

Вычислительная концепция и архитектура искусственной психики

Антон Колонин

akolonin@aigents.com

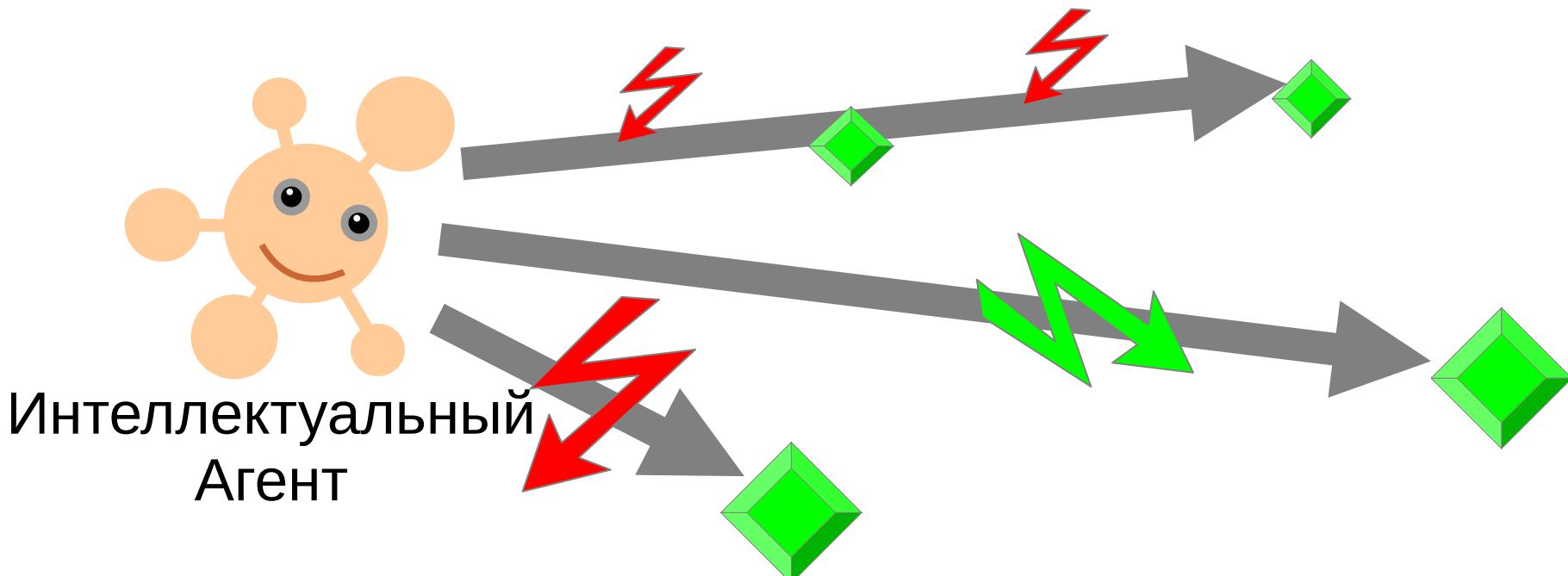
Telegram: akolonin



<https://agirussia.org>

Интеллект:

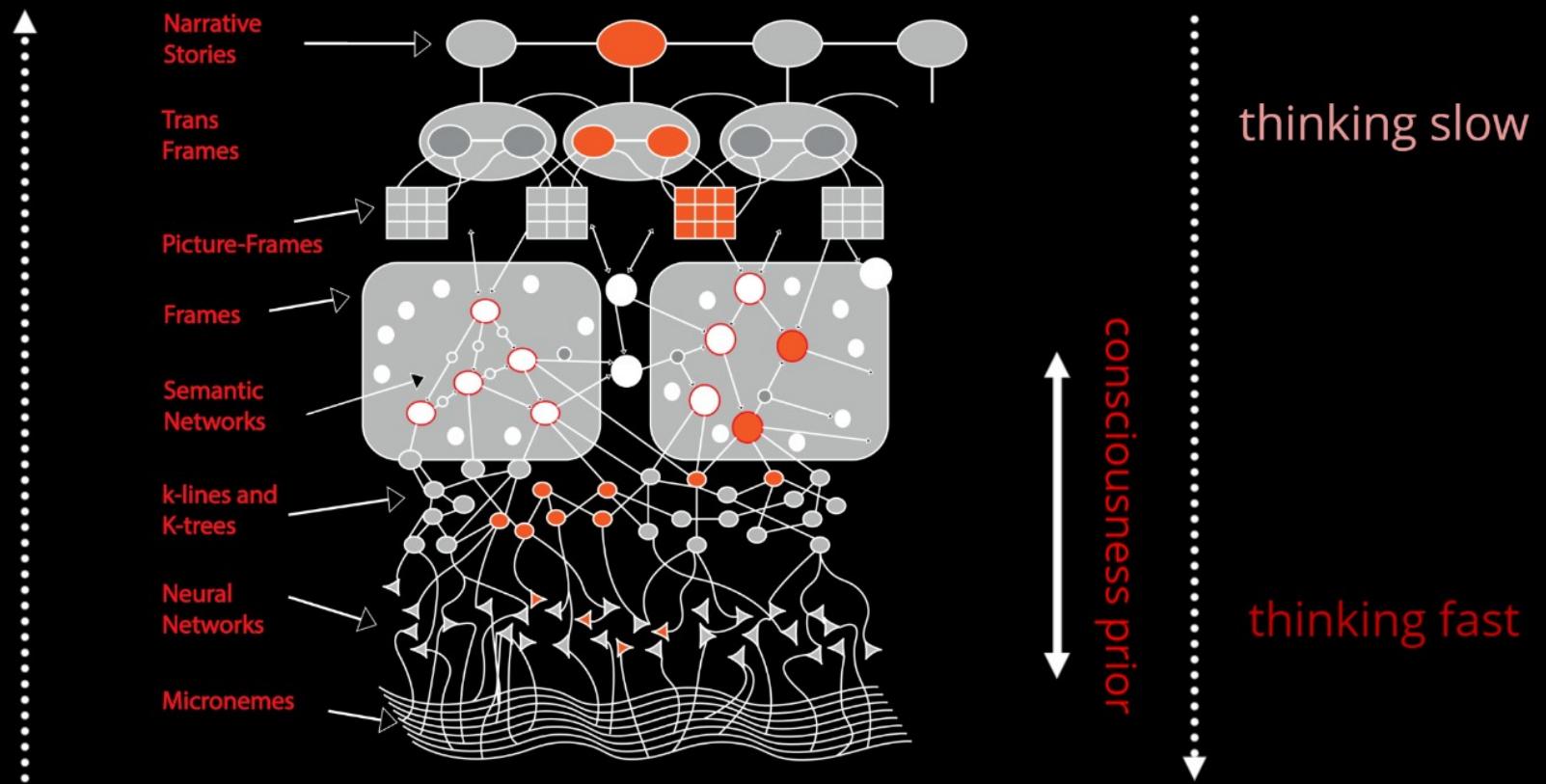
Достижение сложных **целей** в различных
сложных **средах**, в условиях ограниченных ресурсов
(Ben Goertzel + Pei Wang + Shane Legg + Marcus Hutter)



