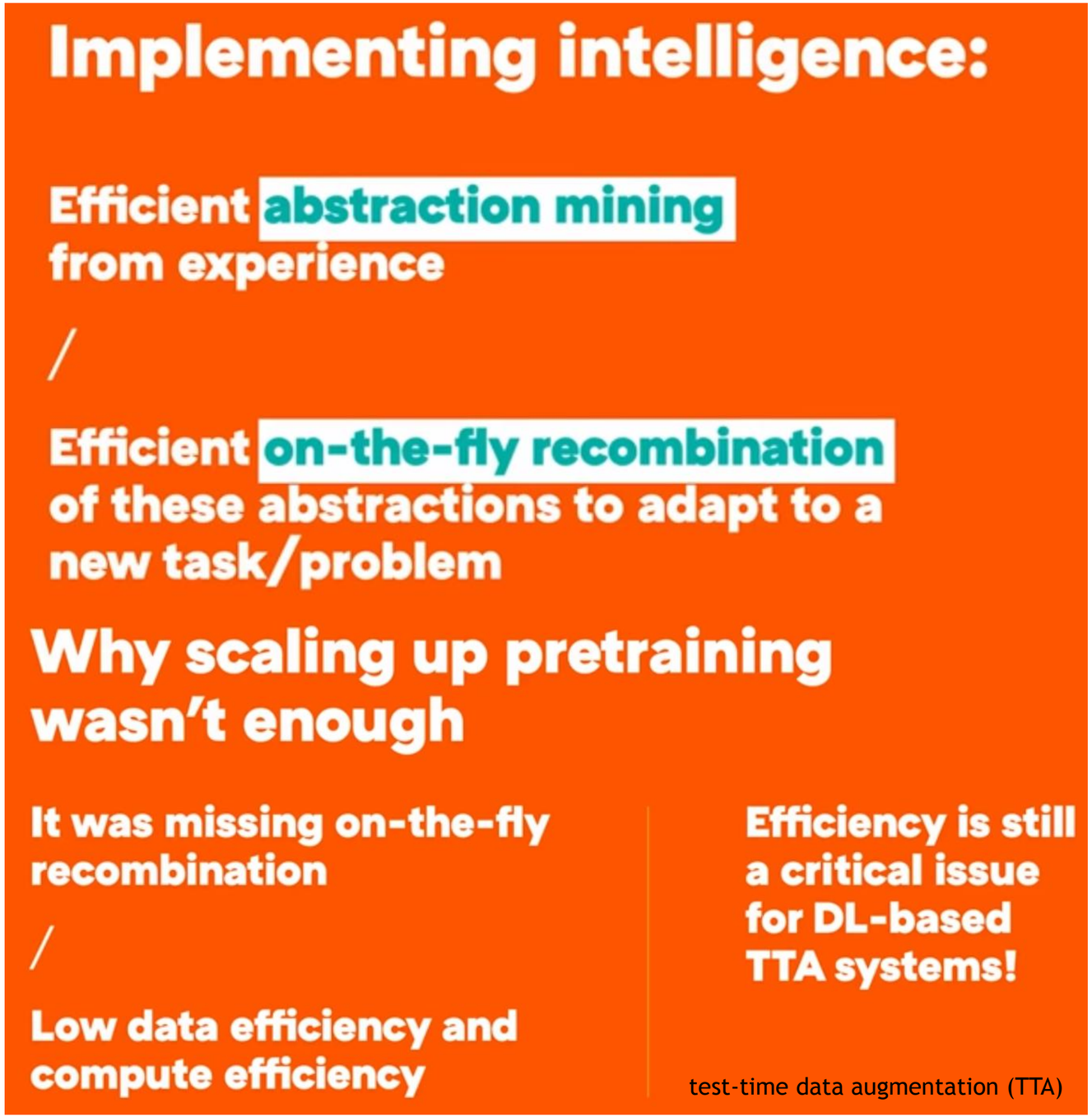


ARC-AGI-3 – а будут и 4, и 5



Франсуа Шолле:
Что же не хватает для построения AGI?
Возможности эффективно использовать свой опыт для встречи с будущим (*решать ранее не встречавшиеся задачи в сложной среде*).
Ничто не бывает по-настоящему новым (!?)

Чего не хватает для AGI?



***Франсуа Шолле*:**
Что же не хватает для построения **AGI?**
1. Эффективного формирования абстракций на основе накопленного опыта;
2. Эффективная рекомбинация этих абстракций на лету для адаптации к новой задаче/проблеме.
(Рекомбинация предполагает декомпозицию…)

Два типа абстракции

The two poles of abstraction: type 1 vs type 2

Prototype-centric (value-centric) abstraction

Set of prototypes + distance function
Example: classify face vs. non-face using abstract features

Abstract wrt details not present in the prototypes

Obtained by clustering concrete samples into prototypes
This is a value analogy!

Program-centric abstraction

Graph of (usually discrete) operators where input nodes can take different values within a type
Example: function that sorts a list

Abstract wrt input nodes values

Obtained by merging specialized functions under a new abstract signature
This is a program analogy!

```
a = [4, 5, 2, 6]
result = 0
for i, e in enumerate(a):
  result += i * e

b = [6, 3, 4, 7]
result = 0
for i, e in enumerate(b):
  result += i * e
```

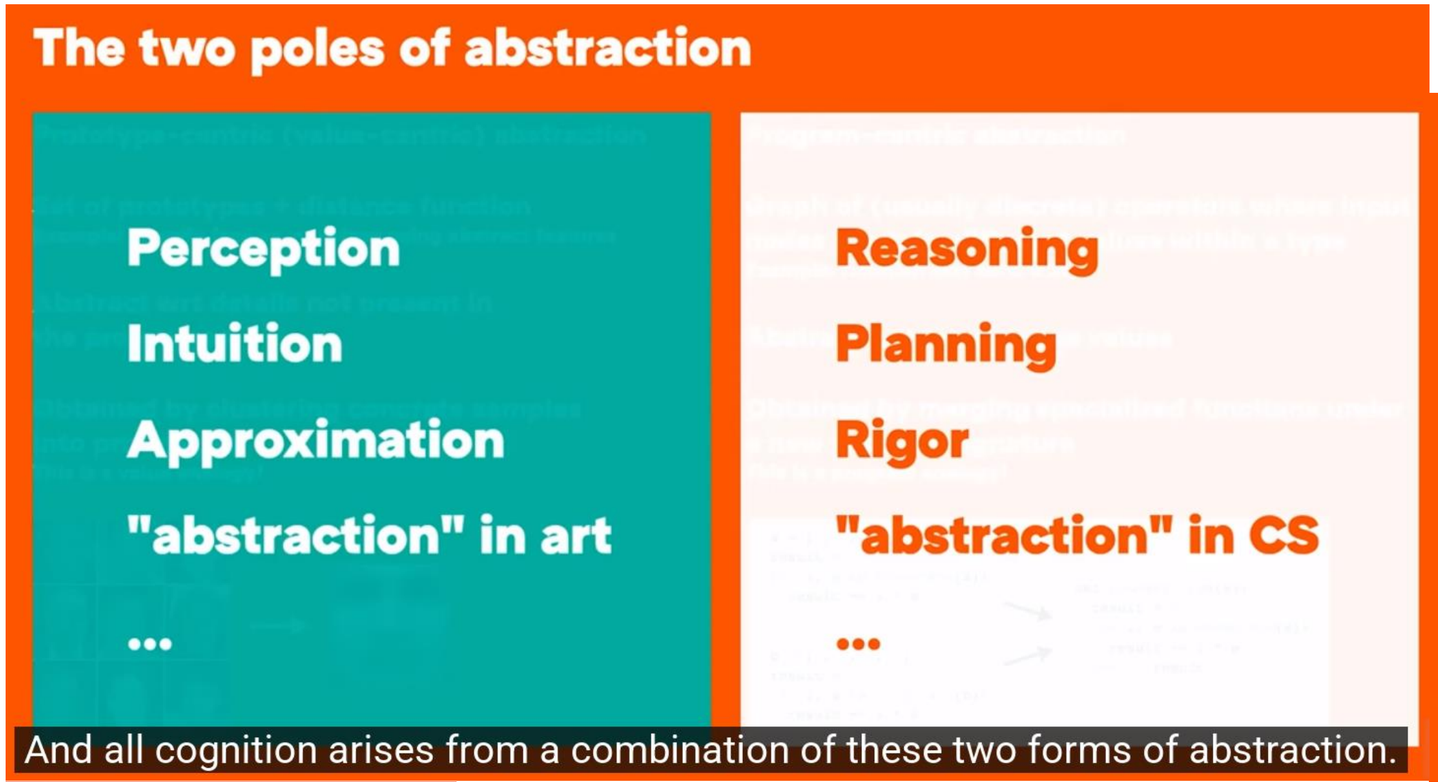
```
def process_item(x):
  result = 0
  for i, e in enumerate(x):
    result += i * e
  return result
```

It's also what software engineers do when they're refactoring some code.

Франсуа Шолле:
Абстракции:
- прототипо-центричные и
- программно-центричные.

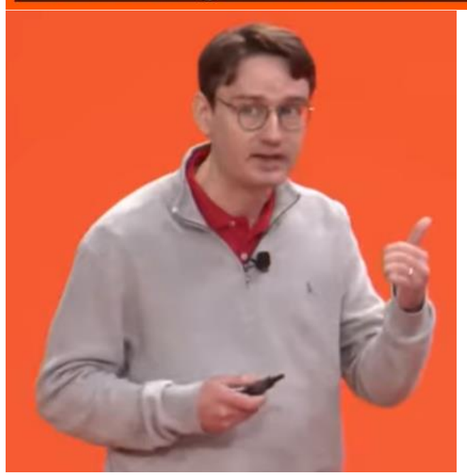
(Отмечу, что нейросети не усредняют образы и не считают расстояния с центром кластера, а аппроксимируют преобразования. Но действительно, аппроксимация сильно отличается от большинства алгоритмов, используемых в подходе GOFAI).

