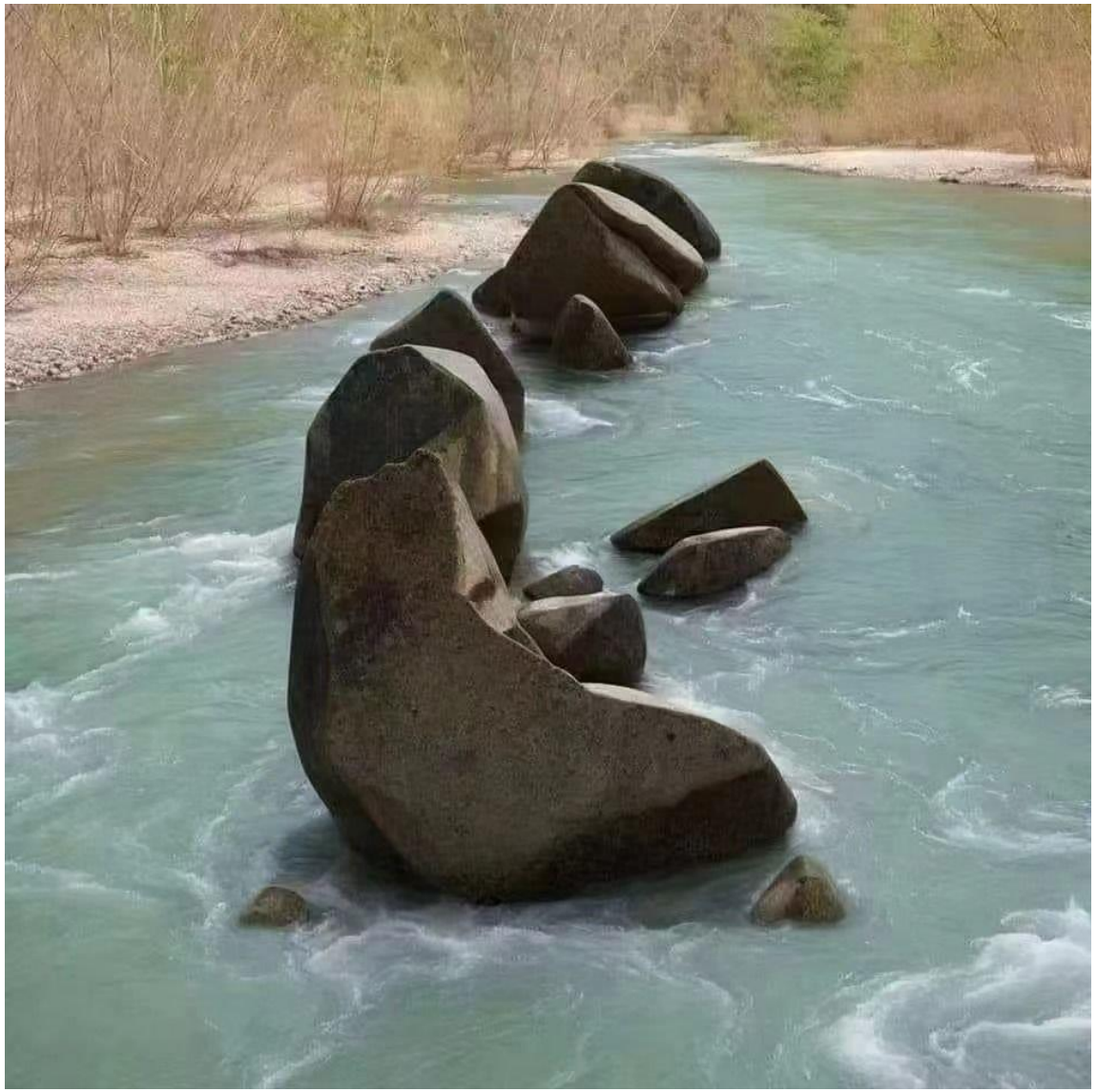
ПОЛИТЕХ
Санкт-Петербургский политехнический университет Петра Великого

Курс: управление научными проектами

Занятие 4

научные проекты РАН
ПО ПРОБЛЕМАМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И КОГНИТИВНЫХ ИССЛЕДОВАНИЙ

16 октября
2024



ошибка многих исследователей «сознания» состоит в их желании редуцировать СОЗНАНИЕ до каких-то измеримых физических процессов, чтобы найти СОЗНАНИЮ соответствующий коррелят в материальном мире.

ПРЕЗИДИУМ
РОССИЙСКОЙ АКАДЕМИИ НАУК
НАУЧНЫЙ СОВЕТ ПО МЕТОДОЛОГИИ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
И КОГНИТИВНЫХ ИССЛЕДОВАНИЙ
НСМИИ РАН
109240, г. Москва, ул. Гончарная, д. 12, стр. 1, www.scmai.ru, mail@scmai.ru

МЕЖДУНАРОДНАЯ МЕЖДИСЦИПЛИНАРНАЯ КОНФЕРЕНЦИЯ
«ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СОЗНАНИЕ

ПРОГРАММА КОНФЕРЕНЦИИ

23 – 24 октября 2024 г.,
,
Ленинский просп., дом 14,
Президиум РАН

Сайт:
https://phAI.info

г. Москва

ПЛЕНАРНОЕ ЗАСЕДАНИЕ

Макаров Валерий Леонидович, академик РАН, доктор физико-математических наук, профессор, заместитель председателя НСМИИ РАН, научный руководитель Центрального экономико-математического института РАН, г. Москва

Движение искусственного сознания и интеллекта от индивидуума к группам

Ушаков Дмитрий Викторович, академик РАН, доктор психологических наук, профессор, директор Института психологии РАН, член НСМИИ РАН, г. Москва

Естественный и искусственный интеллект: сходства и различия

Анохин Константин Владимирович, академик РАН, доктор медицинских наук, профессор, директор Института перспективных исследований мозга МГУ имени М. В. Ломоносова, член бюро НСМИИ РАН, г. Москва

Грани сознания в естественных и искусственных системах

Гончаров Сергей Савостьянович, академик РАН, доктор физико-математических наук, профессор, заведующий лабораторией теории вычислимости и прикладной логики Института математики имени С.Л. Соболева, г. Новосибирск

Задачный подход в математике и в приложениях к искусственному интеллекту

ПЛЕНАРНОЕ ЗАСЕДАНИЕ

Финн Виктор Константинович, доктор технических наук, профессор, член НСМИИ РАН, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, г. Москва

Интеллект и продуктивное сознание с точки зрения искусственного интеллекта

Целищев Виталий Валентинович, доктор философских наук, профессор, научный руководитель Института философии и права Сибирского отделения РАН, г. Новосибирск

Человеческая логика Гёделя

Михайлов Игорь Феликсович, доктор философских наук, ведущий научный сотрудник сектора методологии междисциплинарных исследований человека Института философии РАН, г. Москва

Рекурсивные алгоритмы сознания

Цели исследований РАН

- Рассмотрение феномена сознания как ОБЪЕКТА исследований в области технологий искусственного интеллекта с целью
  - создания принципиально новых технических устройств, воплощающих возможности работы информационно открытых (самообучающихся) систем
  - и самоприменимых вычислителей, работающих в режиме накопления данных о своем функционировании с целью повышения эффективности процессов вычислений

Мир, в котором «работает» и физическая сила и память, радикально другой

- Мир рассматриваемый как гиперсеть сложных систем является носителем процессов, динамика которых нелинейна и спорадически (от случая к случаю) переходит в стадию хаоса….то есть
- …образует поток бифуркаций и непрогнозируемых изменений состояний, приводящих к ситуации «радикальной неопределенности» в которой:
  - факты неточны, критерии спорны,
  - ошибки «дорого стоят», но все решения должны приниматься быстро.

