

Курс: управление научными проектами

Занятие 5

ИИ и сознание

23 октября  
2024

## Введение



Ошибка многих исследователей «сознания» в их желании редуцировать СОЗНАНИЕ до каких-то объективных физических процессов , то есть найти СОЗНАНИЮ соответствующий **коррелят** в материальном мире.





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 НСММИ 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 НСММИ 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 НСММИ 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*

# Обсудим трудные проблемы «компьютерных науки» и «физики процессов вычислений»

- **Надо научиться правильно задавать вопросы, например, возможен ли и за счет чего прорыв в технологиях сверхбыстрых вычислениях ?**



**Гипотеза:** Мозг – матрица, где записаны наиболее важные и возможные решения, которые надо принимать во всех возможных ситуациях.

Это значит, что мозг фактически уже принял решение за долго до того, как мы – разумные существа эти решения начинаем осознать ? и воплощать в форме физических процессов ???

**Вопрос:** Что такое сознание ? – **Ответ:** сознание, есть «**со-знание**».

**Вопрос:** где хранятся копии (или какая-то нематериальная сущность) всех знаний, воплощенных во всех формах материальных тел ?

**матрица понятий:**

- Знание-Сознание-Осознание
- Осознание -Энергия-Смысл
- Смысл-Язык-Программа вычислений

# Компьютерные науки

- Объект : информация
- Предмет : средства обработки информации

Будем исходить из того, что информация – это все то, что сохраняется на основе «законов Природы» или может быть сохранено и доступно для копирования во времени и пространстве.... с помощью «компьютерных систем»

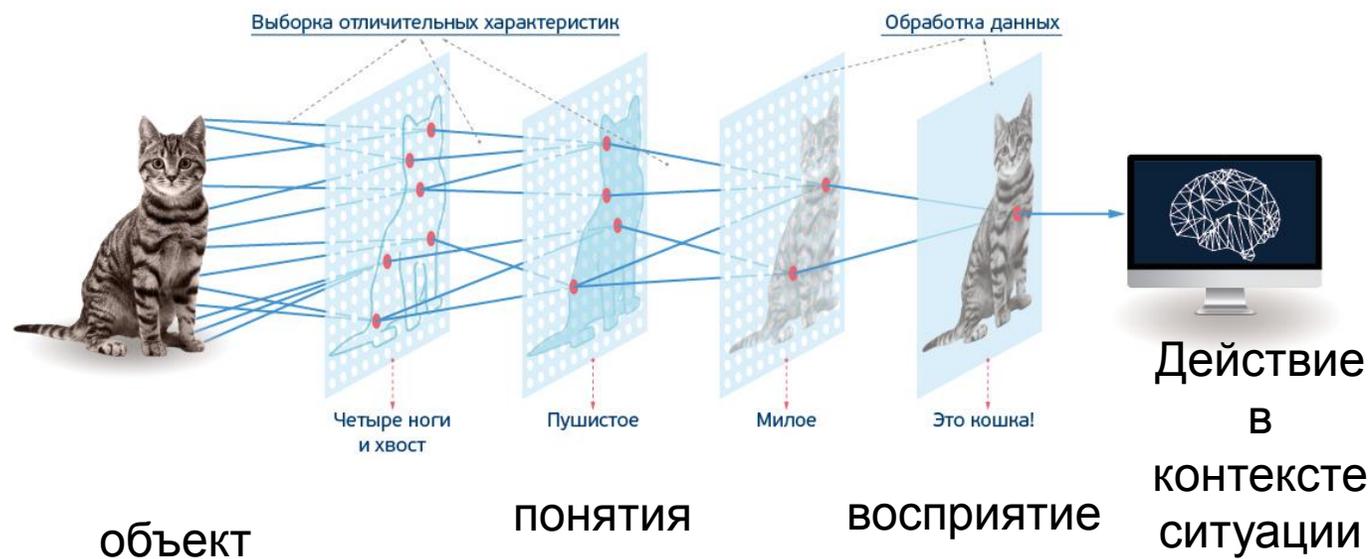
# ПРИНЦИП «шахматной доски» - субъектность элементов



основой научного подхода - возможность разделения окружающего «пространства-время» на **классы эквивалентности**,

т. н. фактор-множества (например, черные/белые клетки), на которых определено множество операций -. Это позволяет выделить подмножества состояний, которые можно рассматривать как **последовательности элементов имеющих «и прошлое, и настоящее, и будущее»**

## КАК РАБОТАЕТ НЕЙРОННАЯ СЕТЬ



Эти элементы образуют базис модели, замкнутый относительно фундаментальных (энергетических, топологических и др.) инвариантов. Возможна и «ИИ реальность», как совокупность инвариантов-цифровых моделей физических

# Физика процессов вычислений

- В основе современной физики как науки лежат **различные законы сохранения**: энергии, импульса, массы и др. Формальная математическая запись законов сохранения суть уравнения мат-физики, решение которых позволяет моделировать физические процессы
- Однако вполне уместен вопрос: «где и как хранится та сущность, которая входит в законы сохранения» ?
- Например, любой изолированный физический объект обладает формой и энергией, а стремление носителя энергии значение «своей» энергии минимизировать «запускает» процессы, которые направлены на то, чтобы сделать внутреннюю энергию объекта была **не отличимой от нуля ???!**
- **(суть второго закона термодинамики)**
- «**стремление**» Природы к **минимизации энергии** — запускает природный «**двигатель**», **который заставляет атомы** любого твердого тела, например куска металла, вместе «работать» над тем, чтобы металлический объект перешел в состояние, в котором энергия его внутреннего магнитного поля неотличима от нуля.
- **Искусственная Нейросеть (ИНС)**, как и твердое тело, также стремится минимизировать энергию «**формы входных данных**» .
- Обучение ИНС направлено на изменение «формы» сети таким образом, чтобы определенные процессы, изменения **формы входных данных** требовали минимума энергии. Суть процесса обучения ИНС в минимизации энергии, затраченной на изменение «форма» .
- Можно ли и **как физические принципы науки** использовать при разработки суперкомпьютеров ???