Два типа абстракции



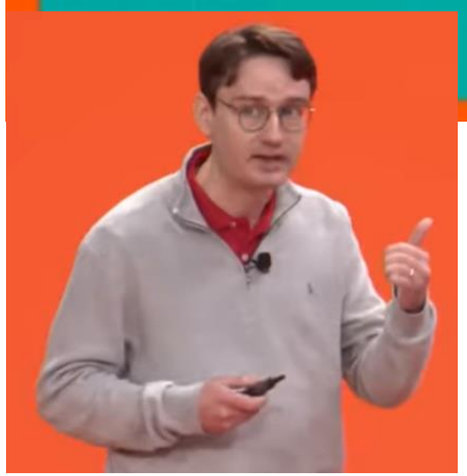
Transformers are a great type 1 abstraction machine.

But how do we get to type 2?



Франсуа Шолле:
Трансформеры прекрасно подходят для абстракций типа 1.
Как работать с абстракциями типа 2? – Есть компьютерсайнс!

Полушарная аналогия

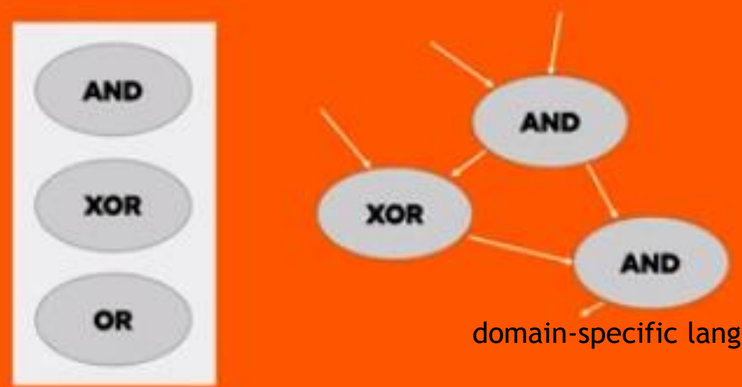


Франсуа Шолле:
Два полюса абстракции прекрасно (?!) соотносятся с полушариями.
Левое полушарие – интуиция, правое - размышления!

Синтез программ

Program synthesis from input/output pairs: overview

Modern machine learning	Program synthesis from I/O pairs
Model: differentiable parametric function	Model: graph of operators from a DSL
Learning engine: SGD	Learning engine: combinatorial search
Feedback signal: loss function	Feedback signal: correctness check
Key challenge: data (obtaining a dense sampling of the problem space)	Key challenge: combinatorial explosion



domain-specific language, DSL

Доказательство концепций, основанное на восприятии и интуиции

Natural intelligence as a proof-of-concept: reasoning in humans is heavily guided by perception and intuition

Examples: high-Elo chess playing, theorem proving... even ARC

We intuit where to go then only discretely enumerate a small number of options

Франсуа Шолле:
Натуральный интеллект человека показывает (?!), что мышление - алгоритмическое (выполнение программ). Мы можем проинтуичить, куда идти, а затем дискретно выбираем небольшое число действий.

Борьба с комбинаторным взрывом

Concretely, how can type 1 help type 2?

Type 2 related techniques involve a lot of enumeration over combinations of things (walking down graphs/trees with large branching factors)

Key issue: combinatorial explosion

Type 1: turn a lot of discrete data into a data structure that enables fast, approximate judgement calls about the target discrete space

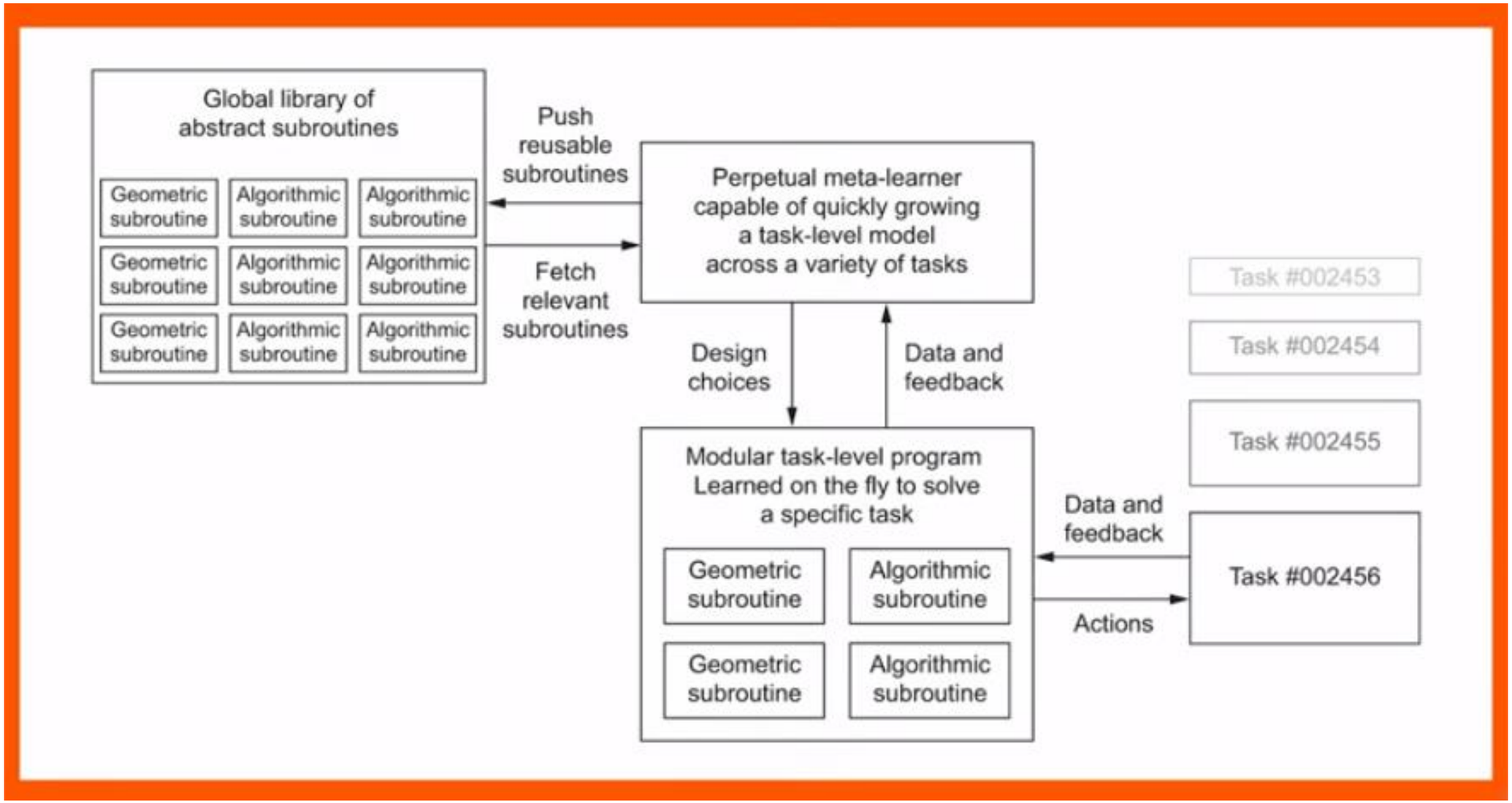
Only feasible if:
The discrete space has some interpolative macro-structure, i.e. the manifold hypothesis applies

A reasonably dense sampling of the space is available (aka lots of data)

Key idea: use these fast-but-maybe-wrong judgment calls to keep combinatorial explosion in check

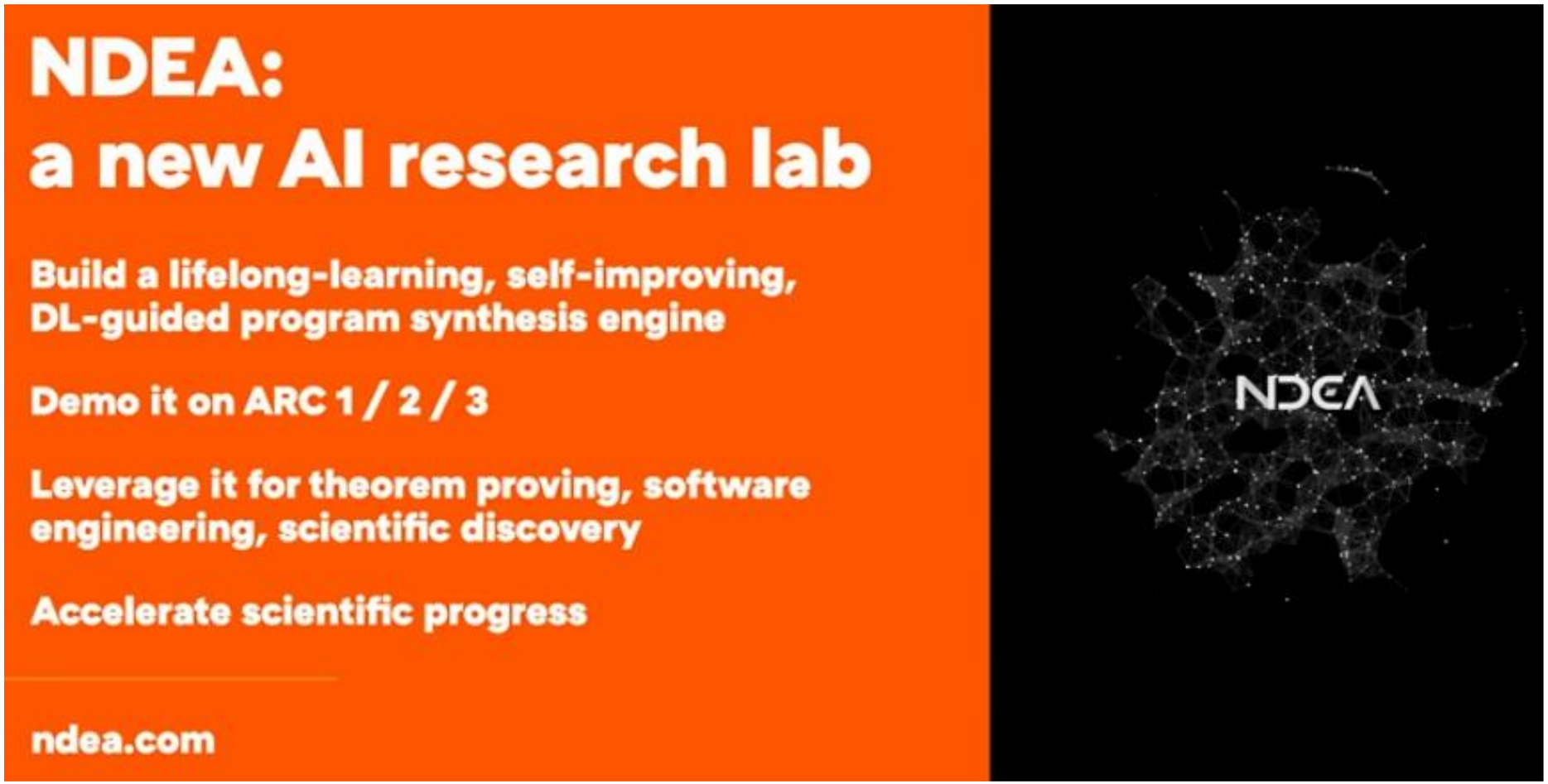
Франсуа Шолле:
Как конкретно тип 1 может помочь работе абстракций типа 2?
Приближённые оценки типа 1 могут быть направлены на предотвращение комбинаторного взрыва выбором ограниченного набора вариантов для сравнениияё