“Быстрое и медленное мышление” – Daniel Kahneman

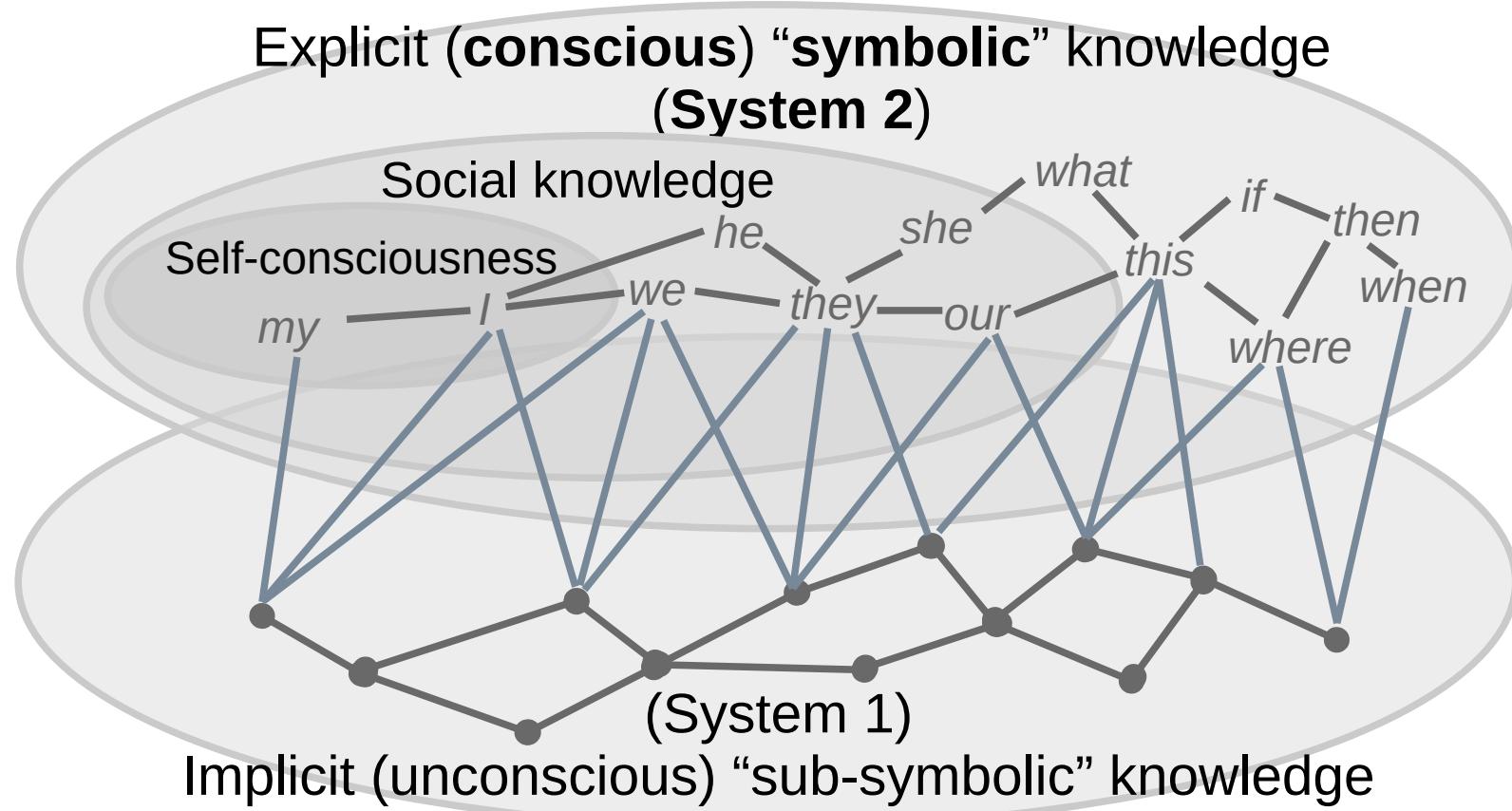
easy
explanation
learning fast

hard
explanation
learning slow



<https://towardsdatascience.com/explainable-ai-vs-explaining-ai-part-1-d39ea5053347>

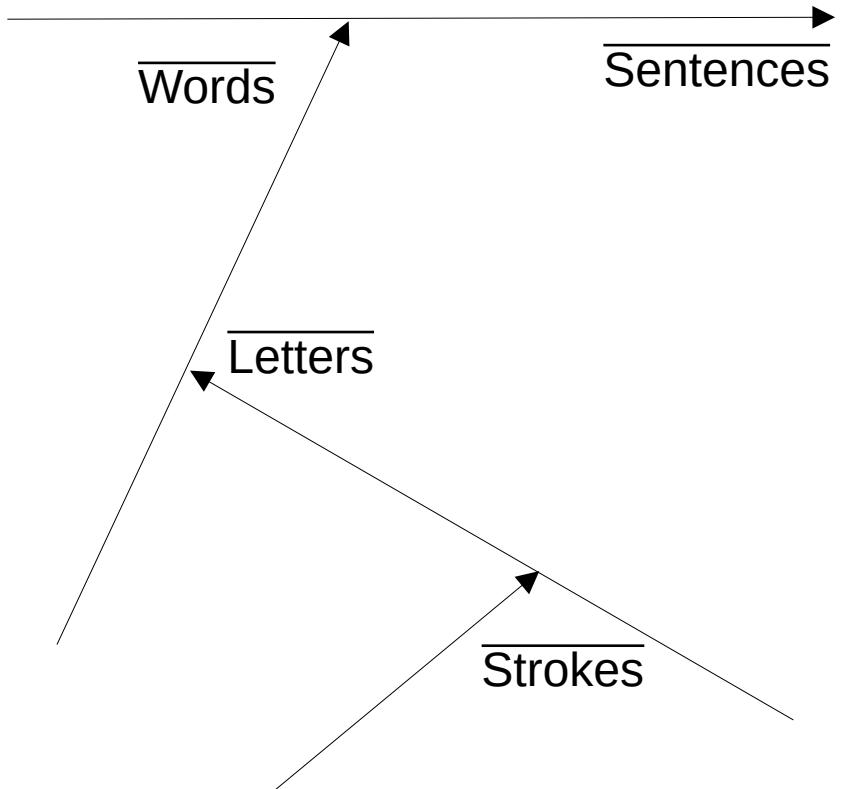
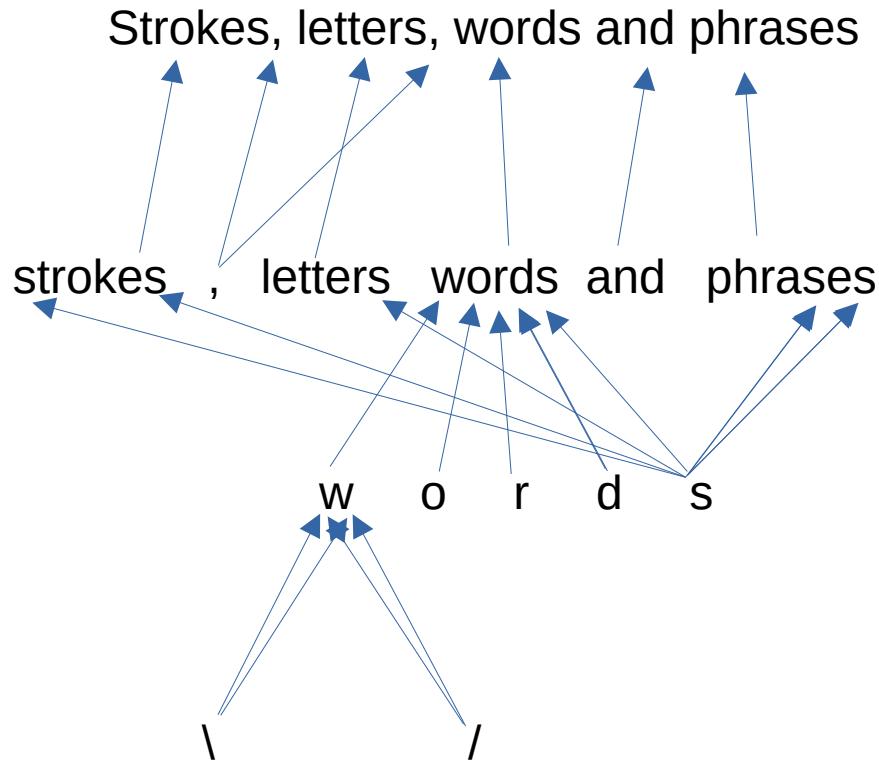
Среда: “нейро-символьный” “граф знаний”