# Мозг - отвечает за все, что мы делаем.



*И.М. Сеченов*

- Нет никакой разницы в процессах, обеспечивающих в мозгу реальные события, их последствия или воспоминания о них.
- Наша зависимость от мозга больше, чем мы привыкли думать
- Мы должны его знать!



Гедель утверждал: никакая система не способна изучить и воспроизвести свойства системы более сложной, чем она сама.

В задачах ИИ мы и пытаемся заниматься с помощью машин Тьюринга смоделировать мозг и его функции .

## Дж. Хинтон лауреат Нобелевской премии по физике, 2024 года: в ожидании «третичной теории»



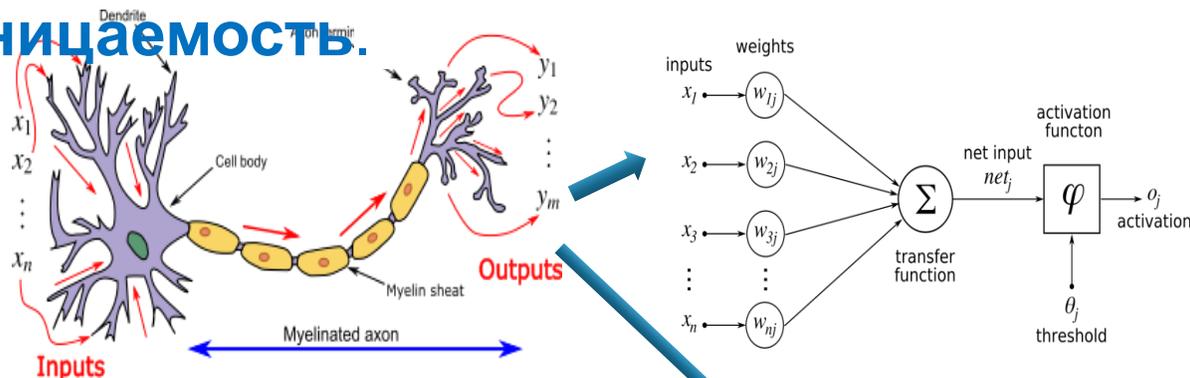
Сложные отношения ИИ к физическому миру не выражаются через диадические понятия

Требуется ревизия того, как мы представляем сознание и ее материальные носители

Суть новой **третичной теории** - **введение третьего понятия** – информации, понятия ментального и физического логически объединить в общем «пространстве состояний»

# Биологический нейрон и его математические модели :

## важнейшим свойством мембраны является избирательная проницаемость.



Модель N1: искусственный нейрон **параметрическая модель избирательной проницаемостью не обладает**.

В клеточные мембраны встроены различные специфические **рецепторы** — белки или гликопротеины, распознающие определенные молекулы (**лиганды**). Это значит, что молекулы и ионы проходят через мембрану с различной скоростью **СВОЙСТВА КЛЕТОЧНОЙ МЕМБРАНЫ как**

**трансформера:**

1. Обладает **избирательной проницаемостью**, которая изменяется при различных состояниях клетки
2. Имеет **каналы**, через которые проникают ионы:

- Потенциал-зависимые каналы — открываются при **изменении разности потенциалов**;
- Потенциал-независимые (лигандозависимые, гормонрегулируемые) — открываются при **взаимодействии рецепторов с веществами**.

$q$  – query - **запрос**

$k_i$  – key - **ключ**

$v_i$  – value – **значение**

$$y = \sum_{i=1}^N \alpha(q, k_i) v_i$$

«ядро» модели

$$\alpha(q, k_i) = \text{softmax}_i \left( \text{score}(q, k_j) \right)_{j=1}^N$$

Принцип цифрового моделирования: модель получает сигналы «х» от других нейронов и «вычисляют» пороговую функцию выходного сигнала «у», которая передается в другие нейроны ИНС. Принцип когнитивного моделирования: прогнозирования + объяснение вычисляемых решений