Концептуальная структура



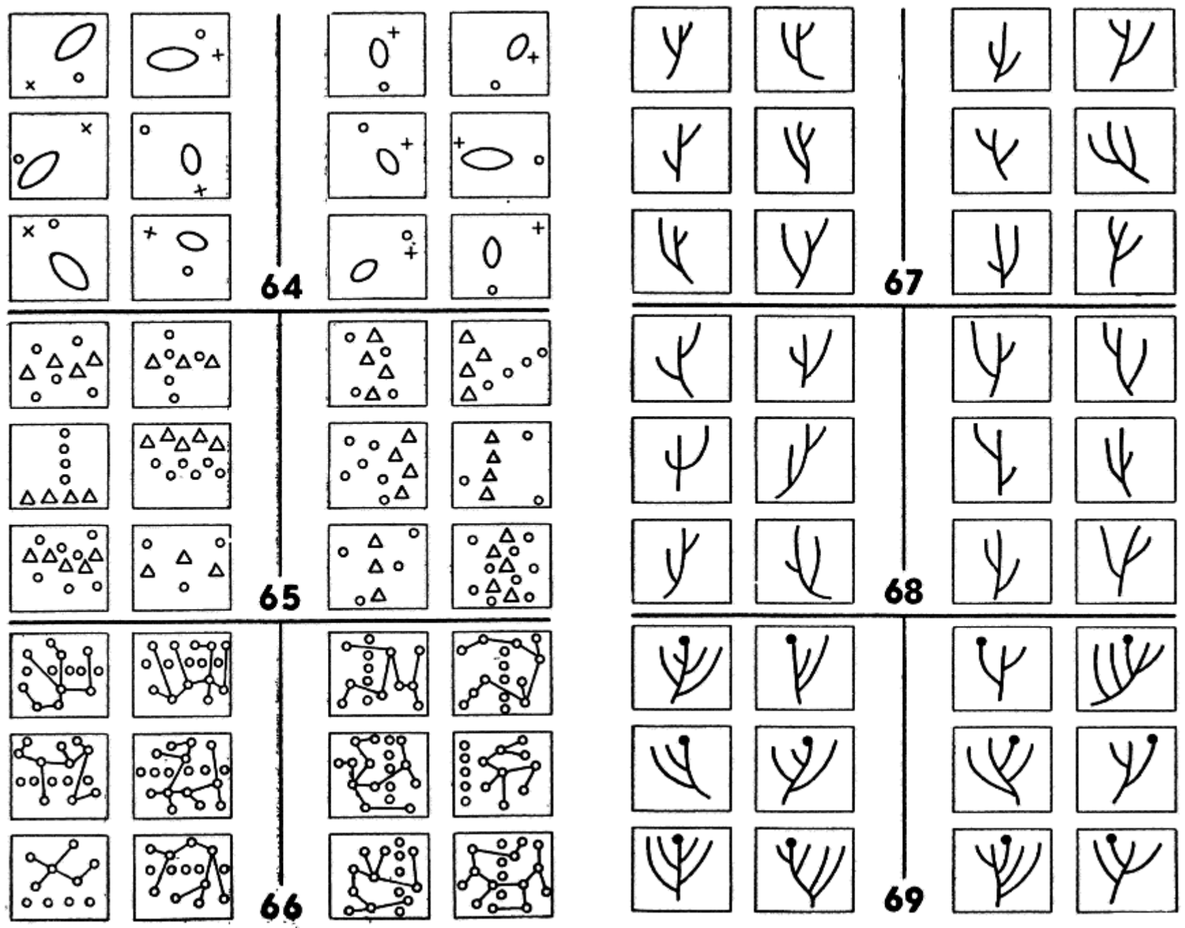
Где происходит обучение?!

Новая лаборатория ИИ - NDEA



We're starting Ndea — an AI research and science lab. The name - like 'idea' with an 'n' - is inspired by the Greek concepts ennoia (intuitive understanding) and dianoia (logical reasoning), capturing our first goal to merge deep learning with program synthesis.

Что было, когда ARC-AGI не было?



М. М. БОНГАРД

ПРОБЛЕМА УЗНАВАНИЯ

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1967

64. На продолжении оси эллипса — крест	—	На продолжении оси эллипса круг
65. Множество треугольников вытянуто горизонтально	—	Множество треугольников вытянуто вертикально
66. Свободные кружки лежат на горизонтальной прямой	—	Свободные кружки лежат на вертикальной прямой
67. Правая ветка начинается выше левой	—	Правая ветка начинается ниже левой
68. Конец правой ветки выше, чем левой	—	Конец правой ветки ниже, чем левой
69. «Ягода растет на главной ветке»	—	«Ягода растет на боковой ветке»

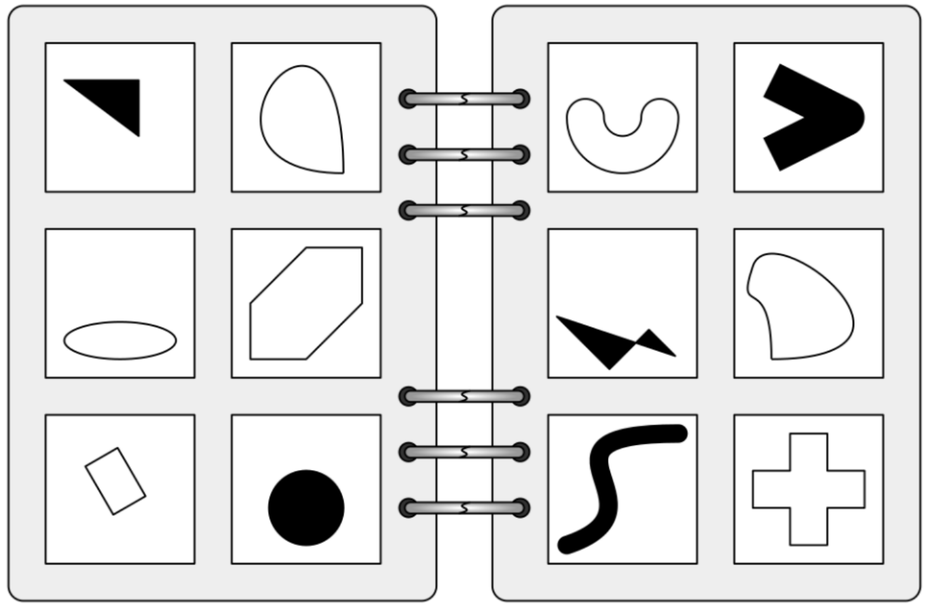
Игры на выявление правил:Zendo, Eleusis, Penultima