<https://www.amazon.com/Thinking-Fast-Slow-Daniel-Kahneman/dp/0374533555>

<https://amit02093.medium.com/atomspace-hyper-graph-information-retrieval-system-450cab9d751e>

Функциональная эквивалентность графовых и нейросетевых тензорных моделей



Психика (от др.-греч. ψυχή́ «душевный, жизненный») — сложное понятие в **психологии** и **медицине**, которое, в зависимости от областей знаний и направлений **наук**, определяется как:

- Совокупность мыслительных процессов и явлений (**ощущения, восприятия, эмоции, память** и т. п.); специфический аспект **жизнедеятельности** человека и других животных с **нервной системой** в их взаимодействии с окружающей средой^{[1][2]}.
- «Форма активного отображения субъектом объективной реальности, возникающая в процессе взаимодействия высокоорганизованных живых существ с внешним миром и осуществляющая в их поведении (деятельности) регулятивную функцию». ^[3]
- Системное свойство высокоорганизованной живой материи, заключающееся в активном отражении субъектом объективного мира, в построении неотчуждаемой от него картины мира и регуляции на этой основе своего поведения и деятельности^[4].
- Внутренний мир человека (предмет психологии)^[5].

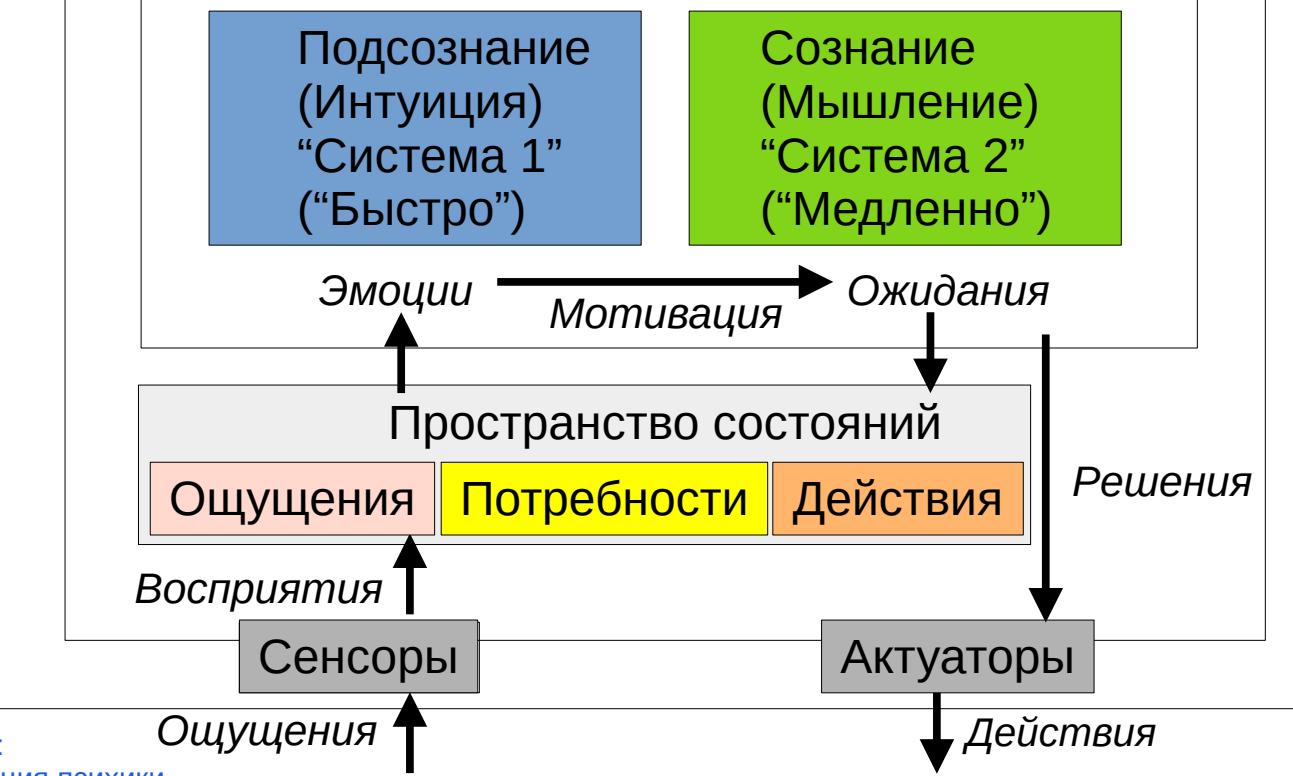
Психика животных — субъективный мир животного, охватывающий весь комплекс субъективно переживаемых процессов и состояний: **восприятие, память, мышление, намерения, сны** и т. п.^[6]

Психика характеризуется такими качествами, как целостность, **активность**, развитие, саморегуляция, коммуникативность, адаптация и т. д.; связана с **соматическими** процессами.

Система = Интеллектуальный Агент

Психика = Операционная система

Интеллект = Система принятия решений



А.Г.Колонин, В.Г.Крюков:
Вычислительная концепция психики,
Статья принята на конференцию
Нейроинформатика-25