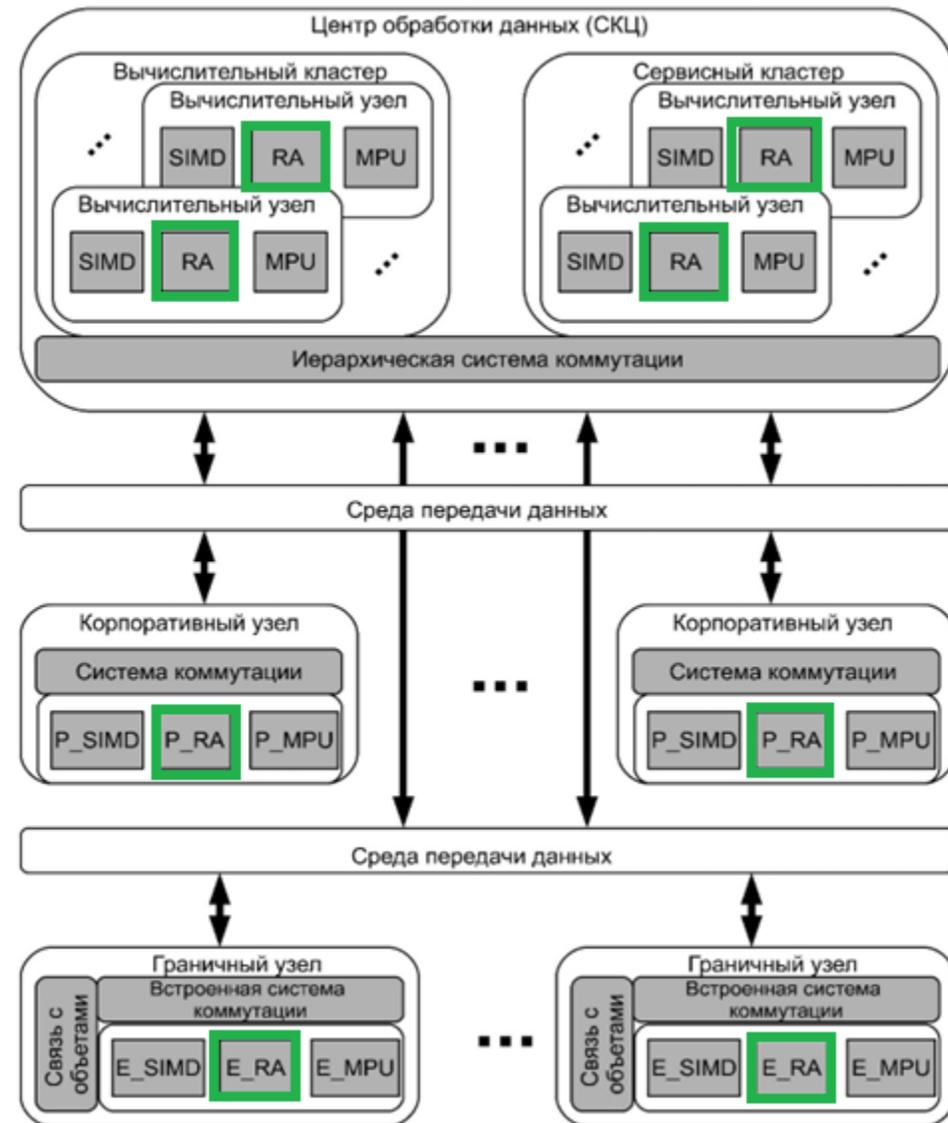
Модель N2: искусственный нейрон **непараметрическая модель избирательной проницаемостью обладает**

# «ТРИТИЧНАЯ» гибридная модель «машины Геделя»

Уровень «объяснения»  
результатов

Уровень машинного  
обучения и обобщения

Уровень машины  
Тьюринга: бинарных  
логических выводов и  
индуктивной  
предобработки



# Суперкомпьютера ... is nothing but трансформер возможных состояний

## ВЫЧИСЛИТЕЛЬНЫХ ПРОЦЕССОВ В МЫСЛИМЫЕ ОБЪЯСНЕНИЯ

Деревья классификации возможных состояний системы

Имеется

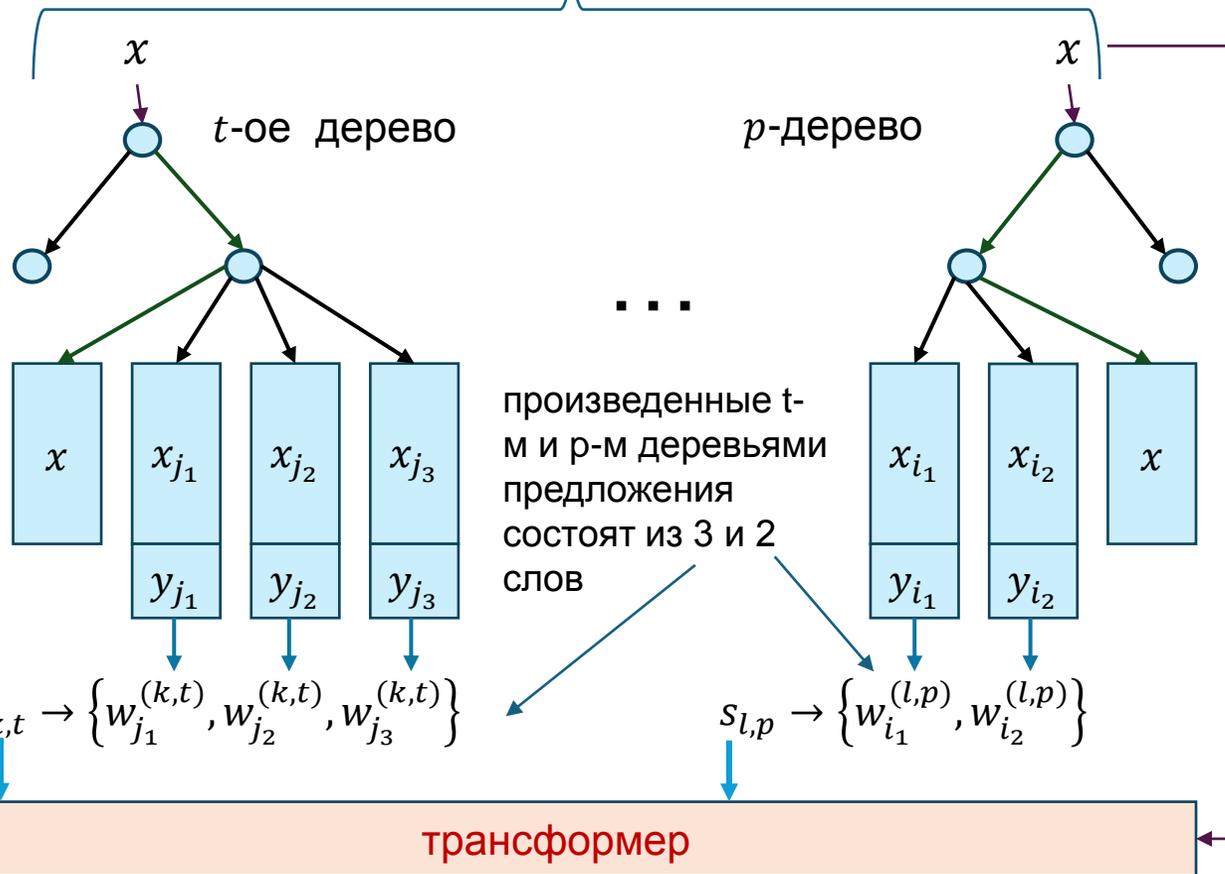
- 1) обучающая выборка  $(x_j, y_j)$ , где  $x$  – вектор признаков
- 2) индексы  $j \in J_k(x)$ , попадают совместно с вектором признаков  $x$  в один и тот же  $p$ -й лист  $k$ -го дерева классификации.
- 3)  $(x_j, y_j)$ , представим в виде предложения  $s$ , состоящего из слов, обозначаемых как  $w_i^{(k,p)} \in J_k(x)$ ,
- 4) из этих слов можно «собрать» объяснение вычисленного результата