Поиски ядра "чистого" интеллекта

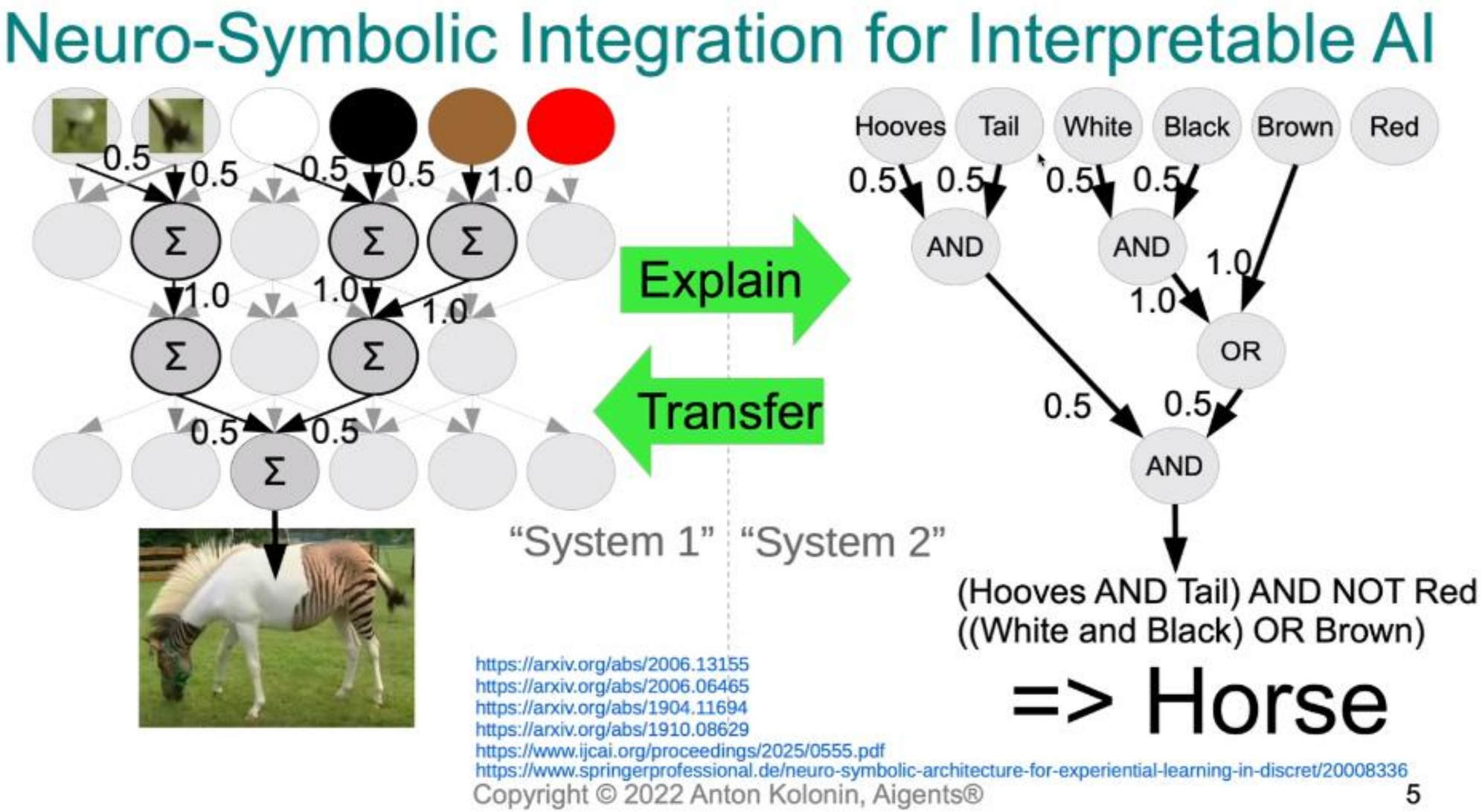
Идея задач Бонгарда состоит в том, чтобы представить два набора относительно простых диаграмм, скажем, A и B.

Все диаграммы из набора A имеют общий фактор или атрибут, который отсутствует во всех диаграммах из набора B. Проблема заключается в том, чтобы найти или убедительно сформулировать общий фактор и были популяризированы благодаря опубликованной в 1979 году книге Дугласа Хофштадтера "Гедель, Эшер, Бах", который сам был автором проблем Бонгарда. По словам Хофштадтера, "***умение решать задачи Бонгарда лежит очень близко к ядру "чистого" интеллекта, если такое вообще существует***".

На более глубоком уровне книга представляет подробное освещение концепций, на которых основываются математика, симметрия и разум.



Популярная идея нейро-символьной интеграции



Задачный подход (task-oriented approach)

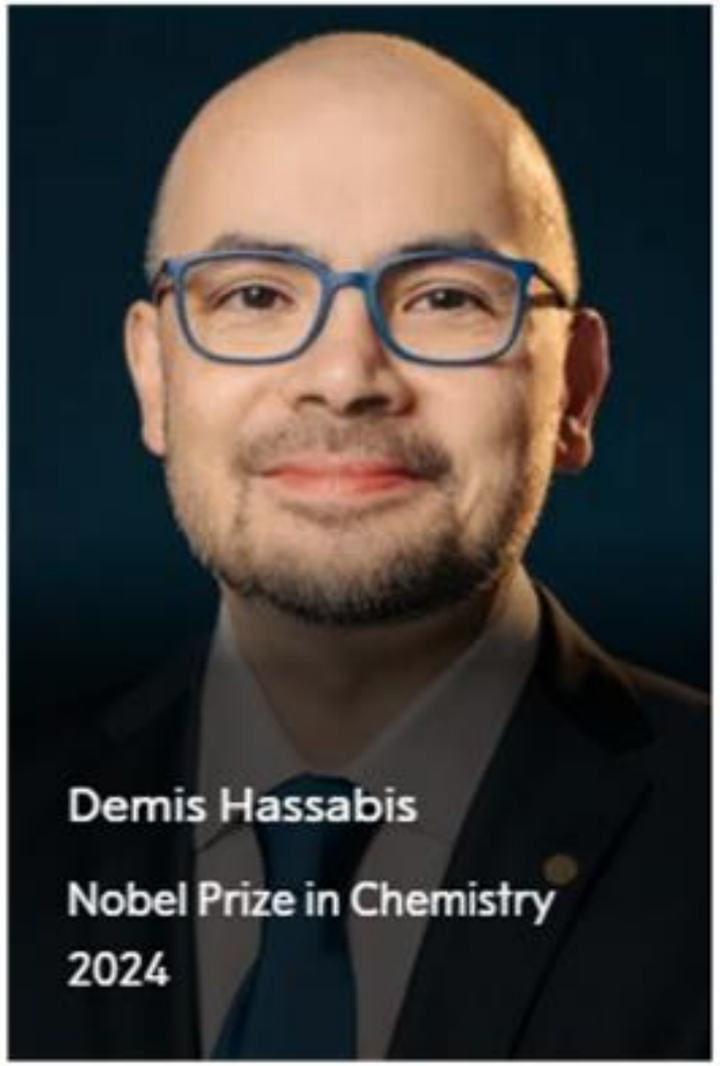
Одна из фундаментальных и наиболее практичных парадигм в разработке ИИ

В основе задачного подхода лежит простая идея: **ИИ создается для решения четко определенной, конкретной задачи (или набора задач)**, а не для имитации общего человеческого интеллекта.

•**Аналогия:** Это не попытка построить универсального робота-человека (как в научной фантастике), а создание «умной» стиральной машины, навигатора, спам-фильтра или ***системы для игры в го***.

•**Контраст:** *Противопоставляется **общему ИИ(AGI)*** — гипотетической системе, способной решать любые интеллектуальные задачи, как человек.

Почему у решающих всё более сложные задачи гибридных систем LLM+RAG (Retrieval-Augmented Generation) **и им подобных** сознание и сильный ИИ не могут возникнуть ***эмерджентно***?



Синергетический подход

Синергетический подход — это **взгляд на мир (и на ИИ) как на иерархию сложных, открытых, нелинейных систем, способных к *самоорганизации и порождению нового*.**

Применительно к ИИ это означает переход от разработки «решателей задач» к исследованию **систем ИИ,** где взаимодействие компонентов воспроизводит известные и порождает новые формы действий по сохранению и развитию порядка.

Синергетический подход не отменяет задачный, а поднимает на мета-уровень, позволяя видеть за деревьями отдельных моделей — лес **проблем искусственного интеллекта**.

Это путь к постановке задач ИИ на системной основе.

Сравнение задачного и синергетического подходов

	Критерий	Задачный подход	Синергетический подход
1	**Фокус**	На **конкретном результате** (решение задачи).	На **процессе, взаимодействии и возникновении** новых свойств.
2	**Метод**	Инженерия, оптимизация, декомпозиция.	Наблюдение, моделирование, выявление параметров порядка, управление через слабые воздействия.
3	**Вопрос**	**«Как решить эту задачу?»**	**«Как здесь возникает порядок? Какие скрытые законы управляют системой?»**
4	**Отношение к сложности**	Стремится **упростить и свести** к измеримой задаче.	Принимает сложность как данность и изучает, **как из нее рождается организация.**
5	**Пример в ИИ**	Обучить модель классифицировать спам.	Исследовать, как ***знания формируют процессы живой самоорганизации.***