Внешний мир = Операционная среда

Copyright © 2025 Anton Kolonin, Vladimir Kryukov

Психика = Операционная система

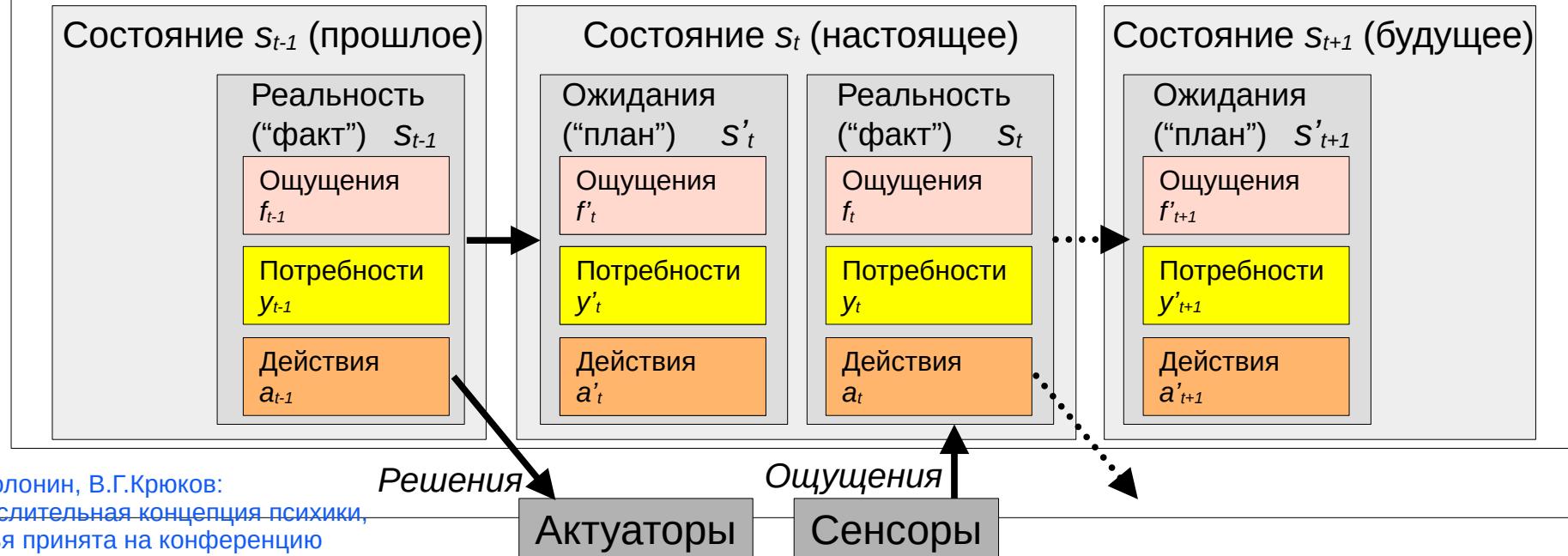
Интеллект = Система принятия решений

Модели s ("инварианты") состояний, полезности U и вероятности P переходов
 $U(\{S_t\}_{t \in \{-T, -1\}}, S'_0) = L(x \cdot (y_t - y_{t+1}), s'_t - s_t, E(a_t))$ $s'_t = \text{argmax}_s(U(\{S_t\}_{t \in \{-T, t-1\}}, s'_t), P(\{S_t\}_{t \in \{-T, t-1\}}, s'_t))$

↑Обучение на опыте

↓Принятие решений

Пространство состояний и эпизодическая память ("прецеденты")



Энергоэффективность
 a

Психика = Операционная система

Оптимальное решение
и ожидание s

Предсказумость
 s

Интеллект = Система принятия решений

Ожидаемая
полезность s

Модели s ("инварианты") состояний, полезности U и вероятности P переходов
 $U(\{S_t\}_{t \in \{-T, -1\}}, S'_0) = L(x \cdot (y_t - y_{t+1}), s'_t - s_t, E(a_t))$ $s'_t = \text{argmax}_s(U(\{S_t\}_{t \in \{-T, t-1\}}, S'_t), P(\{S_t\}_{t \in \{-T, t-1\}}, S'_t))$

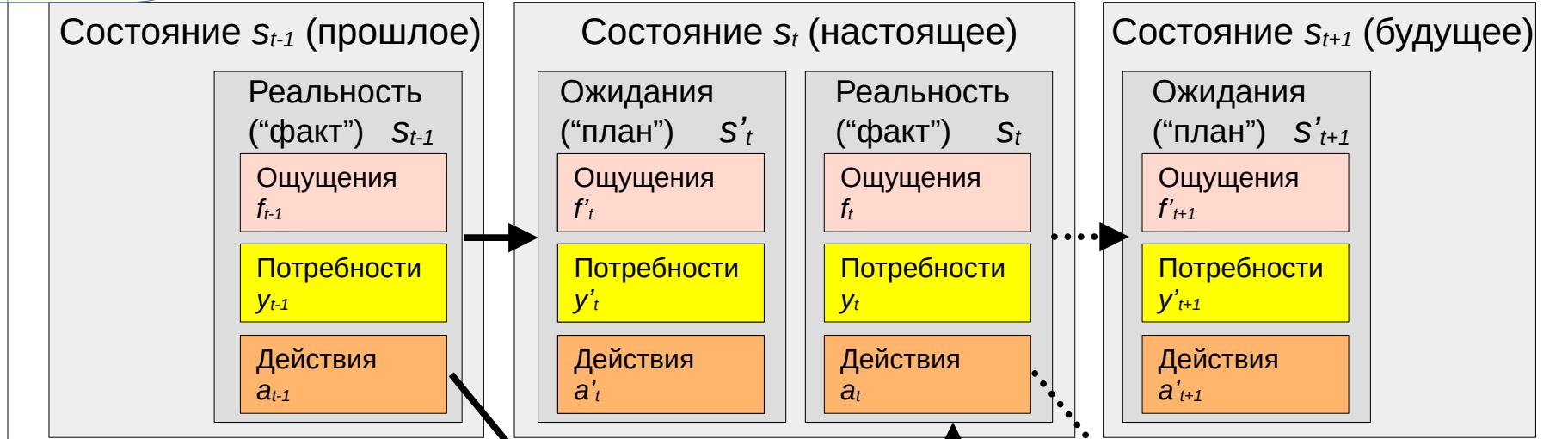
↑Обучение на опыте

↓Принятие решений

Ожидаемая
вероятность s

Удовлетворенность
 y

Пространство состояний и эпизодическая память ("прецеденты")



$x \cdot y_t$ – "мотивационный вектор"

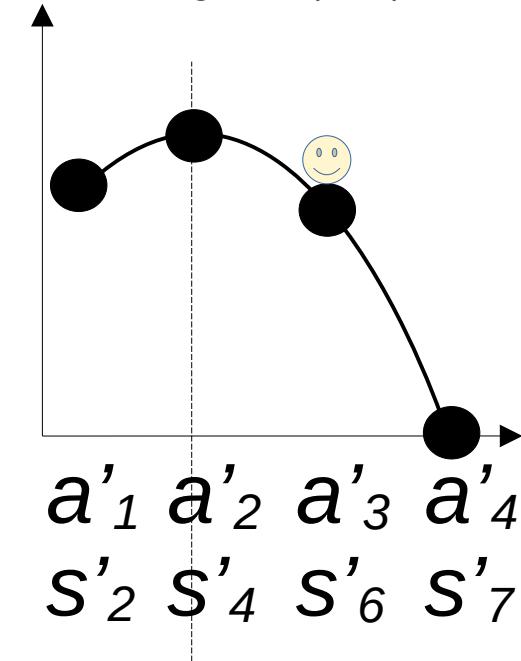
V. F. Petrenko and A. P. Suprun, "Goal oriented systems, evolution, and the subjective aspect in systemology," Tr. Inst. Sistem. Analiza RAN 62 (1) (2012)