$i \in J_k(x)$ ,

$$s_{k,p} = \{nw_{i1}^{(k,p)}, \dots, w_{ip}^{(k,p)}\}_{1ir}$$

«Предложение»  $s_{k,t} \rightarrow \{w_{j_1}^{(k,t)}, w_{j_2}^{(k,t)}, w_{j_3}^{(k,t)}\}$

$k$  - номер дерева;  $p$  - номер листа;  $i$  - индекс примера (индекс слова в  $p$ -м предложении), который попадает в  $p$ -й лист.



произведенные  $t$ -м и  $p$ -м деревьями предложения состоят из 3 и 2 слов

$$s_{l,p} \rightarrow \{w_{i_1}^{(l,p)}, w_{i_2}^{(l,p)}\}$$

ИЗ КАЖДОГО «ЛИСТА» ДИРЕКТЬЕВ МОЖЕТ ВЫХОДИТЬ РАЗНОЕ КОЛИЧЕСТВО «ВЕТВЕЙ»:

- 1) из листа  $t$ -ого дерева выходит три «слова» описания реакции, организма, а из листа  $p$ -дерева выходит описание, состоящее из двух «слов»
- 2) На вход системы подается текущий вектор « $x$ » и

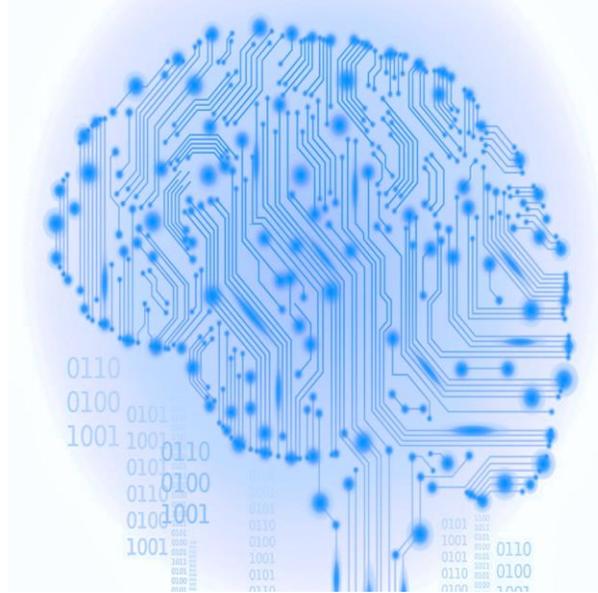
### Задача ИИ системы

- 1) вычислить оценку  $[x^{\wedge}]$
- 2) По этой оценке рассчитать значения индексов Робинсона, Старра, Кердо, то есть  $\hat{y}$  и выбрать вид и форму воздействия  $\mathbf{z} = \mathbf{q}(\hat{y}, \mathbf{d1}, \mathbf{d2}, \mathbf{d3})$ , где коэффициент  $d_i$  получаются также в процессе обучения для каждого

## Часть 2 (по материалам ак. К. В. Анохина)

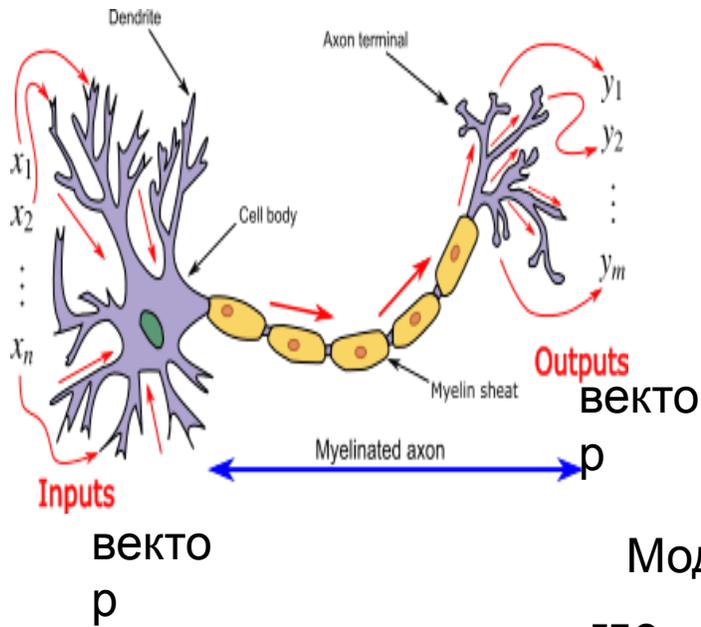
### Человеческий мозг—самоприменимая сложная адаптивная система

Мозг состоит из приблизительно 90 миллиардов нейронов, которые структурированы триллионами взаимосвязанных синапсов. Модель нейроморфных вычислений предполагает выполнение вычислений биологически правдоподобным образом. Частью нейроморфных вычислений являются спайковые нейронные сети. Спайковая нейронная сеть (СН) является одним из кандидатов для преодоления ограничений нейронных вычислений и эффективного использования алгоритма машинного обучения в реальных приложениях.