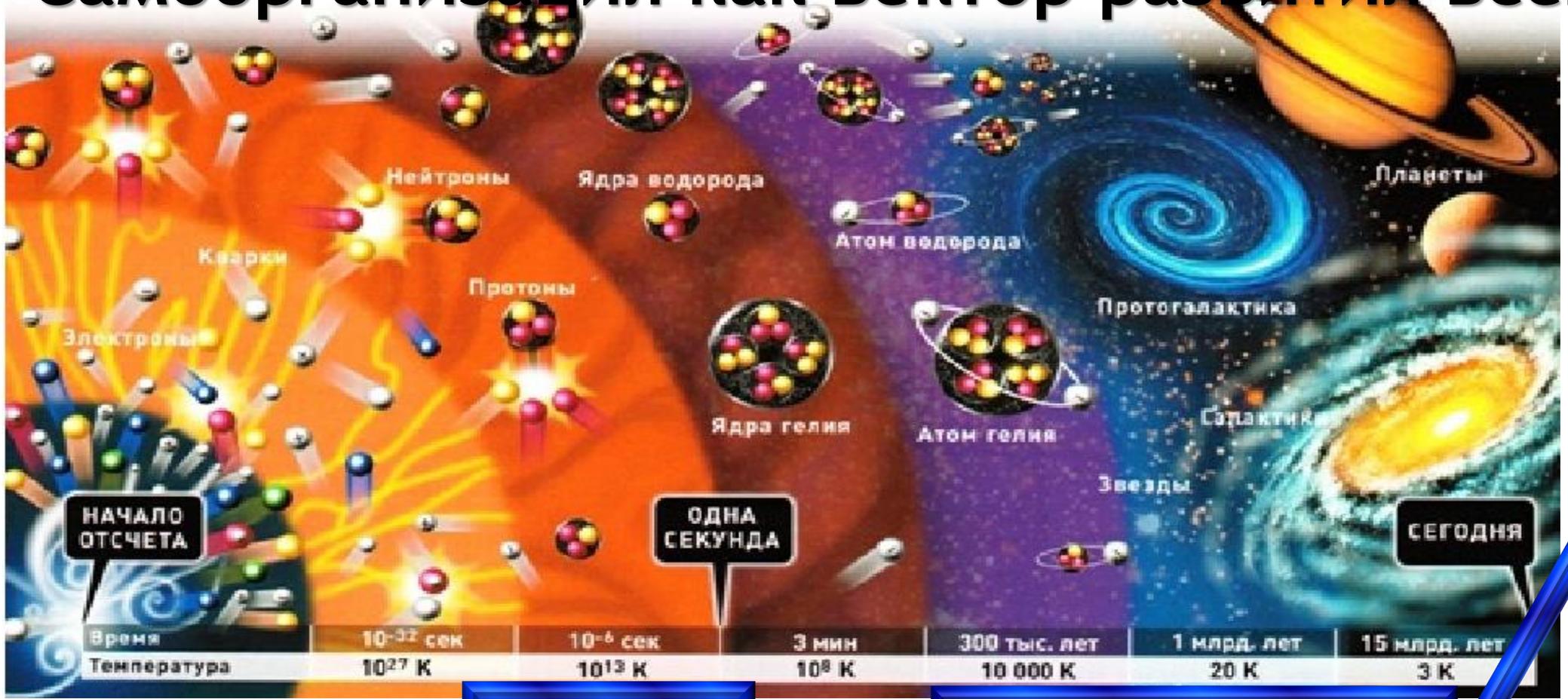
Синергетика – наука о развитии систем

Развитие как рождение нового качества (эмерджентность)
Синергетика изучает свойства, которых не было и не могло быть у её частей.
•Это не количественный рост, а **изменение природы системы**.
•В этом смысле синергетика — наука о **креативности** самой природы, о возникновении нового.

Синергетика изучает, как из хаоса и беспорядка спонтанно, без внешнего «режиссёра», возникает новый порядок, часто более сложный, чем предыдущий.
•Это противоречит взгляду, что порядок может быть создан только сверху или что системы стремятся к хаосу. В **открытых системах**, обменивающихся энергией/информацией 2 закон термодинамики не выполняется.

Синергетика применима к разным областям знаний, но будут рассмотрены ***эволюция жизни и развитие цивилизации***, даны определения таких понятий как ***знания, понимание, интеллект.***

Самоорганизация как вектор развития вселенной



Тепловая смерть

???

сии

Хаос

Самоорганизация

Зарождение

Жизнь

Животные

Цивилизация

Самоорганизация на основе знаний

По Г. Хакену, самоорганизация — это ***процесс упорядочения*** в открытой системе, ***основанный на согласованном взаимодействии*** её элементов.

В ***живых*** системах ***согласование*** между элементами системы и со средой осуществляется с использований ***знаний*** (ДНК, НС, записи).

Цель жизни, как процесса самоорганизации

Свойство естественного отбора:

Выживают (и передают свой код ДНК) ***только наиболее приспособленные***.

Это ***свойство*** принято рассматривать как ***цель*** эволюции.

И ***познания в целом***.

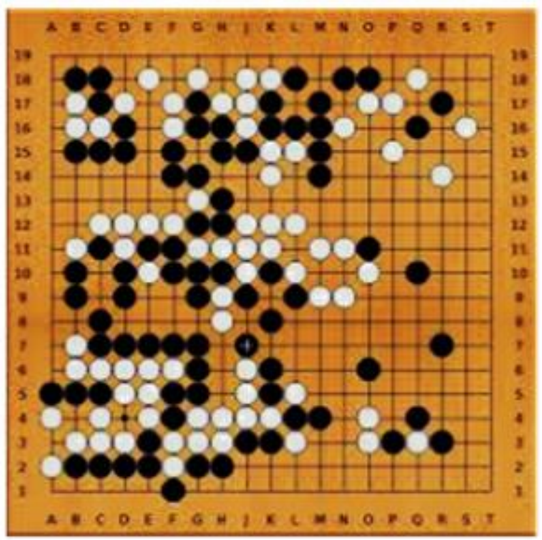
Неограниченность процесса познания

Повторяемость объектов и явлений наука связывает с атомарной структурой вещества и наличием небольшого числа физических взаимодействий. Если мы познаем все законы физики и откроем полный список элементарных частиц – останется ли тогда что ещё изучать?

Синергетика утверждает, что да – количество вариантов комбинаций объектов, способов и последовательностей их взаимодействий *бесконечно. Каждый вариант комбинации может дать синергию или эмерджентность.*

10^{47}

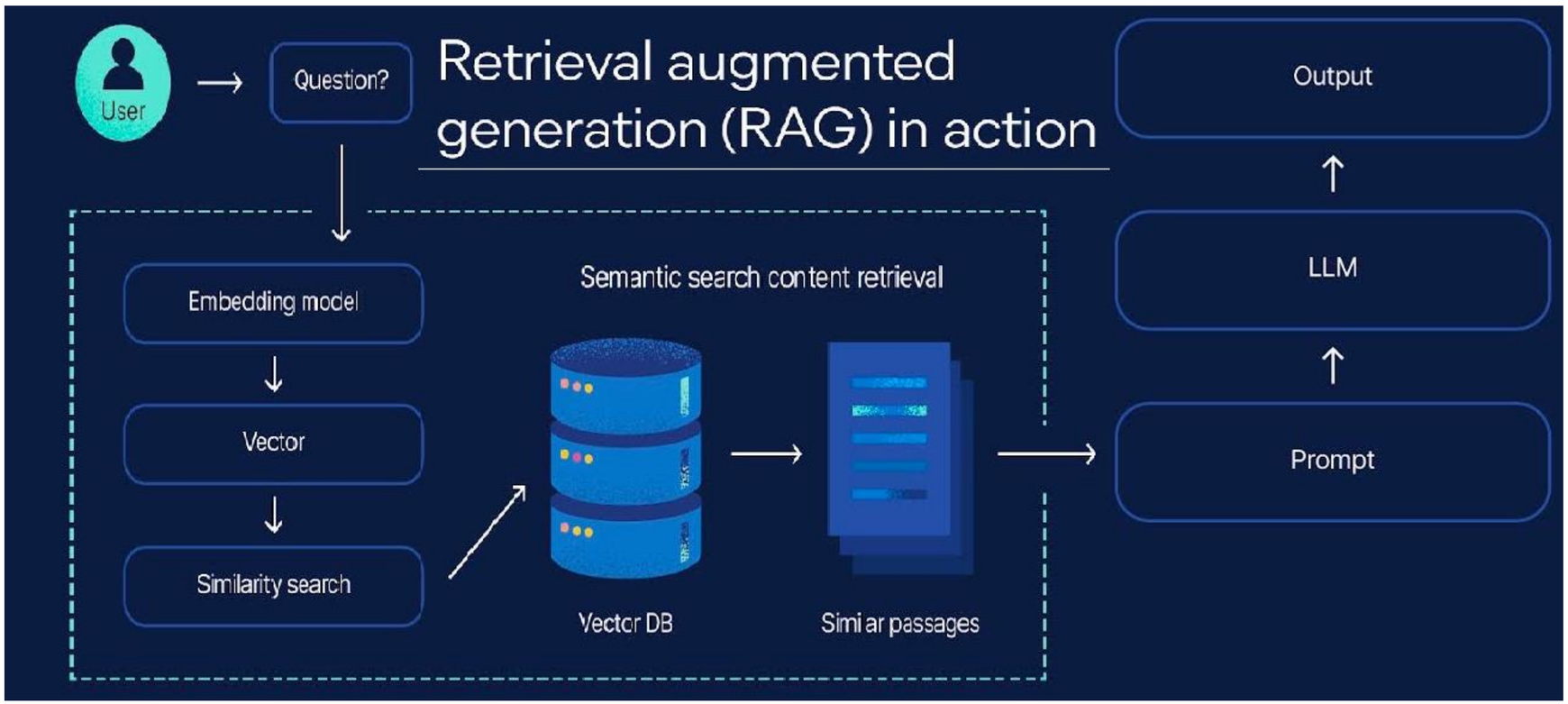
10^{171}



Пример синергии

Синергия в гибридных системах (LLM + RAG + Tools + Agents)
Это **пример синергетического подхода в современной AI-инженерии**:
Элементы: LLM (генератор), векторная БД (память), исполнители инструментов (tools), планировщик (agent loop).
Взаимодействие: Сложный, нелинейный цикл «мышление-поиск-действие-анализ».
Эмерджентное свойство: Способность решать сложные многоэтапные задачи, которые не под силу ни одному из компонентов в отдельности.

LLM без инструментов — болтун без доступа к реальности.
Инструменты без LLM — набор бесполезных программ.
Вместе они порождают ***синергетического агента,*** способного действовать в сложной среде.



Разделение данных, знаний и информации

Описания **— представление свойств объектов и явлений другими материальными носителями**

- ***Данные*** **— описания в любой форме;**
- ***Знания*** **— описания для выполнения действий с данными;**
  (управляют преобразованиями данных);
- ***Информация*** **— описания для выбора знаний**
  (управляет использованием знаний).

Знания **и** ***информация*** **обязательно должны быть согласованы (по форме представления) с системой, которая их использует.**

Откуда появляются знания?

How Much Information is the Machine Given during Learning?