В такой обстановки единственного обоснованного верного решения нет, а выбор постоянно надо делать на границе зоны критических рисков (устойчивости).

Неопределенность , порождаемая причинной моделью генерации воспринимаемых данных

Существовать – значит быть воспринимаемым

Дж. Беркли (17 в.)

- Мозг можно представить как «само-применимую» динамическую систему, которая используя ресурсы своей памяти постоянно обновляет представления/предположения/модели как о:
  - внешнем мире

  так и
  - «самом себе»

на основе данных, хранящихся в памяти (условный вывод) и непрерывно поступающих от сети сенсоров.

- Соединение принципа свободной (информационной) энергии, заключенной в любой структуре, и условного (байесовский) вывода на основе данных, поступающих от сети сенсоров, позволяют формализовать функционирование мозга как непрерывный процесс минимизации информационной неопределенности хранящихся и воспринимаемых данных об окружающей среде

Статистическая vs «радикальная неопределенность»: реификация в результаты вычислений

Существовать – значит быть мыслимо вычислимым

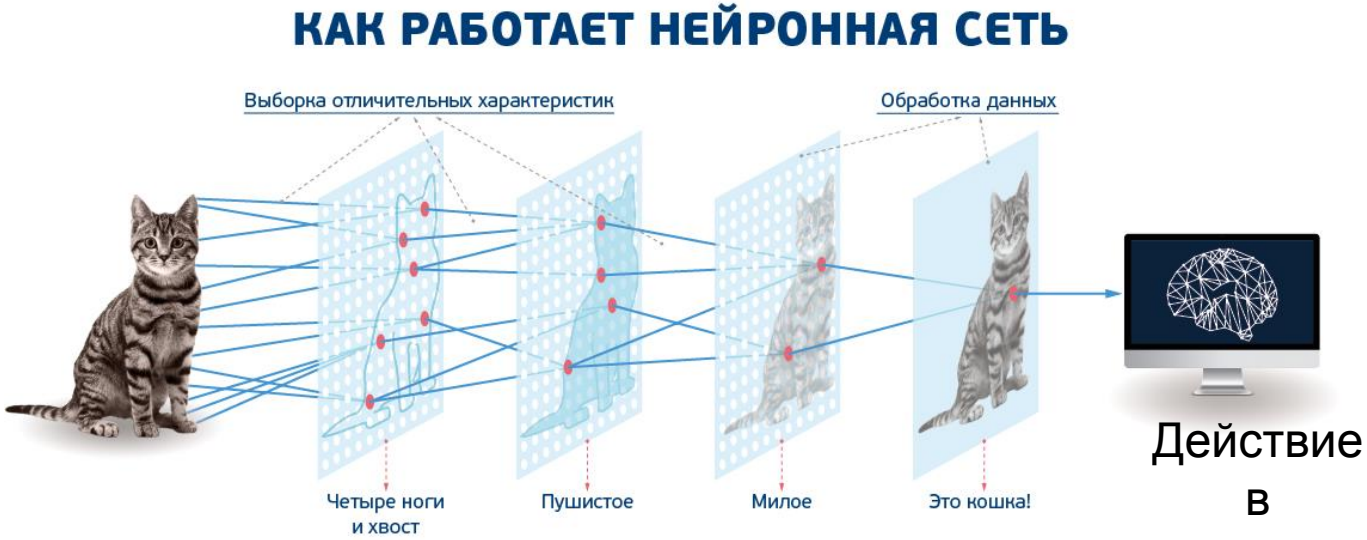
XXX (21 в.)

- классическая **парадигма рациональности современной науки**, а именно: существует то, что воспринимаемо, не предлагает эффективных механизмов для «**вычисления**» решений в ситуациях «радикальной неопределенности».
- В 90-х годах 20 века Дж. Хинтон, лауреат Нобелевской премии по физике 2024 года, предложил считать, что «перцептивная система мозга представляет собой механизм статистического **вывода** вероятных причин сенсорного **входа**.

  Взаимосогласованные модели Дж. Хинтона:
  - модель распознавания (перцепции) описывают процесс «вывода вероятностного распределения гиперпараметров (причин) наблюдаемого сенсорного входа»,
  - модель порождения используется для «обучения модели распознавания» и самообучения
- Итак, понятие введенное в 19 веке Г. фон Гельмгольцем о «бессознательном выводе» в 21 веке Дж. Хинтон «наполнил» статистическим смыслом.
- Мы покажем как этот смысл **реифицируется в результаты вычислений**

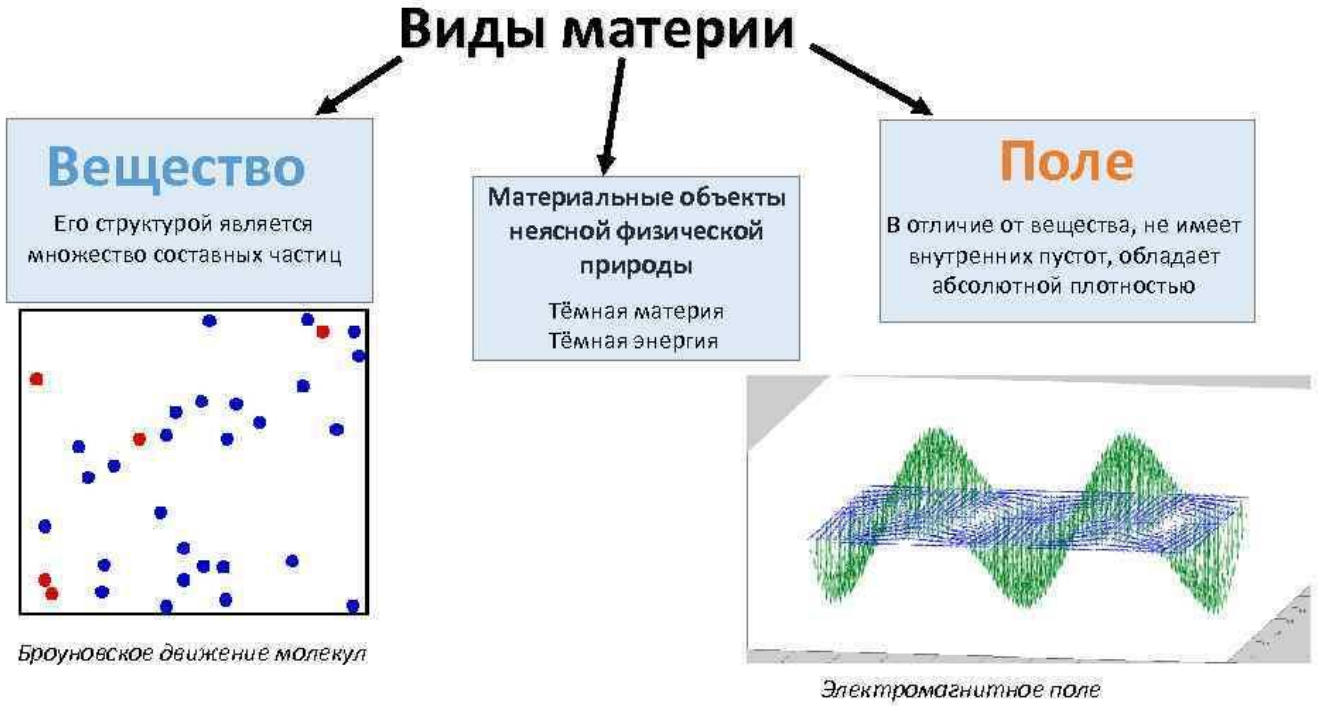
ПРИНЦИП «шахматной доски» - субъектность элементов

возможность разделения окружающего «пространства-время» на классы эквивалентности, т. н. фактор-множества (например, черные/белые клетки), на которых определено множество операций - является основой научного подхода. Это позволяет выделить подмножества состояний, которые можно рассматривать как пределы последовательности элементов. Эти элементы образуют базис научной модели, замкнутый относительно фундаментальных (энергетических, топологических и др.) инвариантов. Альтернатива такому подходу– например «ИИ реальность, инвариантов.