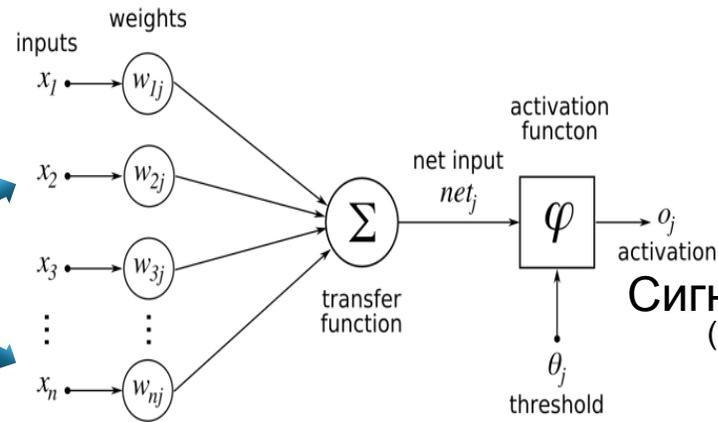
*Spiking neural network, SNN*) — третье поколение искусственных нейронных сетей (ИНС), которое отличается от бинарных (первое поколение) и частотных/скоростных (второе поколение) ИНС тем, что в нем нейроны обмениваются короткими (у биологических нейронов — около 1—2 мс) импульсами одинаковой амплитуды (у биологических нейронов — около 100 мВ). Является самой реалистичной, с точки зрения физиологии, моделью ИНС

# Модели процессов фазовой синхронизации больших популяциях нейронов



Модель N1

$\{x\}$



Модель N1: искусственный нейрон -параметрическая модель

Модель N2: искусственный нейрон непараметрическая модель

Сигнал активации  $u_i$   
(амплитуда, фаза)

Модель N2

$$y = \sum_{i=1}^N \alpha(q, k_i) v_i$$

«ядро» модели

где  
 $q$  – query -запрос  
 $k_i$  – key -

ключ

$v_i$  – value –

значение

$$\alpha(q, k_i) = \text{softmax}_i \left( \text{score}(q, k_j) \right)_{j=1}^N$$

Несколько сигналов активации нейронов м.б. фазово-синхронизированы – то есть соотношение их фаз не меняется со временем или фазы всех сигналов зависит от какой-либо одной детерминированной функции.

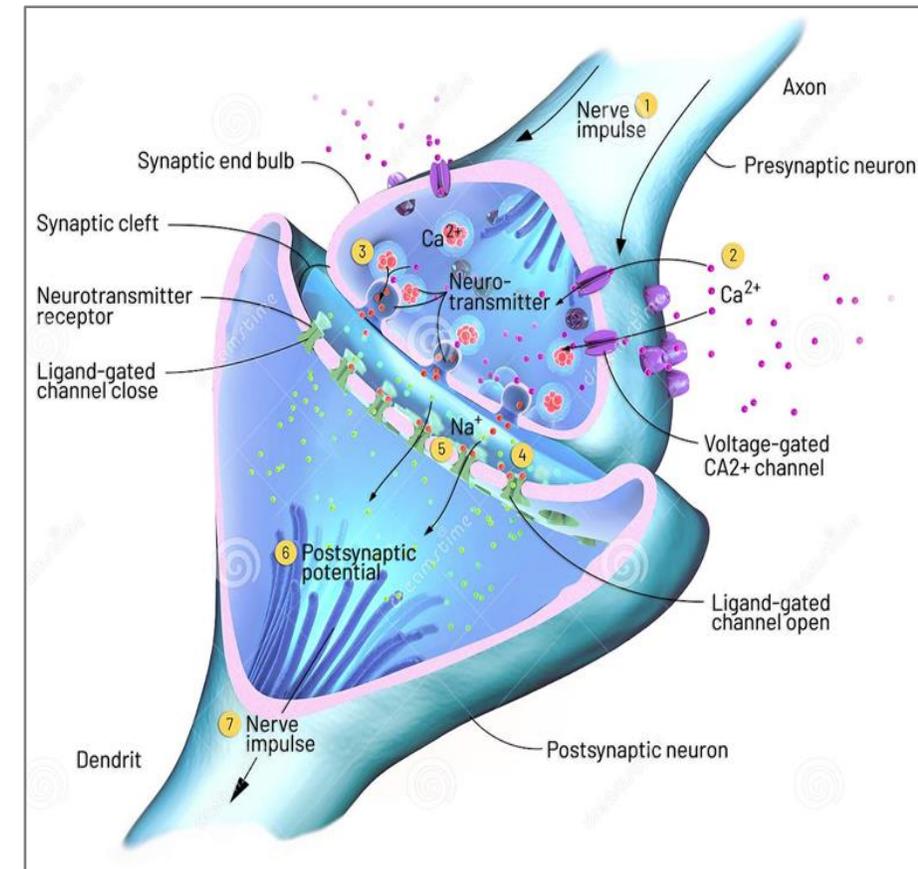
Синхронизированная активность больших популяциях нейронов является основным механизмом образования макроколебаний или ритмов. В этом случае каждый нейрон получает входные микросигналы  $\{x\}$  от других нейронов и «вычисляют» пороговую функцию выходного сигнала « $y$ », которая передается в другие нейроны сети, а синхронизация фаз этих сигналов приводит к макроколебаний

# 1. Сигналы между нейронами имеют химическую 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 возбуждения и торможения 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».