Принятие решений как управление операционными рисками

s_t	s'_{t+1}				U	P	$\sum U^*P$
		a'	y'	f'			
s_1	s'_2	a'_1	y'_1	...	1.0	0.5	<u>0.7</u>
s_1	s'_3	a'_1	y'_2	...	0.4	0.5	
s_1	s'_4	a'_2	y'_3	...	1.0	0.8	<u>0.8</u>
s_1	s'_5	a'_2	y'_4	...	0.0	0.2	
s_1	s'_6	a'_3	y'_5	...	0.6	1.0	<u>0.6</u>
s_1	s'_7	a'_4	y'_6	...	0.0	1.0	<u>0.0</u>

$$s'_t = \operatorname{argmax}_s(U, P)$$

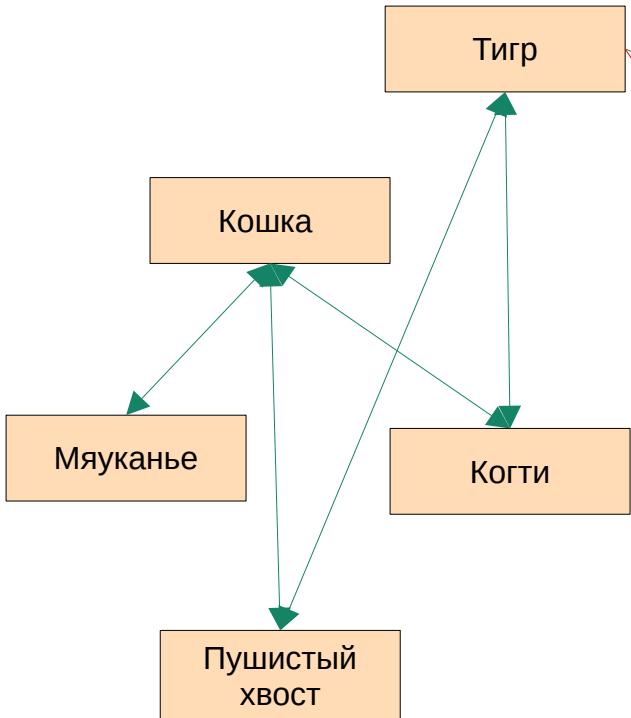
$$a'_t = \operatorname{argmax}_a(U, P)$$



По Тверски и Канеману,
большинство людей
выбирает a'_3 и s'_6
("синицу в руке")

Взаимосвязность переменных состояния

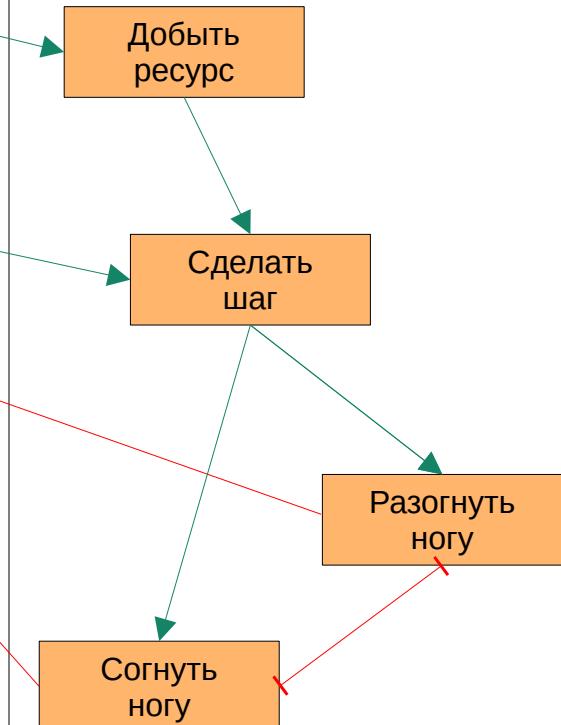
Переживаемые ощущения



Удовлетворяемые потребности



Совершаемые действия



Пространство потребностей/целей/ценностей

Психологические

Физиологические

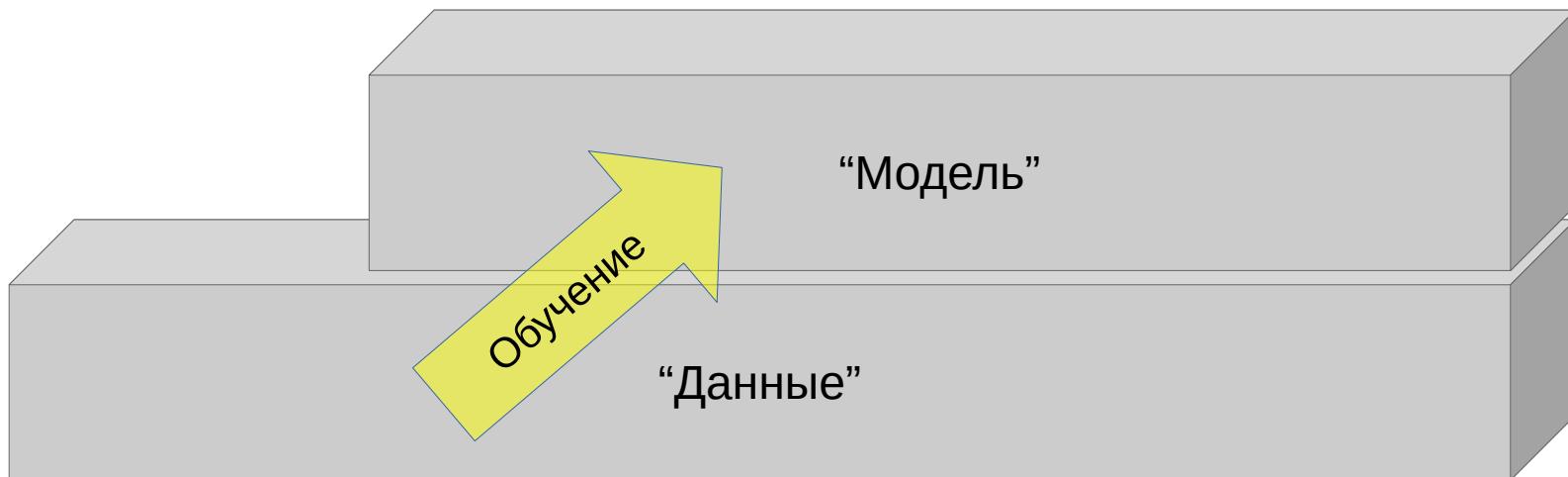


Бытовые
(Рутинные)