Y. LeCun

- **"Pure" Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

© 2019 IEEE International Solid-State Circuits Conference 1.1: Deep Learning Hardware: Past, Present, & Future 59

При решении задачи главное – это наличие знаний как её решать, какие данные для этого нужны. Шалле справедливо отмечает, что RL требует много времени и мастера ценятся имеющимися умениями, а не перспективными возможностями к обучению.
Но кроме обладания знаниями необходима способность не только применить их к конкретной, ранее не встречавшейся задаче, но и возможность найти новое решение там, где знаний оказалось недостаточно.
И эффективнее ищут знания те агенты, которые обладают большим объёмом знаний в предметной области.

Ян ЛеКун: основа появления знаний – *наблюдение* за свойствами окружающего мира (и использование *генеративных моделей*).

Материальность знаний

Д. Чалмерс

Лёгкая проблема сознания – то, что можно измерить, то что решается традиционными научными методами (то, что относится к способностям или функциям мозга)

Трудная проблема сознания – в непонятности того, почему физический процесс переработки информации даёт начало внутренней жизни.

Знания и информация **— это формы данных, имеющие дополнительные (по отношению к просто данным) свойства:**

- **Согласованность с системой и обрабатываемыми данными;**
- **Полезность для процессов самоорганизации.**

А любые данные могут существовать только на материальных носителях.

Изменение знаний (как и любых других данных) связано с изменением состояния их носителя.

Материалистический взгляд на знания и сознание, как систему управления использованием знаний, делает трудную проблему Чалмерса лёгкой.

Три типа знаний: способы получения и использования

Генетические **знания — наследуются, но медленно меняются.**
Нейросетевые **знания — оперативны, но плохо наследуются.**
Цивилизационные **знания объединяют потенциал обеих форм.**

Генетические знания — сперва рекомбинируется или мутирует генетический код, а затем на его новых знаниях строится изменённый организм и опытным путём (попыткой выжить и оставить потомство) выясняется конкурентоспособность новых генетических знаний.

Нейросетевые знания — накапливаются путём статистической обработки входных сигналов (преимущественно в виде значений весов связей). В сложной среде для накопления статистики необходима её декомпозиция на простые компоненты. Для обеспечения возможности накопления статистики наблюдений.

Цивилизационные знания — накапливаемые и передаваемые в процессе развития цивилизации знания и методы их получения.

Отличия цивилизационных знаний от нейросетевых

Нейросетевые знания накапливаются в мозге человека и животных в результате настройки преобразований в нейросетевых структурах. Аппроксимации преобразований могут быть результатом обучения, но не обеспечивают высокой точности при выполнении сложных преобразований. Новые знания получаются на основе статистического анализа поступающих данных.

Цивилизационные знания тоже могут быть результатом обучения, но они реализуют не просто обработку статистики наблюдений, а методы проектирования и расчёта.
Проектирование (как правило) эффективнее аппроксимации, но используется людьми совместно с нейросетевыми (и генетическими) знаниями.

До появления информационно-вычислительной техники обработка и передача цивилизационных знаний ***была невозможна без участия человека***. Также, как нейросетевые знания у организмов невозможны без генетических знаний

Отличия от представлений Шолле

ИИ может быть построен исключительно на цивилизационных знаниях, но получение и применение этих знаний в сложной среде требует использования аппроксимации и декомпозиции. Что пока что лучше всего выполняется нейросетевыми алгоритмами.

Вместо двух типов описаний (абстракций…) – два типа знаний, плюс необходимость аппроксимации и получения новых знаний (с применением декомпозиции).

Аппроксимация проще написания программ при малой размерности задачи

Аналог теорем Тьюринга и Гёделя

Метод диагонализации как инструмент для доказательства существования невычислимых функций (Тьюринг) и недоказуемых истин (Гёдель).

Можно написать любой алгоритм или проверить истинность любого утверждения, но нельзя написать универсального алгоритма для решения произвольной задачи и установить истинность или ложность для всех утверждений.

***Аналогично*, ИИ может решать всё более широкий круг задач по мере освоения всё большего объёма знаний, но никакого объёма знаний не будет достаточно для решения произвольной задачи**

Как же человеку удается строить целенаправленное поведение в сложной среде?! – *Декомпозицией и аппроксимацией*…

Теоремы о полноте и неполноте

Нарративом или программой можно описать любую ситуацию в локальной части сложной среды и действия агента, адекватные сложившейся ситуации.

Невозможно создать универсальный алгоритм или словесную инструкцию для описания действий в произвольной ситуации в сложной среде.

Необходимо иметь возможность получения новых знаний для выбора действий в произвольных ситуациях – как следствие неограниченности познания.

Необходимость декомпозиции (для ускорения аппроксимации)

Сложность аппроксимации возрастает экспоненциально с ростом размерности аппроксимируемого преобразования.

Декомпозиция (***пространственная, временная и иерархическая***) направлена на снижение размерности аппроксимируемых преобразований.

Наиболее важные математические проблемы развития ИИ – это теоретическое обоснование и совершенствование алгоритмов (нейросетевых) аппроксимации, декомпозиции и синтеза прогнозов

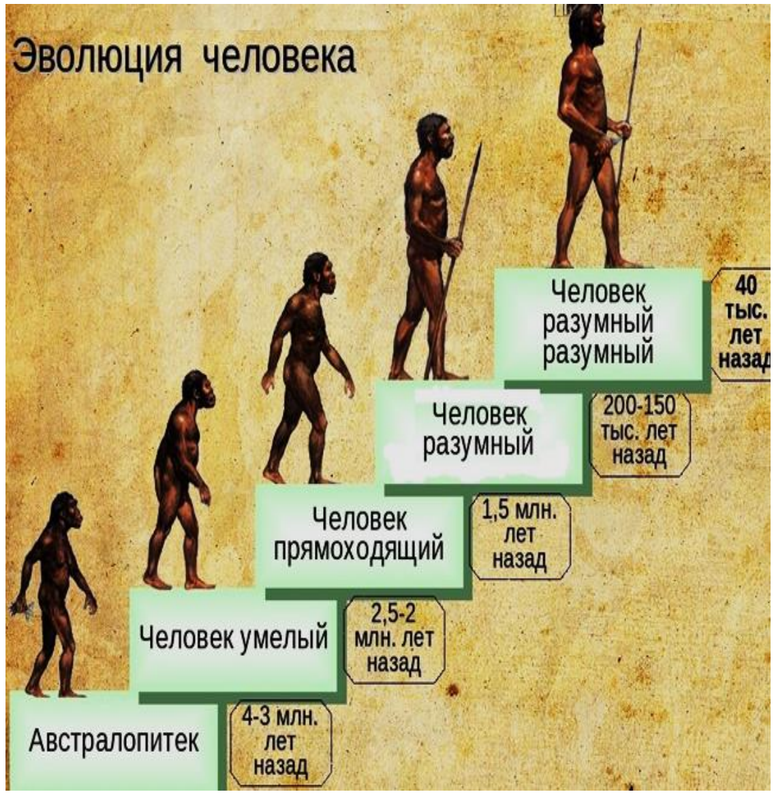
Мир сказок философов, психологов и др.

Самоорганизация, *основанная на знаниях – понятная, но* ***чисто материалистическая идея!***

А мы воспитаны на идеях диамата ***– должно быть ещё и что-то нематериальное!***

Концепции развития ИИ пишут те, кто соответствует общепринятым представлениям: дескать, *мы не понимаем (и, возможно, никогда не поймём…) как работает наш мозг (сознание, мышление, интуиция и пр.).*

Разделим (определим) понятия



Появление *новых свойств* по мере развития и *увеличения сложности* нервной системы:

Условные-рефлексы	Поведение на основе статистически выявленных корреляций	Низшие животные, ИИ
Интуиция	Действия с учётом всей сложности сложившейся ситуации	Низшие животные, ИИ
Понимание	Соотнесение с текущим состоянием моделей объектов и явлений	Высшие животные, ИИ
Мышление	Сравнительное моделирование вариантов развития событий	Высшие животные, ИИ
Сознание	Механизм управления интуицией и мышлением	Высшие животные, ИИ
Разум	Способность к развитию интеллекта	Человек, СИИ
Интеллект	Способность к освоению и выборочному применению и развитию цивилизационных знаний.	Человек, СИИ

Интуицию, понимание, мышление **- биологи находят у высших животных. Но в биологической таксономии видов *разумный* - только человек, как вид.**

Что такое интеллект?

Способность к освоению и выборочному применению и развитию *цивилизационных знаний*.

А *разум* — способность к развитию интеллекта у психически здоровых людей.

Интеллект человек может развить только путём *получением воспитания и образования* в цивилизованном обществе

Сила интеллекта человека связана со степенью успешности применения имеющихся *цивилизационных знаний* и получения новых знаний при решении *цивилизационных задач*.

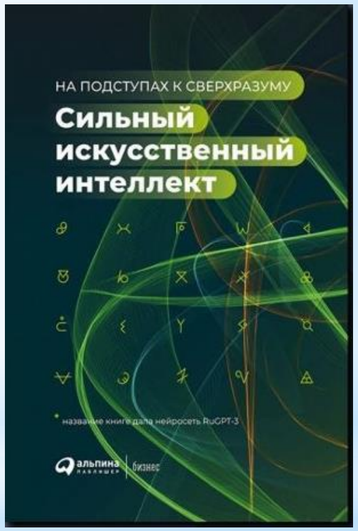
Цивилизационные знания — знания, полученные в процессе развития человеческой цивилизации как *целого*, а не цивилизационных ветвей (как у Н.Я. Данилевского или Освальда Шпенглера).

Определения AGI

Создание AGI является основной темой *исследований* для OpenAI, DeepMind и Anthropic PBC.