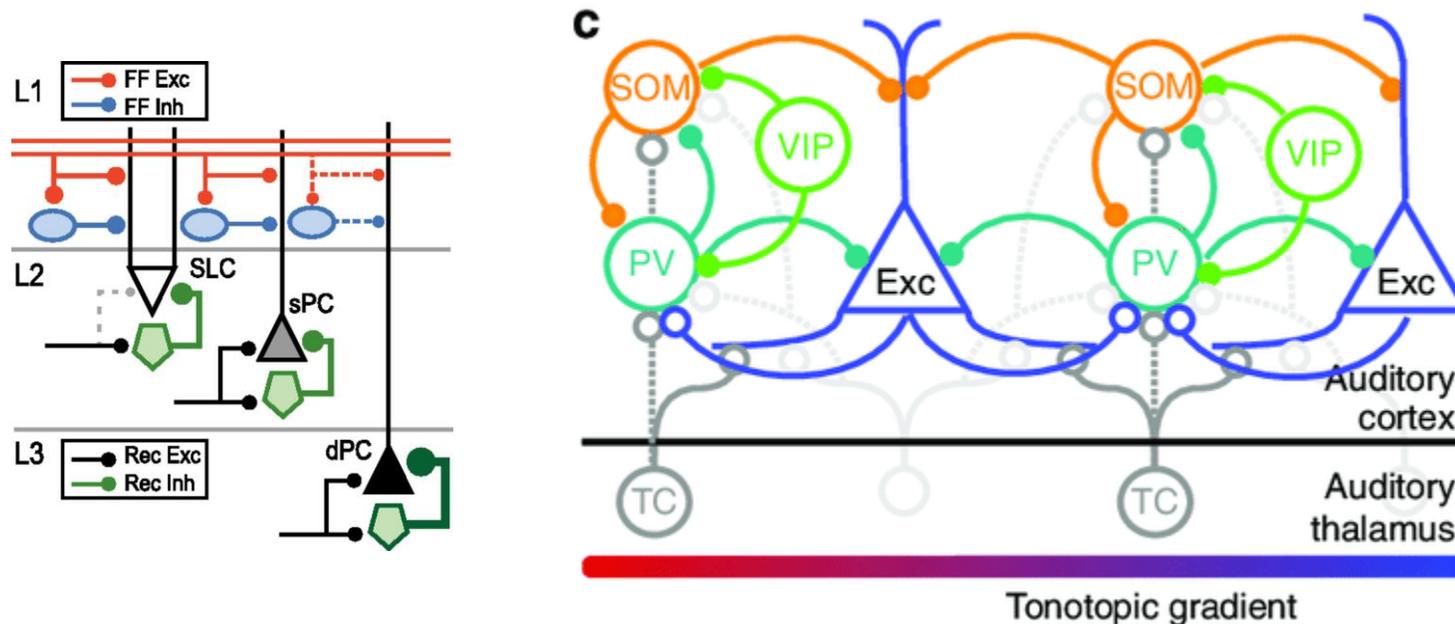


**идея для применения в системах ИИ:**

Химическое кодирование сигналов необычайно расширяет объем передаваемой информации, поэтому обмен информацией в ИИ системах можно реализовать как через с использованием значения «чисел», так и смысла «слов»

## 2. кодирование сигналов между нейронами

Используется два типа воздействий – активация и торможение, но не только по величине сигнала, а по его типу, при этом информация передается одними и теми же электрическими сигналами

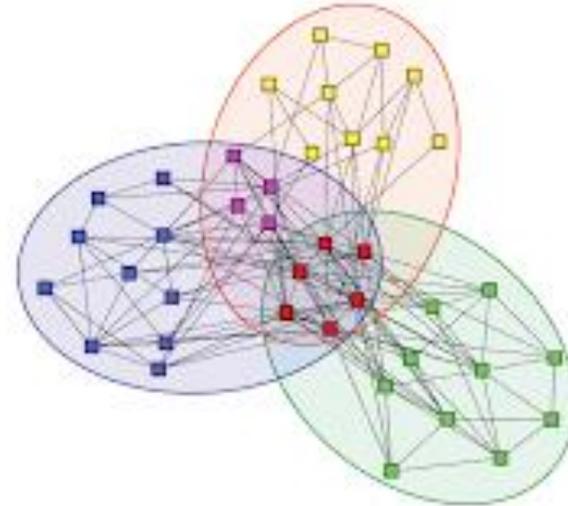
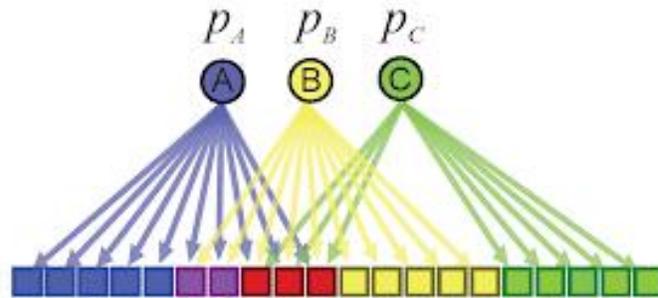