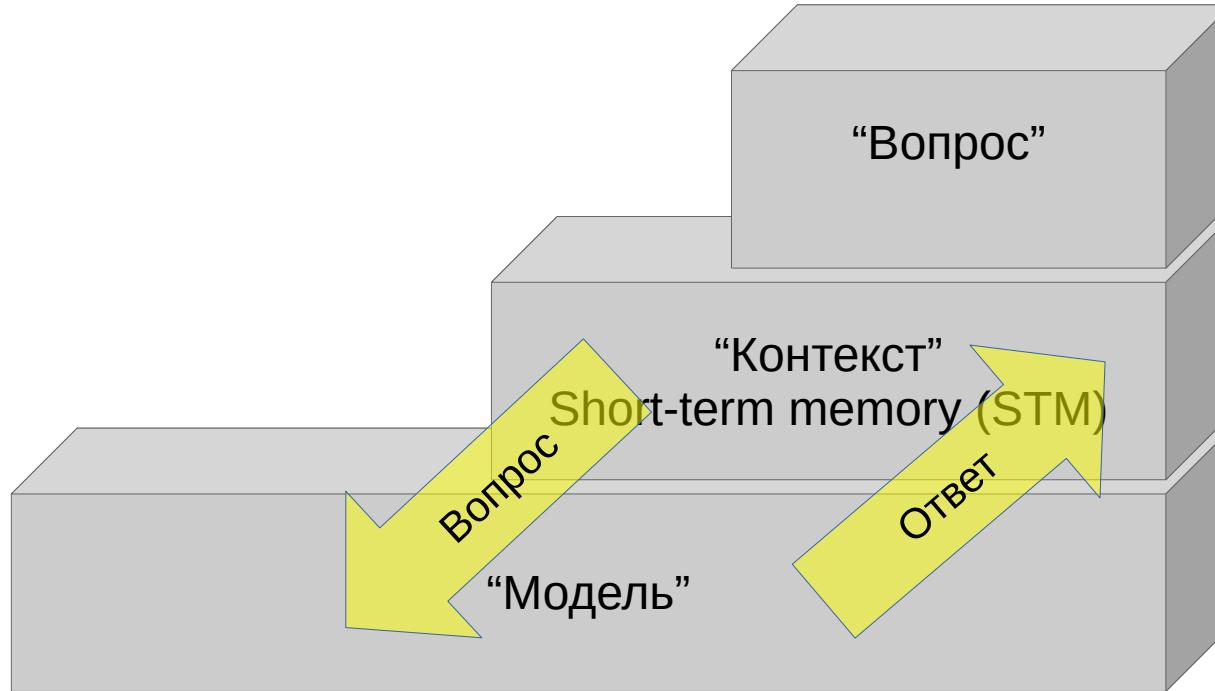
Бытийные
(Экзистенциальные)

А.Г.Колонин, В.Г.Крюков:
Вычислительная концепция психики,
Статья принята на конференцию
Нейроинформатика-25

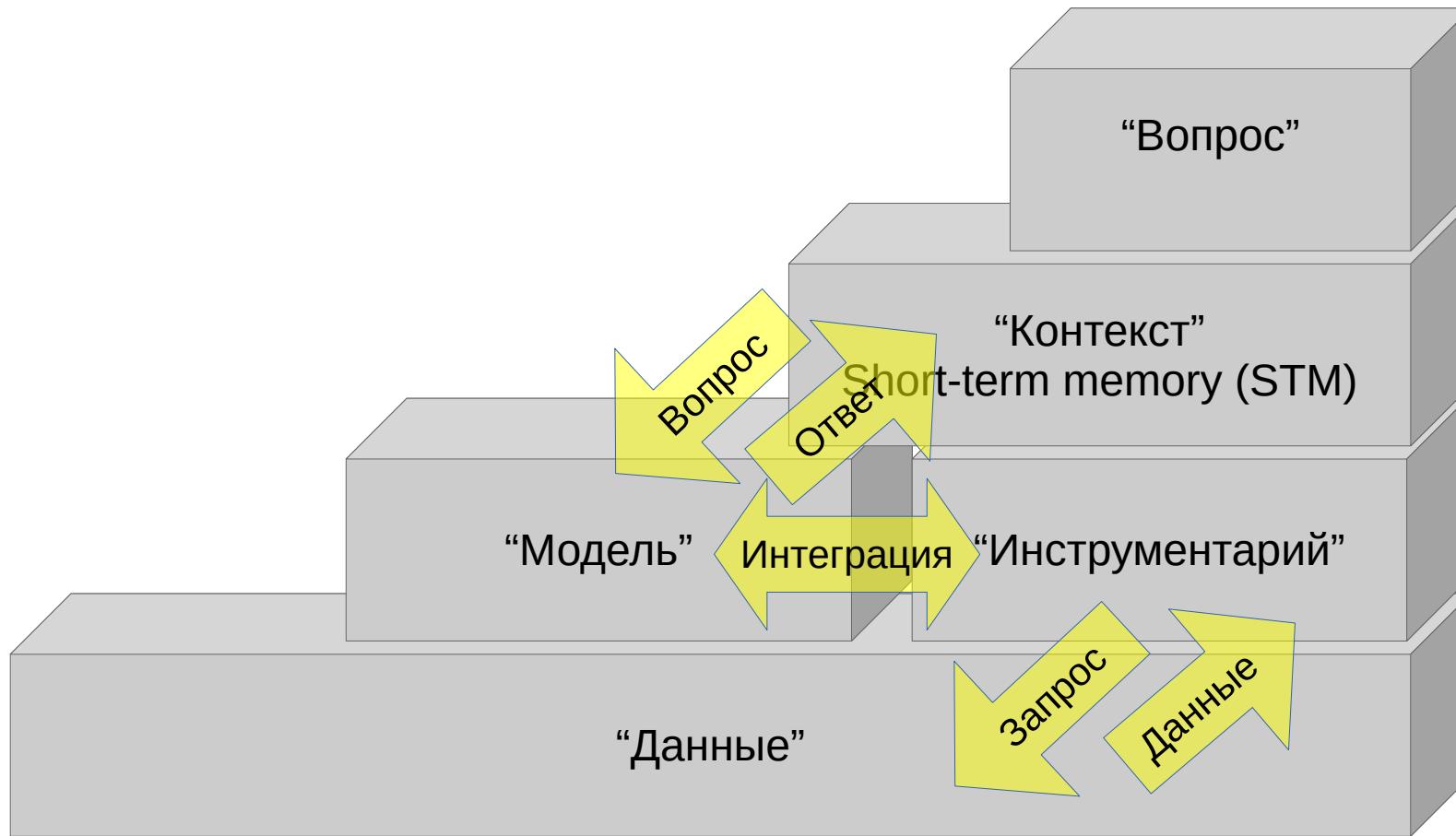
Обучение модели (БЯМ)