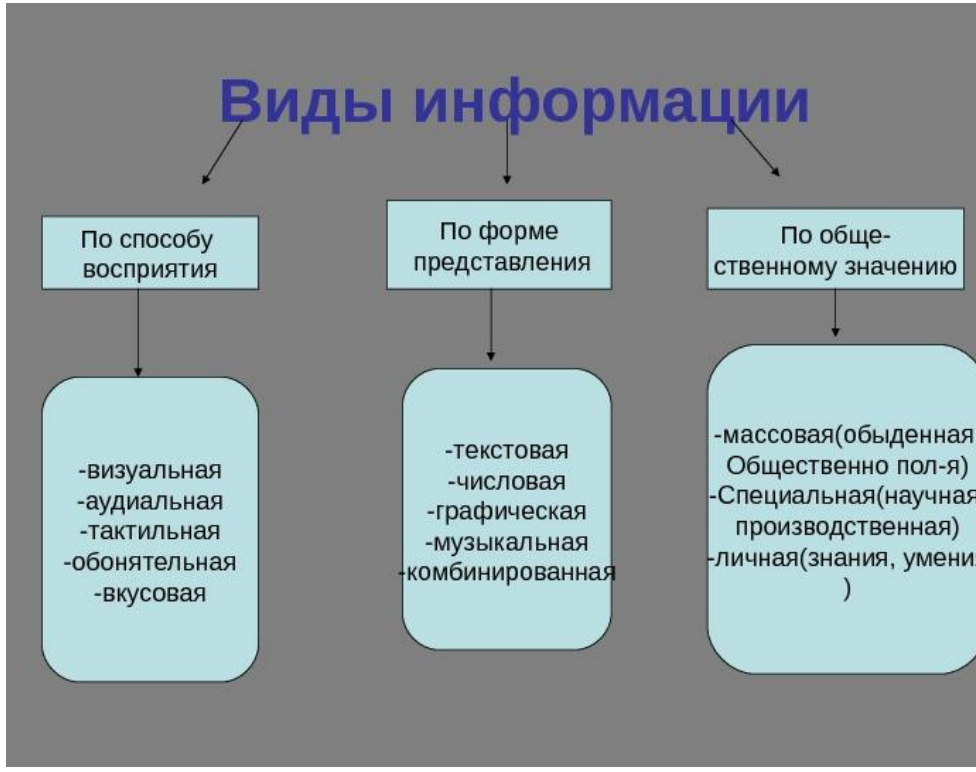


объект понятия восприятие Действие в контексте ситуации

Два цвета на «шахматной доске» Природы…. «материя-информация»



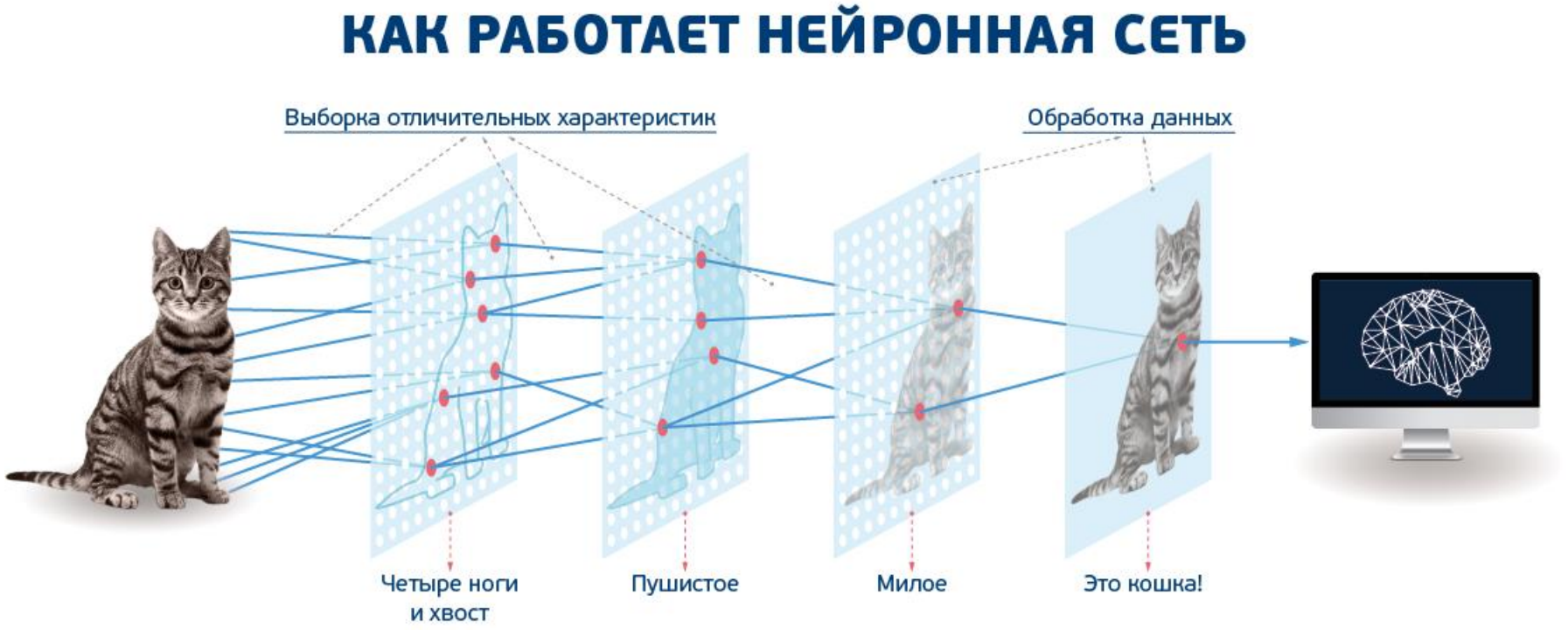
Можно не только переставлять фигуры по правилам игры, но и «переставлять правила»



Математическая основа игры на новой шахматной доске: **теория информации и категорий**

Физическое vs информационное

Физическое – значит локальное контактное воздействие с передачей **энергии**
Информационное – значит воздействие на основе передачи **сообщения**