↑ Beijing Institute for General Artificial Intelligence

Эмблема AGIRussia →



Главные идеи:

Деньги, превосходство, рациональность, саморазвитие и знания (как основа живой самоорганизации)

1. *AGI — прибыль в размере не менее 100 миллиардов долларов (Microsoft и OpenAI).*
2. *AGI — превосходит человека в широком спектре когнитивных задач (Википедия).*
3. *AGI — использует ограниченные ресурсы для задач в сложной среде (AGIRussia).*
4. *AGI — улучшающий структуру своей управляющей системы (Макс Тегмарк).*
5. *AGI — осваивающей имеющиеся и получающий новые знания о компонентах сложной среды и использующий их для задач развития цивилизации.*

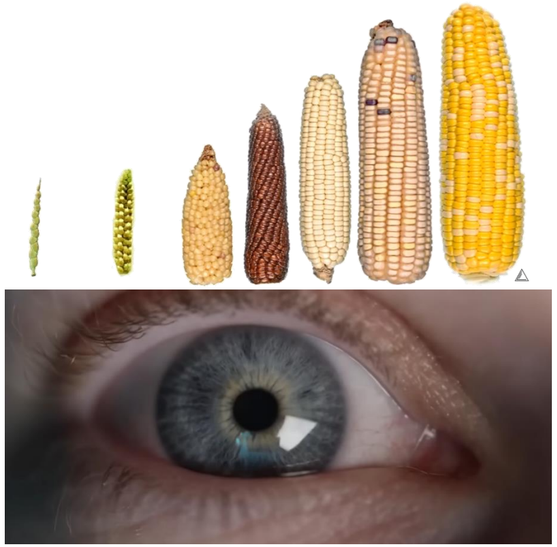
Сильный интеллект - по Тегмарку

Можно ли определить сильный интеллект без использования разума, сознания, эмоций и пр. «человеческих» понятий?

Уровни жизни по М. Тегмарку	Can survive & replicate	Can design its software	Can design its hardware
Life 1.0 (Simply Biological)	✓	✗	✗
Life 2.0 (Cultural)	✓	✓	✗
Life 3.0 (Technological)	✓	✓	✓

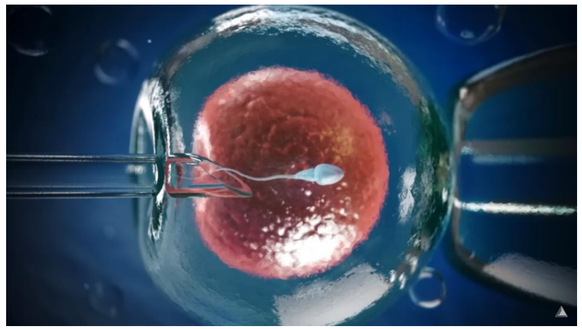
Эволюция по М. Тегмарку:
разные этапы → различные возможности адаптации
Эволюция как приобретение новых знаний о самоорганизации.
Ускорение адаптации по мере развития эволюции.

Углеродная ветвь

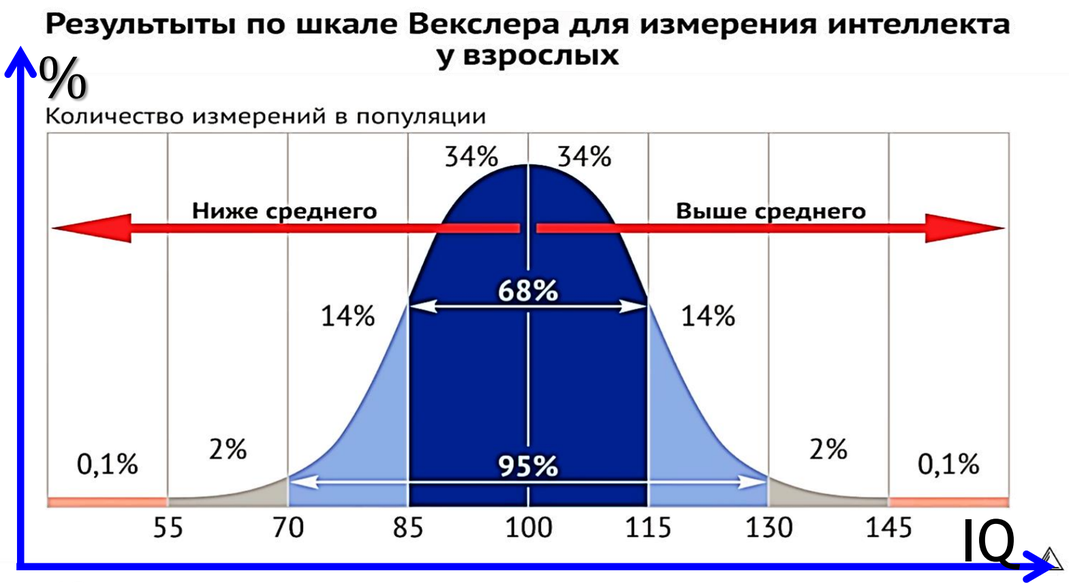


Смещение IQ на 10, 100 и более пунктов

Теренс Чи-Шен Тао, IQ=230, американский математик австралийского происхождения с китайскими корнями.



ИКСИ (интрацитоплазматическая инъекция сперматозоида)



Экстракорпоральное оплодотворение + (плюс) преимплантационная генетическая диагностика = (равно) ускоренный генетический отбор. Это ещё без CRISPR-Cas технологий редактирования генов…

Конец эпохи масштабирования

Илья **Суцкевер,** Safe Superintelligence Inc., or **SSI** Inc., is an American artificial intelligence company founded by Ilya **Sutskever**,

До 2020 г. – эпоха исследований;
2020-2025 – эпоха масштабирования;
Дальше – опять эпоха исследований.

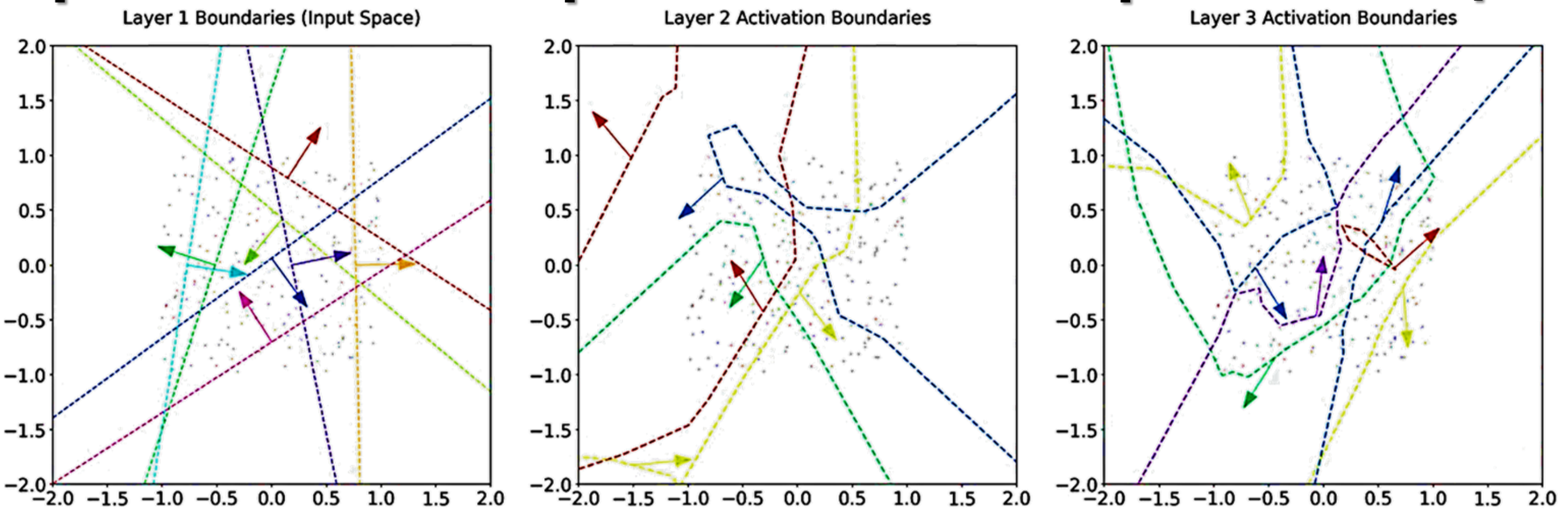
Фундаментально, что нейросетевые алгоритмы обобщают значительно медленнее, чем живой человек.

Ловкость роботов настолько же проигрывает человеку, как человек – белке.

Надо строить ИИ, который будет заботиться не о человеке, а о разумной жизни вообще.

Сэм Альтман: удача с законами масштабирования казалась чудесной и не верилось, что она повторится.