**идея для применения в системах ИИ: :**

В биологической нервной сети эти два типа эффектов создают огромный репертуар регуляторных цепей и механизмов, поэтому ИНС надо использовать как численные, так и лингвистические контура управляющих воздействий

### 3. Многообразие виртуальных (химических) сигналов между нейронами

Нервная система использует «конечное» количество типов медиаторов - возбуждающий и тормозящий, которые способны кодировать десятки и сотни разных типов химических сигналов, образующих управляемые информационные суперпозиции

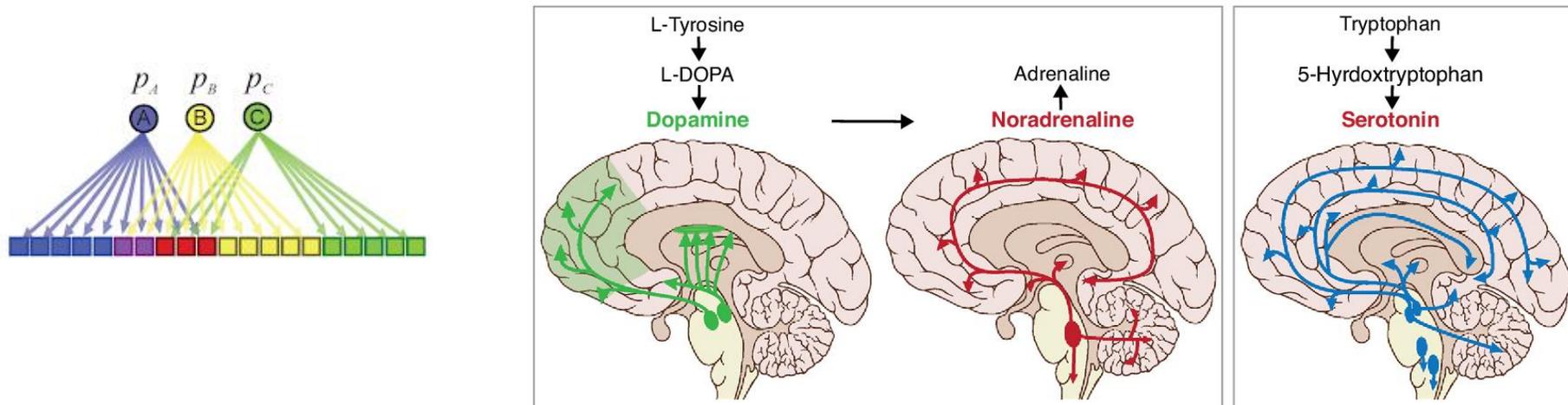


**идея для применения в системах ИИ:**

Необходимо кодировать как количество, так и качество передаваемой информации, например, используя «цвет канала»

## 4. Функциональное кодирование химическим типом сигнала

В режиме суперпозиции разные нейромедиаторы - сигнальные молекулы используются для кодирования не только отдельных сигналов, но и функциональных подсистем - носителей различных видов реакций и поведения

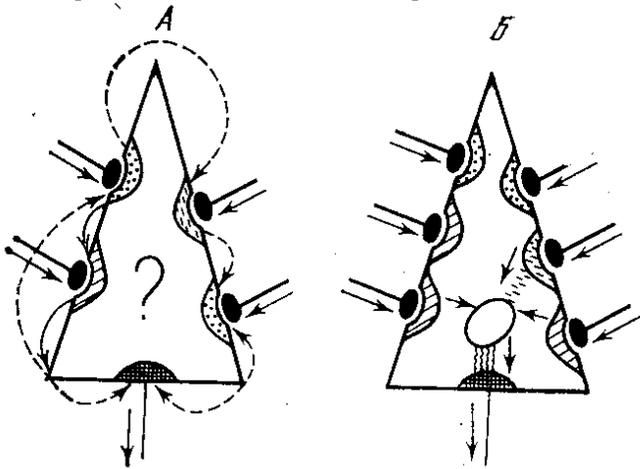


**идея для применения в системах ИИ:**

кодирование функций канала информации или «цветом канала» или «словом»

## 5. Нейроны интегрируют сигналы не только «на себе», но и «внутри себя» - режим «самоприменимости» результатов «вычислений»

Самоприменимость нейросетей – фактор, который позволяет каждой нервной клетке иметь два режима работы с информацией 'in/out', а также работать в режиме мультимодальной логики



мультимодальная логика сигналов, а именно:

- электрической (А)
- химической (Б)