Физический объект

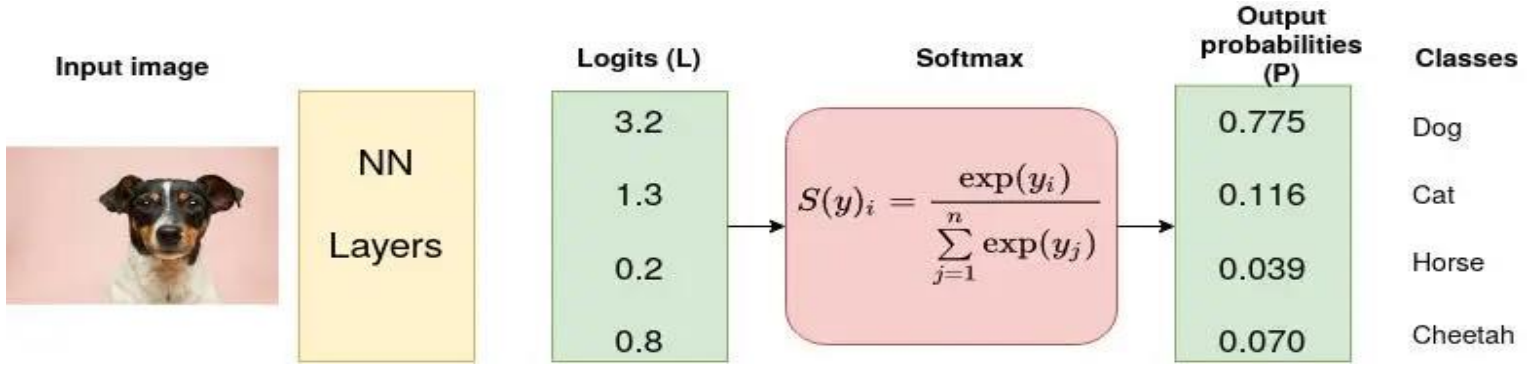
Отложенное действие в контексте ситуации

Функция «мягкого разделения» в задачах машинного обучения

Предположим, что у нас есть непрерывная переменная **x**, которую мы хотим разбить на **N + 1** интервалов. Необходимо задать **n** точек, которые являются «обучаемыми переменными». Обозначим эти точки **[β1, β2,…, βn]** как монотонно возрастающие, то есть **β1 < β2 < . . . < βn**. Во время обучения величина **β** может быть изменена с помощью функции влияния модели классификации для каждого из 4 классов.

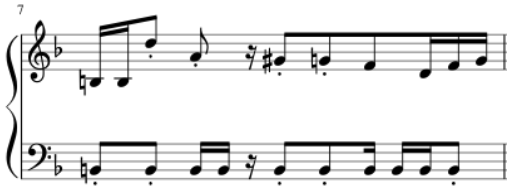
[W] – константа вектор, b = [0,b1, b2, …bn], τ – фактор напряжённости. Функция Softmax, преобразует компонент вектора в «вероятности». Эти «вероятности» являются прогнозами

$$T = T_{w,b,\tau}(z) = \mathrm{Softmax}((wx + b)/\tau)$$



Два когнитивных «цвета» : «код – процесс»

Понятие «информационно-воздействие» можно определить формально как уменьшение количества равновероятных исходов. Большинство определений не конструктивно до тех пор, но возможности точно определить, что является **информационной сущностью** объекта : **код**, субъект воспринимающий код , инструмент реализующий код , физический процесс (сигнал), как **воплощенный код** .



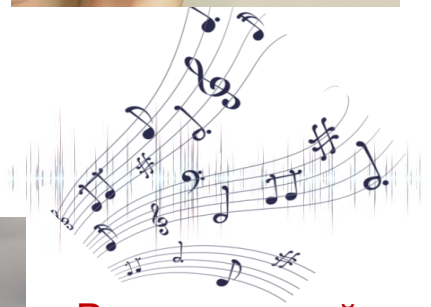
Мыслимый Код
физического процесса

Интеллектуальный субъект, «понимающий» код

Инструмент «расшифровки» кода

Инструмент записи кода

восприятие музыки



Воплощенный код: физический «процесс»

Пример: Музыка – это то, что присуще определенным наборам звуков . Ни каждый набор частот колебаний струн инструмента можно связать с понятием музыка. Каждому музыкальному произведению сопоставляется мыслимый нотный код, некий дискриптор – название и автор.
Аналогично – научный проект : гармония частей, образующих некоторое целое, имеющее авторское имя

Базовая гипотеза

Гипотеза : **научный проект это код для исполнения** в гибридной «вычислительной» среде, объединяющей «нейроструктуры мозга с различными техническими средствами обработки и хранения информации.

Следствие :

Результат научного проектирования - реификация мыслимого кода, полученная с использованием «встраиваемого ПО» гибридного (экзо-интеллектуального) нейрокомпьютера мозга человека.

Вопросы: кто может «написать» и как встроить» это «ПО» в такую гибридную «вычислительную» среду ?

Выводы: Фундаментальная ограниченность научных знаний: проблема неполноты формальных моделей

И. Кант: возможности познавательной деятельности в отрыве от знаний, получаемых эмпирическим путем, **ограничены**

Первая Теорема Геделя:
если формальная арифметика непротиворечива, то в ней существует невыводимая и неопровержимая формула (как ее найти ?)

Вторая Теорема Геделя:
если формальная арифметика непротиворечива, то в ней невыводима некоторая формула, содержательно утверждающая непротиворечивость этой арифметики

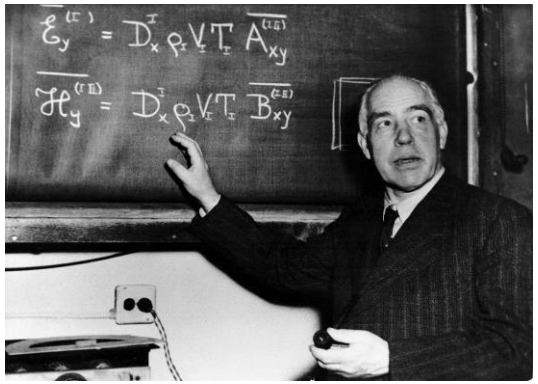
Вывод:

Описание свойств природы как целостной системы в форме точно сформулированного научного проекта невозможна. Что делать ?

Легко можно сформулировать вопросы на которые **не возможно ответить** так как они превосходят возможности человеческого разума

Истина в неполноте - физика «дополнительности»

Принцип дополнительности (также принцип комплементарности) — один из важнейших методологических и эвристических принципов науки, сформулированный в 1927 году Нильсом Бором