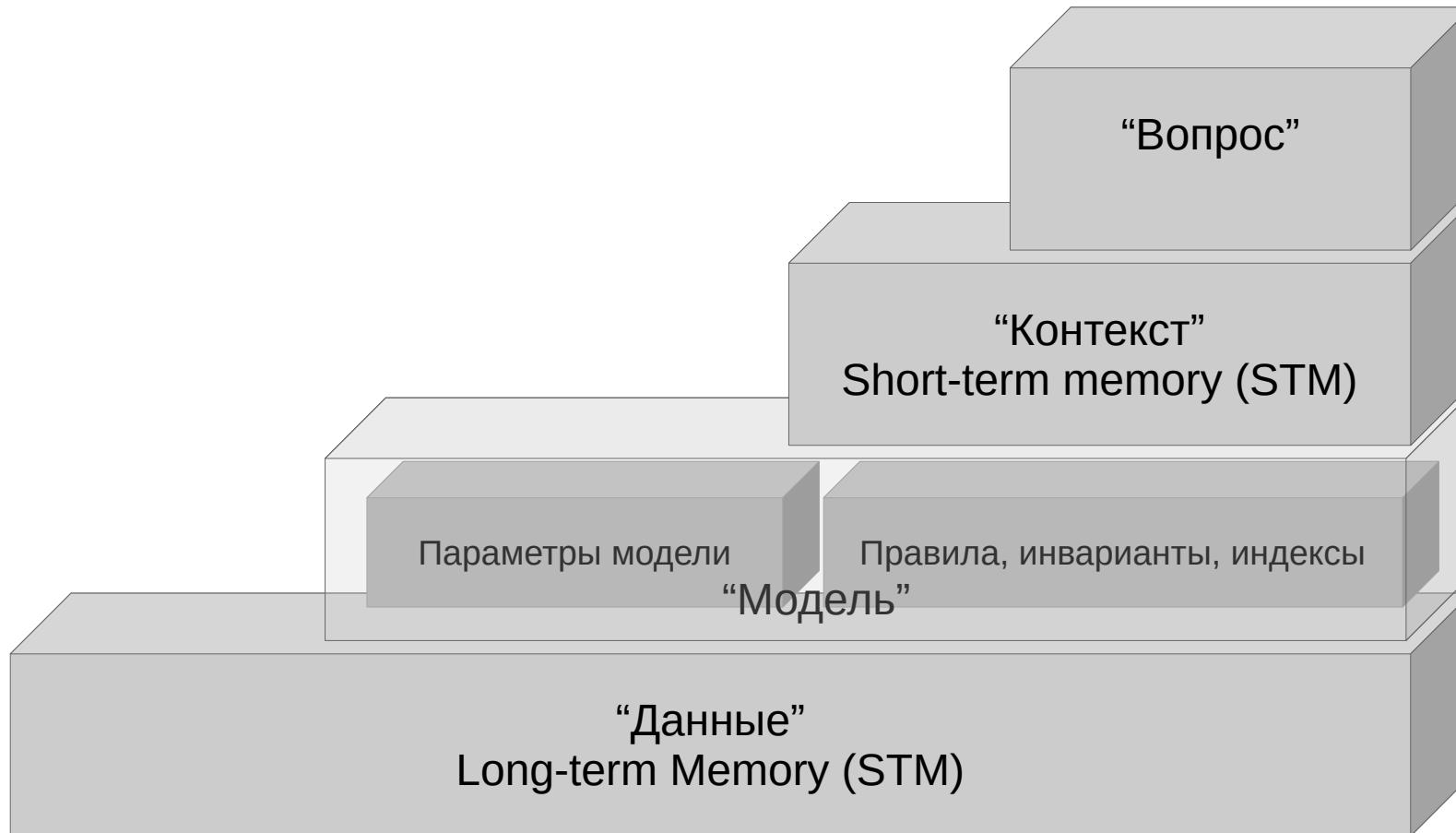
Использование модели (БЯМ)



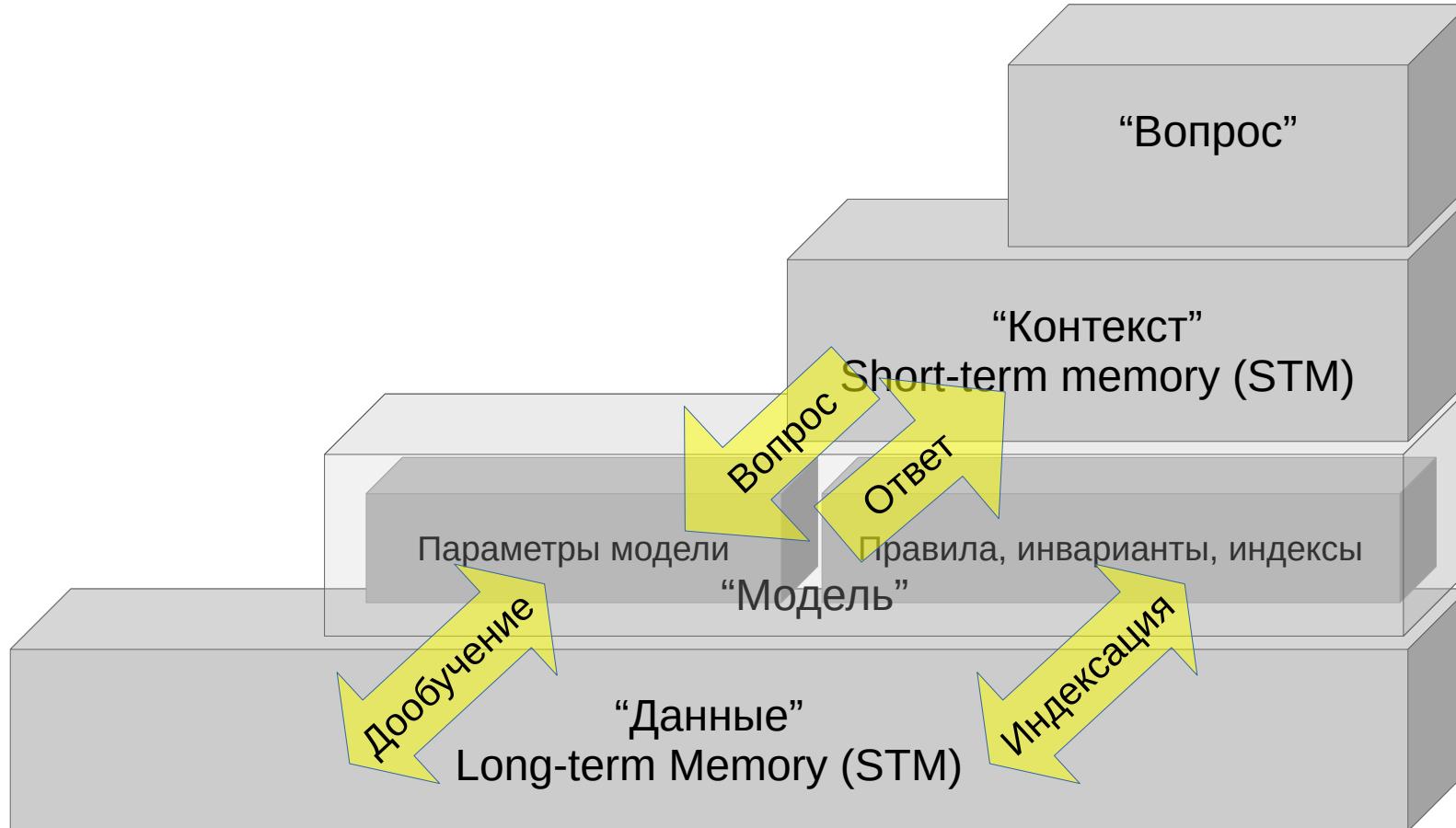
Много-агентные модели (БЯМ+RAG/GraphRAG)