Проблемы нейросетевой аппроксимации



Neuron activation boundaries after Mapping-based training

Нейросетевая аппроксимация хорошо работает для задач распознавания (разделяющих поверхностей) и плохо – для аппроксимации функций. Этому есть ряд причин:

- **Неравномерное участие параметров**
- **Непрерывность настройки**
- **Недостаточность числа параметров**
- **Формирование кривизны**
- **Случайность ортогонализации**
- **Вероятностная декомпозиция**

Цели: локальные и глобальные

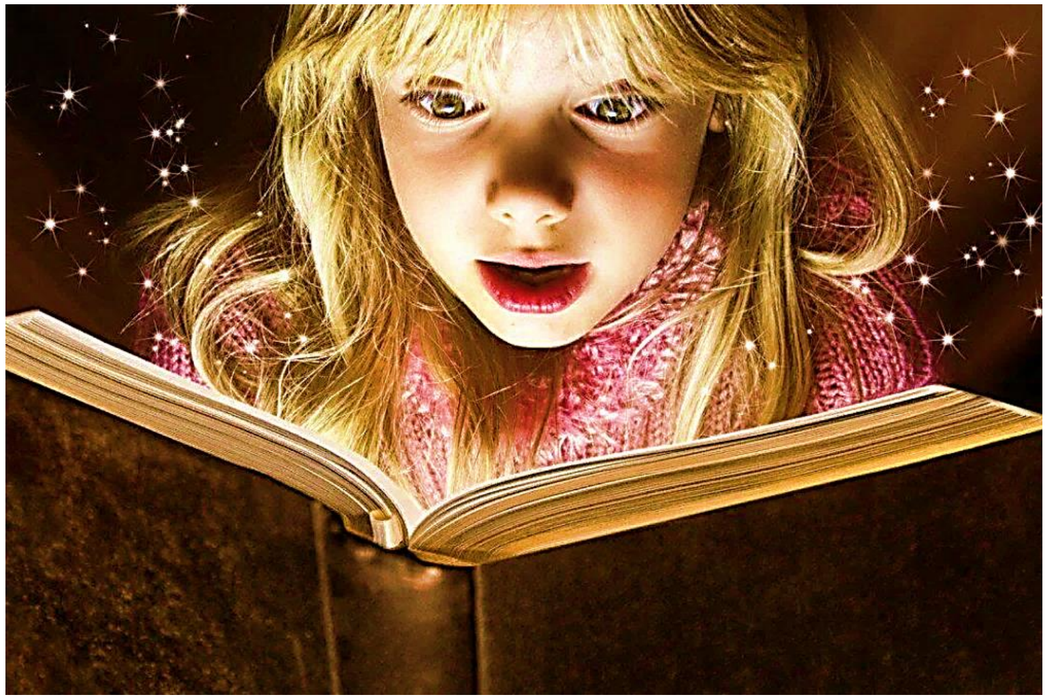
AGI как система, способная ставить цели.

Цели человека формируются на основе знаний, содержащихся в ДНК, соотнесёнными НС с реалиями сложной среды и с использованием цивилизационных знаний.

Аналогично ИИ может на основе знаний формировать цели, как локальные, так и глобальные.

Основной «аппаратно» заложенной целью должно стать развитие процессов самоорганизации. А подцелями – сохранение работоспособности, обеспечение энергией и другие.

Зло от книг, интернета, ИИ

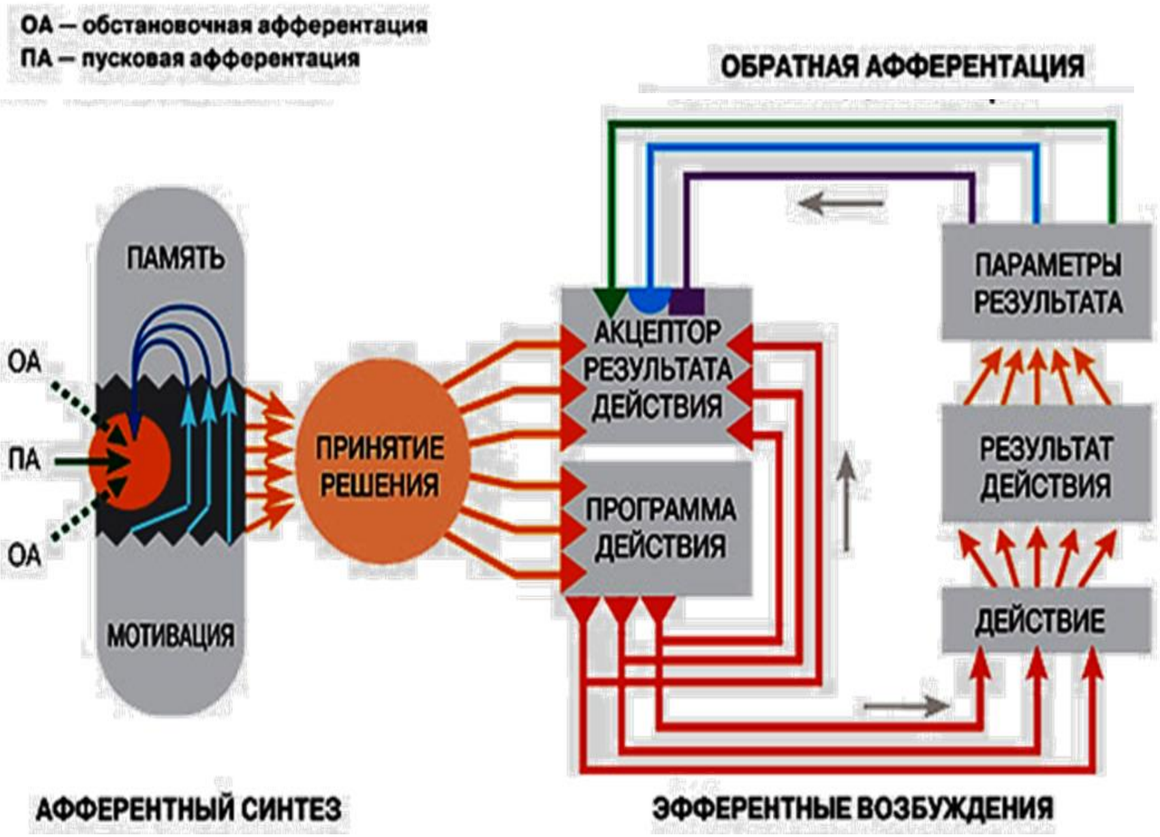


- ***Религия,***
- ***магия,***
- ***философия,***
- ***алхимия,***
- ***наука, книги,***
- ***радио, кино,***
- ***телевидение,***
- ***компьютеры,***
- ***интернет***

Книги снижают нагрузку на память, что позволяет людям с ограниченными возможностями памяти (а таких – большинство!)решать вопросы, которые без книг были бы недоступны.

Включая занятия наукой.

Синтез идей Анохина, нейросетей и цивилизационных знаний

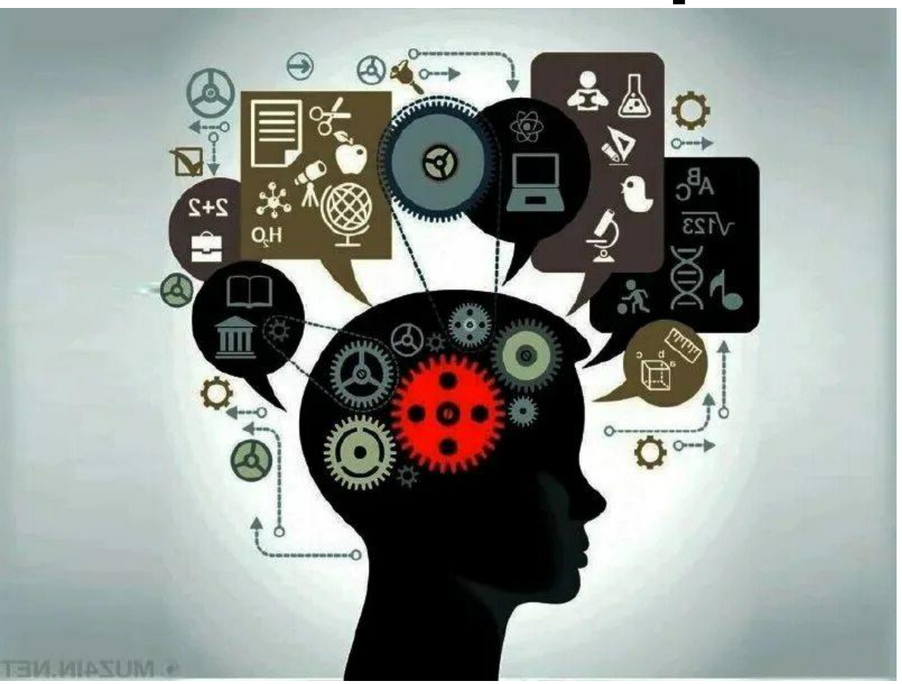


Принципиальная схема функциональной системы по П. К. Анохину. Каждый из её компонентов многократно закольцован благодаря наличию прямых и обратных связей.

Наиболее перспективный путь к AGI — реализация ***иерархического*** «акцептора действия» (по П.К. Анохину) средствами нейросетей, где непрерывное управление целями на основе обратной связи выполняет функции сознания.

Современные успехи нейросетевого ИИ позволяют практически реализовать блоки функциональной системы П.К. Анохина, на основе накопленных знаний генерировать образы «результатов действия» и отслеживать процесс достижения целей.

Центральная задача построения AGI — сбалансированная работа со знаниями



Нет и не может быть *единого универсального* решения проблемы построения AGI.

Системный подход состоит в нахождении баланса применения наиболее эффективных методов работы со знаниями.

Путь к AGI — это не только совершенствование генерации контента или средств обработки данных на основе имеющихся знаний, а получение новых свойств за счёт совместного использования (синергии) ряда способностей, таких как:

- ***Автономное получение новых знаний (путём декомпозиции и аппроксимации)***
- ***Построение локальных и глобальных целей действий***
- ***Использование и совершенствование цивилизационных знаний***
- ***И других***

Теоретическая лёгкость прогресса

Когнитариат — это наемные работники, продающие интеллектуальный труд. С развитием информационных корпораций самостоятельный труд ***когнитариев*** становится неконкурентоспособным, вынуждая их наниматься на «когнитивные фабрики».

Это ведет к ***прекаризации*** — переходу к неустойчивым трудовым отношениям, лишающим работника стабильного дохода и благополучия. Класс людей в таком состоянии называется ***прекариатом***.

И к предложениям выплачивать БОД – безусловный обязательный доход.

Спасибо за внимание!

Владимир Смолин,

smolin@keldysh.ru

Институт прикладной математики имени М. В. Келдыша РАН, Миусская пл., 4, Москва, 125047, Россия

Семинар AGIRussia, 11 декабря 2025 г. (сетевой, модератор Колонин А.Г.)