Согласно этому принципу, для

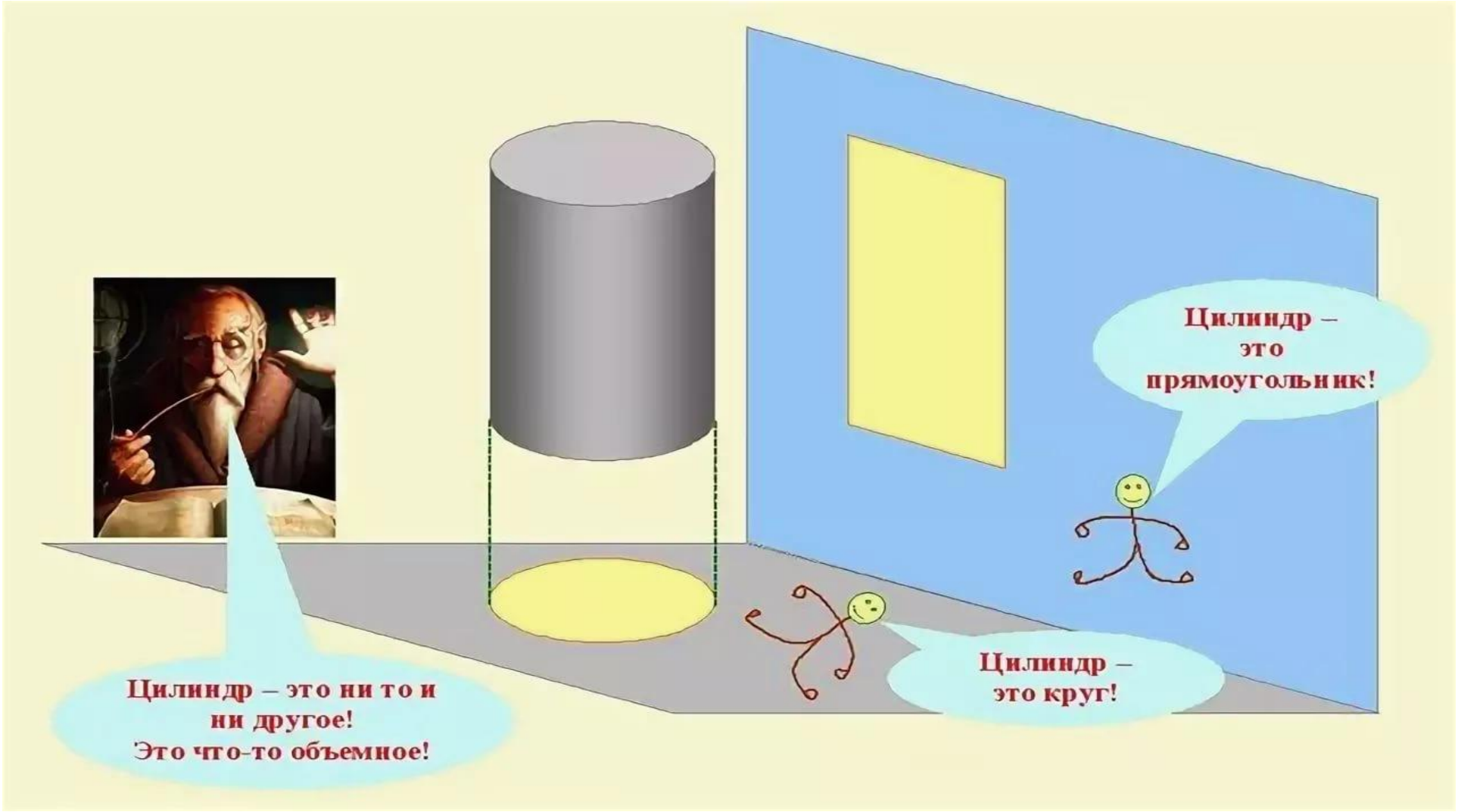
полного описания сложных явлений необходимо применять два взаимоисключающих («дополнительных») набора классических понятий, совокупность которых даёт исчерпывающую информацию об этих явлениях как о целостных.

физическая картина явления и его математическое описание дополнительны.

Суть принципа – использовать взаимоисключающие классы понятий, каждый из которых применим в особых условиях, но их совокупность позволяет воспроизведение целостности данных объектов.

«Истина в неполноте ?!»

К. Гедель



итак.

Обработка информации в режиме «мягких вычислений» , а не не только по одному алгоритму, Это требует создания принципиально новых технических устройств, воплощающих возможности работы информационно-открытых систем.

Тогда:

- Интеллектуализация вычислителя на основе обучения приводит к изменениям в его физической структуре и программном обеспечении
- Обученная «Машина» обретет способность моделировать процессы мышления, если реализует свойства «процессора управляемого входными данными» под контролем методов интеллектуальной регуляризации – объяснения результатов

что надо учесть

При научном проектировании надо иметь в виду «принцип «хрупкости» хорошего, сформулированный ак. В. И. Арнольдом…

- *для системы, принадлежащей части границы устойчивости, при малом изменении параметров более вероятно попадание в область неустойчивости, чем в область устойчивости.*

согласно принципу :

всё «хорошее» (например, устойчивость системы) свойство более «хрупко», чем все «плохое». Все «хорошие» объекты **удовлетворяют нескольким требованиям одновременно**, «плохим» же может считаться объект, обладающий хотя бы одним из ряда недостатков (т.н. принцип Анны Карениной).

Приложение

Принцип «Анны Карениной»:

- одновременное сочетание всех необходимых факторов для получения хорошего решения некоторой проблемы является исключением.

А. Н. Толстой : *«Все хорошо адаптированные (счастливые) системы (семьи) похожи (одинаковы), все неприспособленные системы не справляются с адаптацией (не счастливы), но каждая по-своему» (несчастны посвоемому)*

Поведенческая версия принципа:

В благополучные периоды существования все системы ведут себя одинаково, а в моменты кризиса их поведения могжет радикально отличаться.

Вероятностная версия принципа:

Число ситуаций, когда что-то может пойти не так, гораздо больше чем число ситуаций, когда всё идёт как надо.

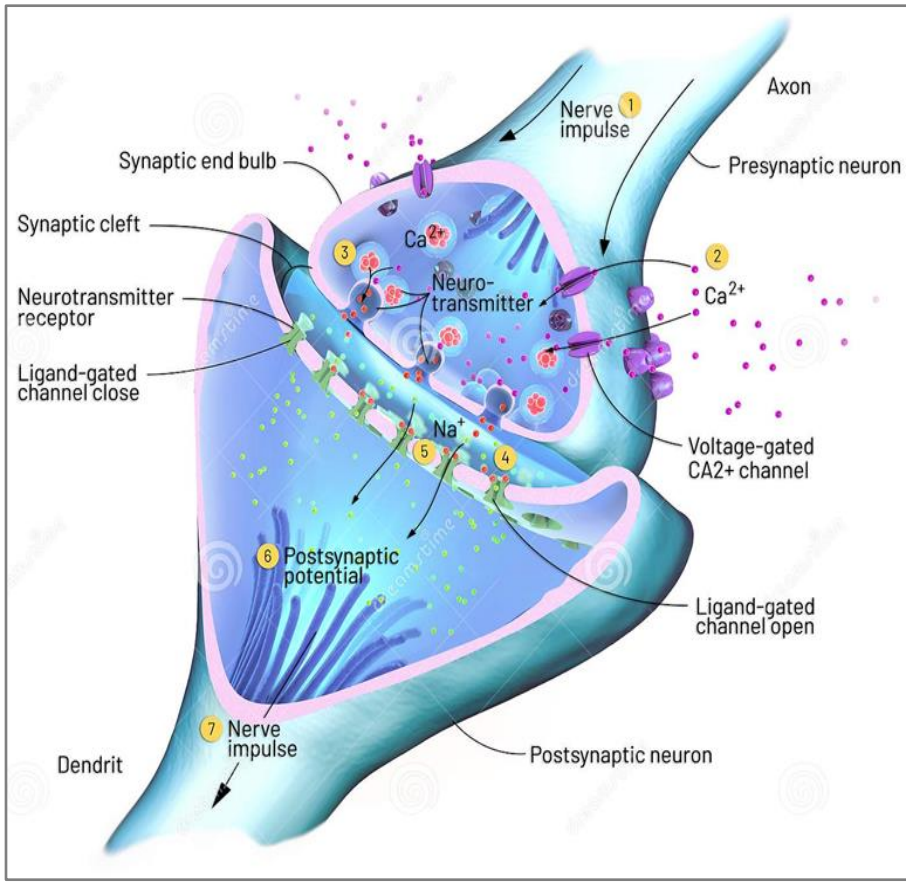
Парадокс: когда различие между системами возрастает (появляются новые признаки) , системы становятся более поведенчески скоррелированными (похожими)

Задача 1. Сигналы между нейронами имеют химическую (лингвистическую) природу

Nobel Prize of **1936** to Henry Dale and Otto Loewi «За открытия, касающиеся химической передачи нервных импульсов»

Nobel Prize of **1963** to John Eccles, Alan Hodgkin and Andrew Huxley «За открытия, касающиеся ионных механизмов возбуждении и торможении мембраны нервных клеток»