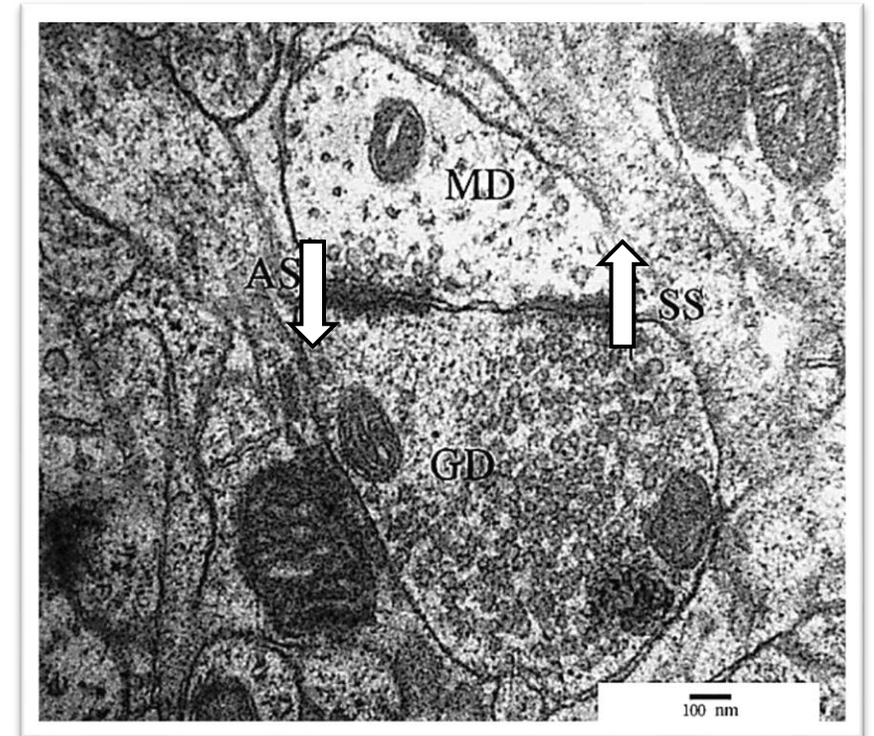
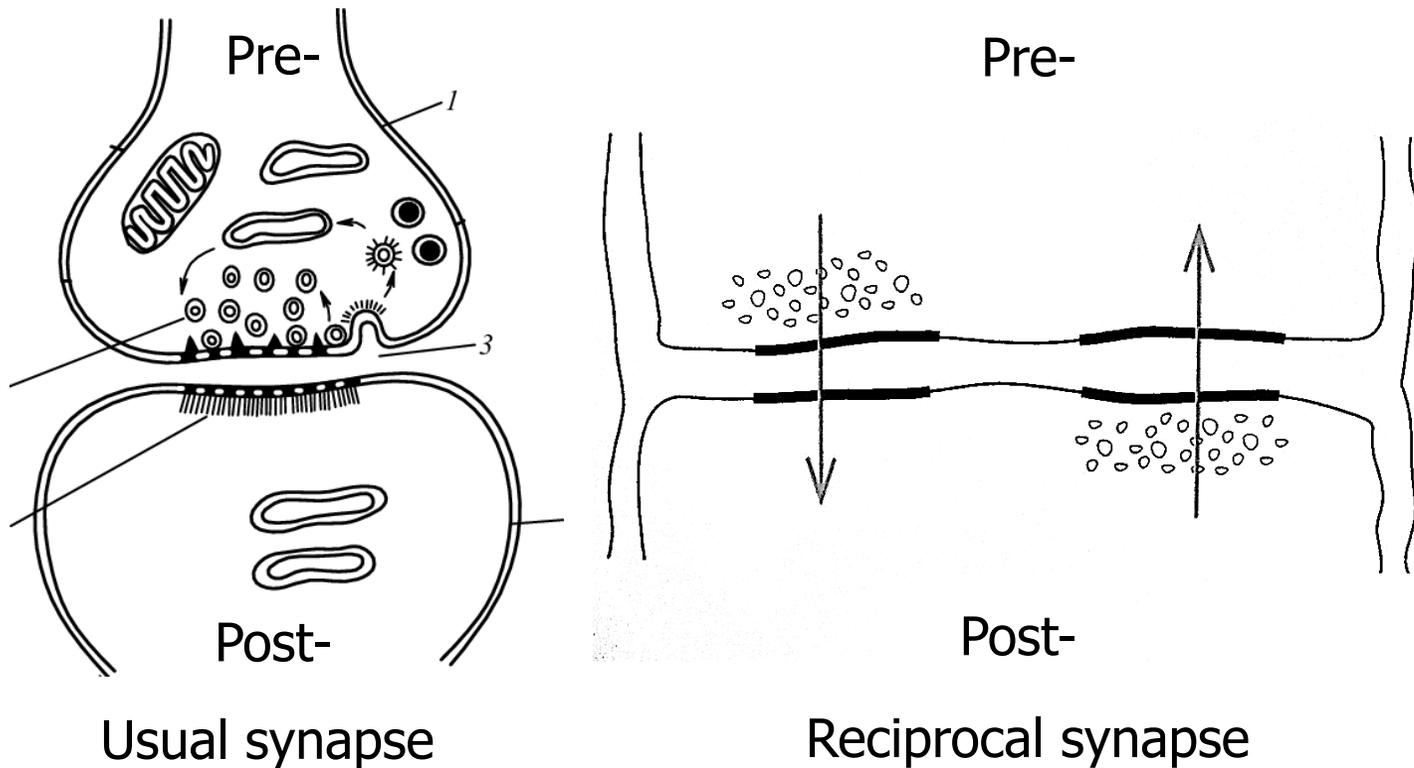
носители информации, генерируемых нейроном сигналов

### **идея для применения в системах ИИ:**

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

## 6. на уровне контакта двух нейронов существует обратное распространение сигнала от пост- к пресинаптическому нейрону

*Реципрокные синапсы с двухсторонней передачей сигнала реализуют принципы «подчиненного адаптивного управления»*



**идея для применения в системах ИИ:** соседние клетки/нейроны в нейронных сетях реализуют рекуррентные (циклические) локальные взаимодействия

## Мыслимость физически невозможного : метафизика «духа и тела»

«сознание следует признать концептуально несводимым аспектом реальности» *Т. Нагель*

**обладать сознанием — значит быть кем-то или чем-то, то есть переживать собственный феноменальный опыт.**

В 17 веке Джон Локк говорил, что разные люди черный цвет могут воспринимать по-разному. Людвиг Витгенштейн утверждал, что невозможно прочувствовать чужую боль — ей можно только сопереживать.