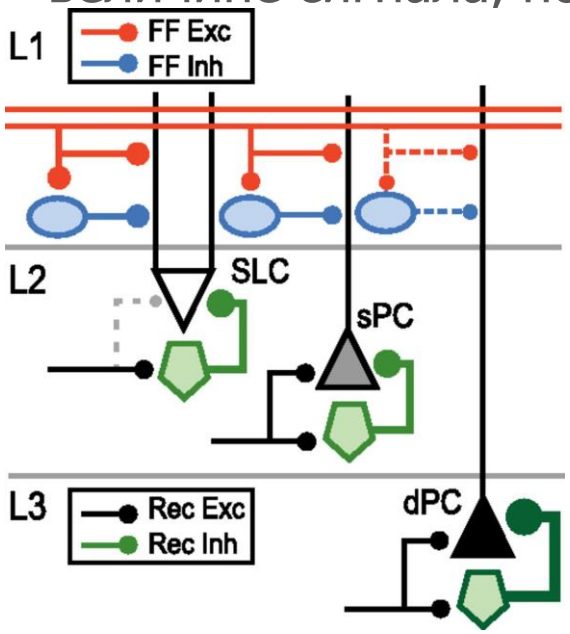
Nobel Prize of **1970** to Julius Axelrod, Ulf von Euler and Sir Bernard Katz «За открытия, касающиеся гуморальных механизмов передачи нервных импульсов химическими веществами - медиаторами, в нервных окончаниях».

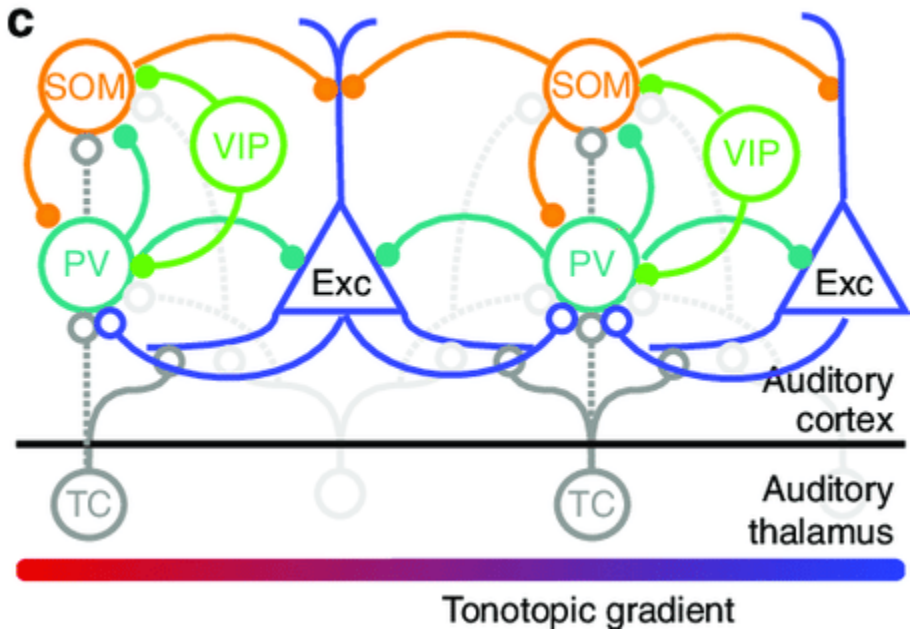


Задача 1 для ИИ:
Химическое кодирование сигналов расширяет объем передаваемой информации, аналогично обучение и обмен информацией в ИИ системах можно организовать как с использованием кодирования с помощью «чисел», так и фразами состоящими из «слов»

Задача 2. Кодирование свойств каналов связи между нейронами типом передаваемых данных

В мозгу используется два типа воздействий – активация и торможение, но не только по величине сигнала, но по его типу передаваемых данных (информация)

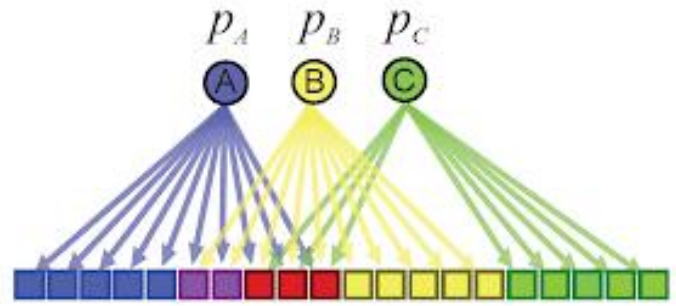


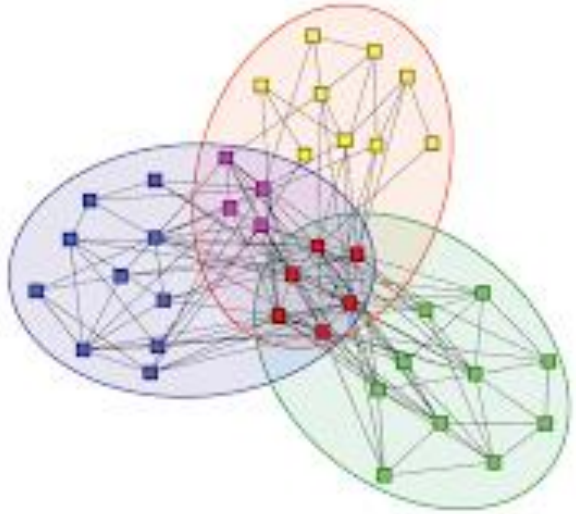


Задача 2 для ИИ:
В биологической нейроморфной сети два типа воздействий создают огромный репертуар регуляторных цепей и механизмов, поэтому в ИНС надо использовать как численные, так и лингвистические контура управляющих воздействий

Задача 3. Многообразие видов каналов и сигналов связи между нейронами

нервная система для передачи информации использует несколько типа нейромедиаторов , кодирующих с помощью химических соединений (слов) сотни разных сигналов (предложений)

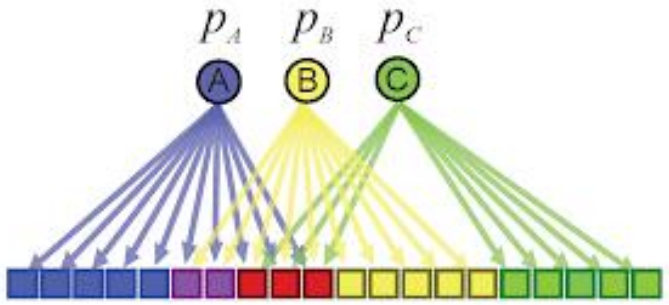


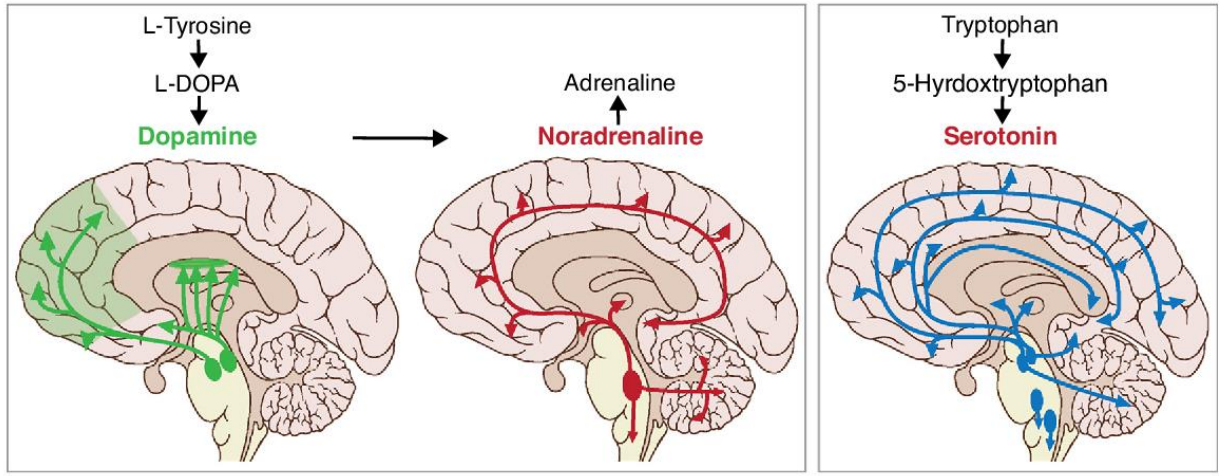


Задача 3 для ИИ:
Необходимо кодировать как количество, так и качество передаваемой информации , например, используя «цвет канала»

Задача 4. Функциональное кодирование типом сигнала

разные нейромедиаторы - сигнальные молекулы используются для кодирования специфических функциональных воздействий, что приводит к воплощению различных видов поведения «субъекта»

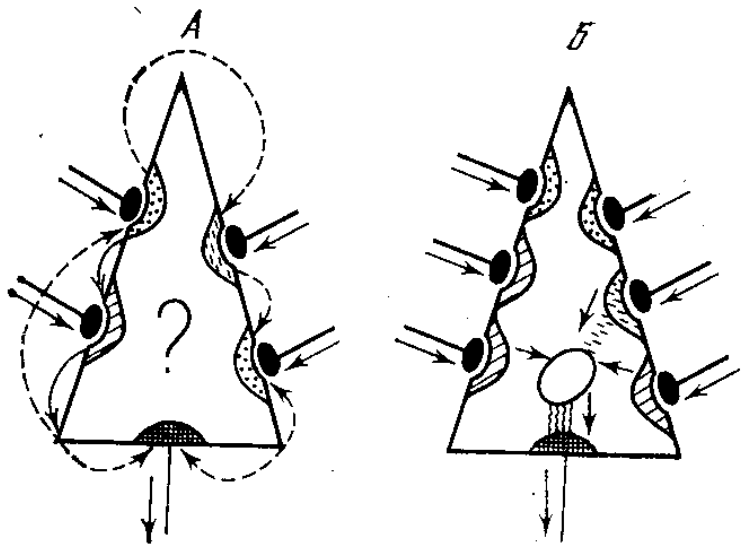




Основная мысль для ИИ:
кодирование функций канала информации выбором его «цвета» или описание дискриптора канала с помощью смылов или набора «слов»

Задача 5. Нейроны интегрируют сигналы не только «на себе», но и «внутри себя»

В итоге у нервной клетке появляется два режима работы с информацией - для себя ‘in / для других out’

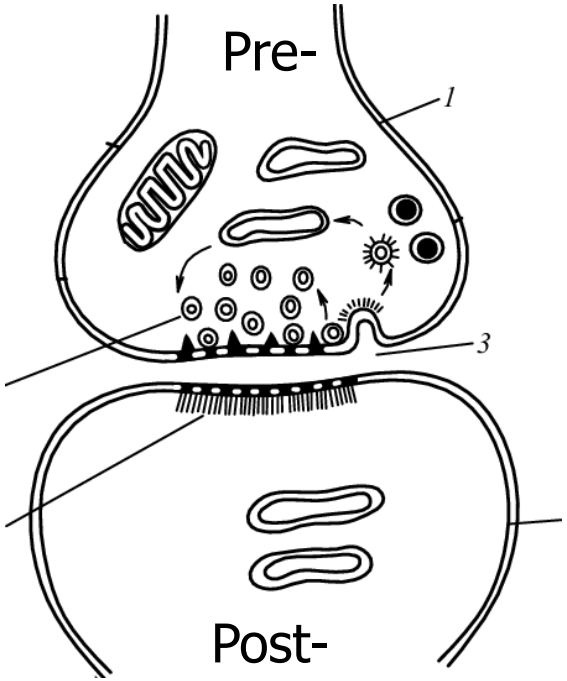


Используется мультимодальная логика кодирования: электрической (А) и химической (Б) формой сигналов

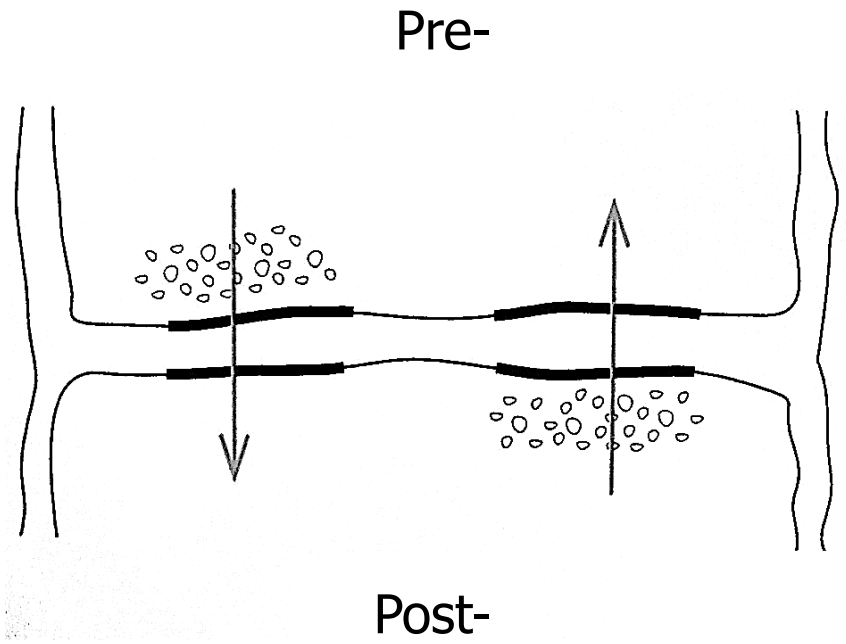
Задача 5 для ИИ:
Помимо суммирования электрических сигналов на поверхности, нейроны передают получаемые ими молекулярные сигналы внутрь клетки, интегрируя их в своей цитоплазме и ядре, собирая информацию о произошедших событиях.

Задача 6. на уровне контакта двух нейронов существует обратное распространение сигнала от пост- к пре синаптическому нейрону

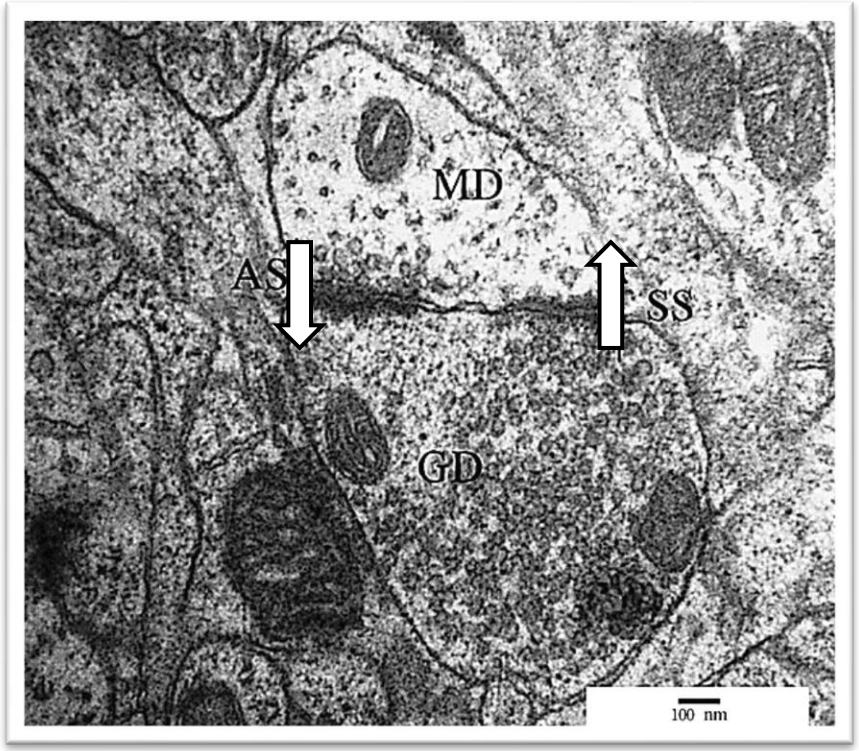
Имеются синапсы с локальной двухсторонней передачей сигнала



Usual synapse

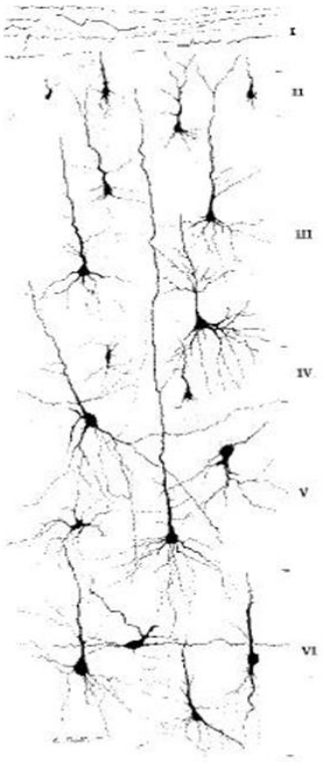


Reciprocal synapse

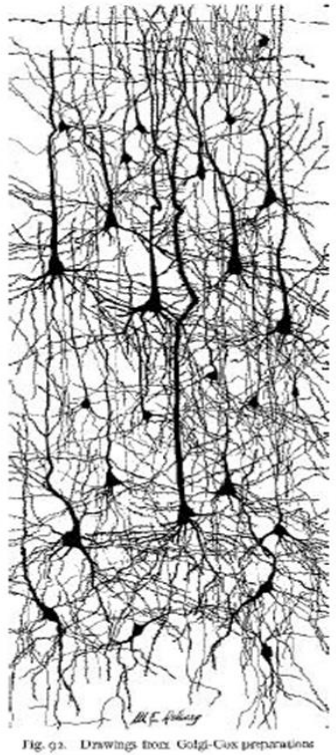


Задача для ИИ: соседние клетки/нейроны в нейронных сетях образуют рекуррентные (циклические) взаимодействия

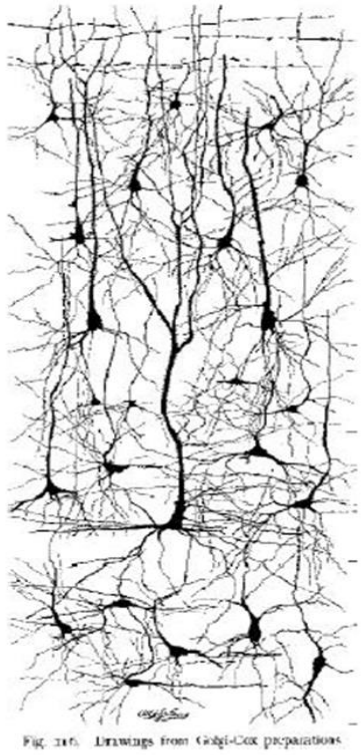
Задача 7. Процессы отбора связей нейронов продолжаются под влиянием обучения и обретенного опыта



Birth



2 years



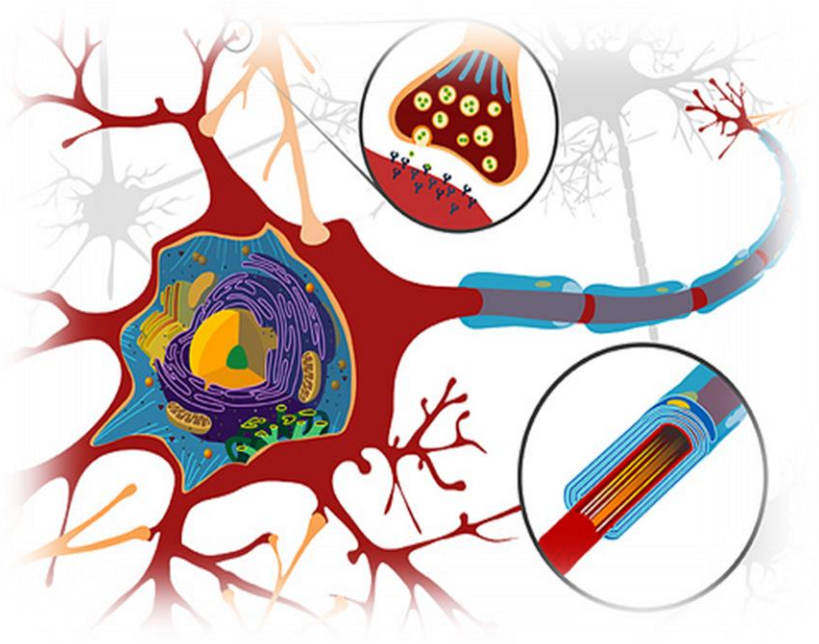
6 Years

Задача 7 ИИ: постоянный встроенный отбор нейронов и связей между ними, как механизм закрепляющий нужные для реализации оптимальных процессов «думания» структуры и параметры

Специфические свойства биологической памяти по сравнению с компьютерной

- **Нерепрезентативность** - она не является точным отражением событий внешнего мира.
- **Реконструктивность** - ее воспроизведение является активным процессом самосборки нейронной системы.
- **Нерепликативность** - каждое ее следующее воспроизведение отличается от предыдущего, вовлекая перекрывающуюся, но отличающуюся популяцию нейронов и синапсов.
- **Рекатегориальность** - каждая ее новая реконструкция при воспроизведении проходит оценку идентичности на весах других, связанных с ней систем.
- **Реконсолидируемость** - каждая новая реконструкция подвергается запоминанию, сходному по своим механизмам с процессами исходного запоминания.

Требования к свойствам нейроморфной памяти по сравнению с оперативной компьютерной памятью



- **вырожденность** - одно и тоже событие хранится в виде множественных неидентичных копий функциональной системы,
- **автоассоциативность** - разные копии одной и той же функциональной системы имеют связи с разнообразными другими системами за счет вырожденности набора входящих в эти копии нейронов,
- **реинтегративность** – целая система может быть извлечена из памяти по возбуждение небольшой части ее элементов,
- **репаративность** – система может восстанавливаться при повреждении части из ее элементов или даже части из ее копий.

Что надо учесть в процессе проектирования

- В настоящее время ИНС не реализуют все известные особенности функционирования мозга. Современные ИНС построены на выявлении корреляций между входными данными при их фиксированном распределении и соответствующих состояний нейронов. Но эти корреляции не отражают причинно-следственную связь (супервентность) между паттернами данных и состоянием ИНС.

- Сформированные на основе корреляций объяснения не выявляют причин, которые вызывает наблюдаемые на «физическом уровне» процессы, что может привести к неправильным обобщениям и искаженному пониманию происходящего.

- Предсказать редкое или нетипичное поведение объекта возможно, но для этого понадобится его мультимодальное описание: объединение корреляционных и физических моделей наблюдаемых состояний