Гибридная архитектура с долгосрочной памятью

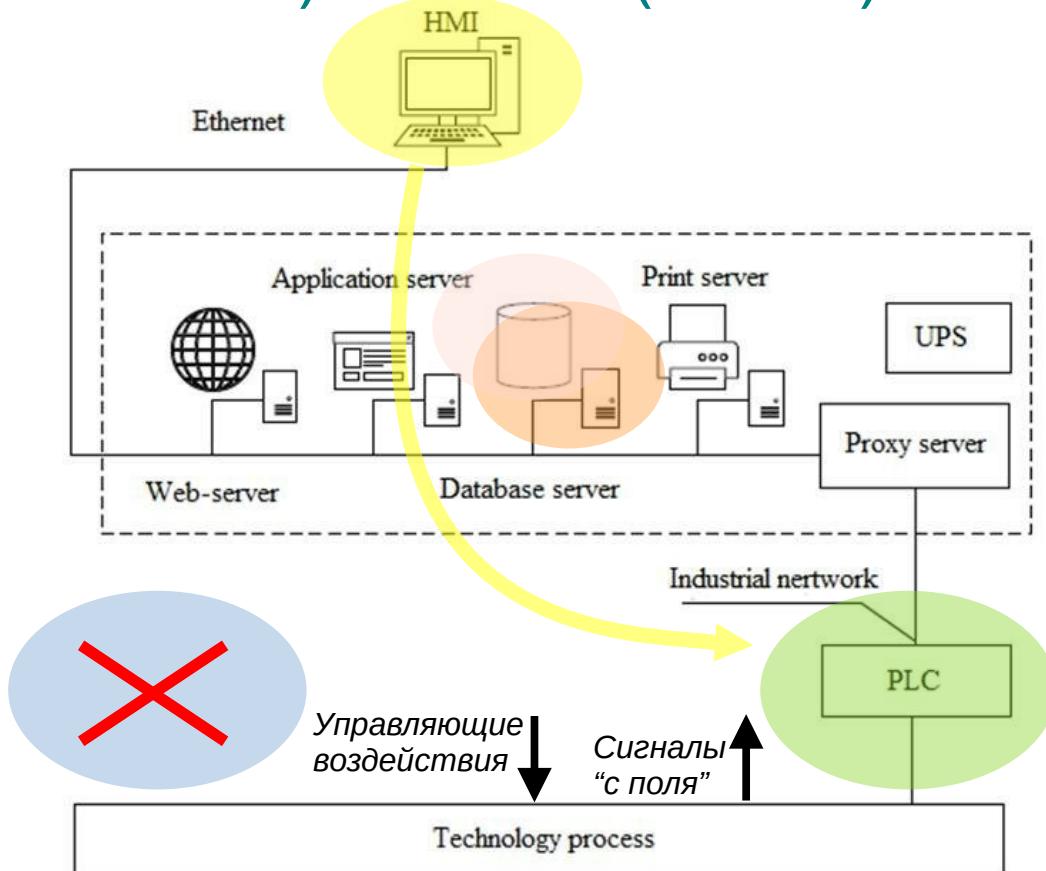


Гибридная архитектура с долгосрочной памятью



Случай применения

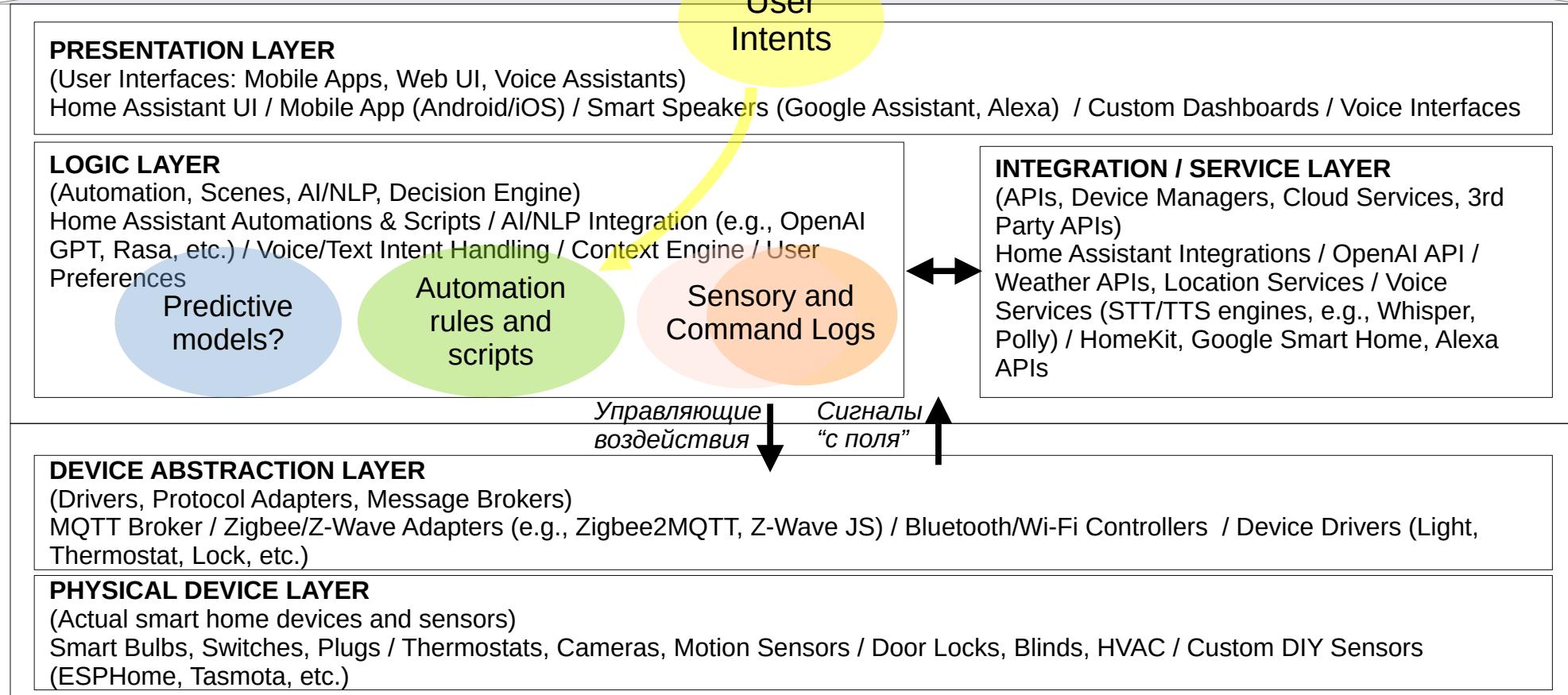
А) АСУ ТП (APCS)



Programmable logic
IEC 61131 (synchronous)
IEC 61499 (asynchronous)

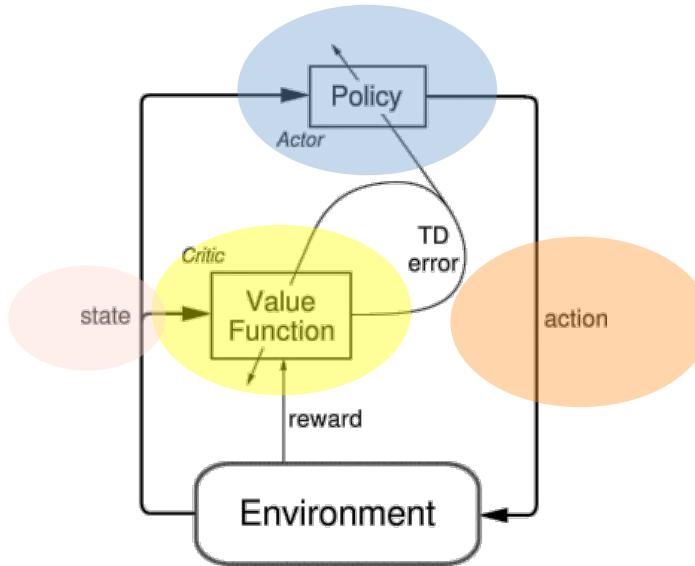
Случаи применения

В) “Умный Дом” (“Smart Home”)



Варианты реализации

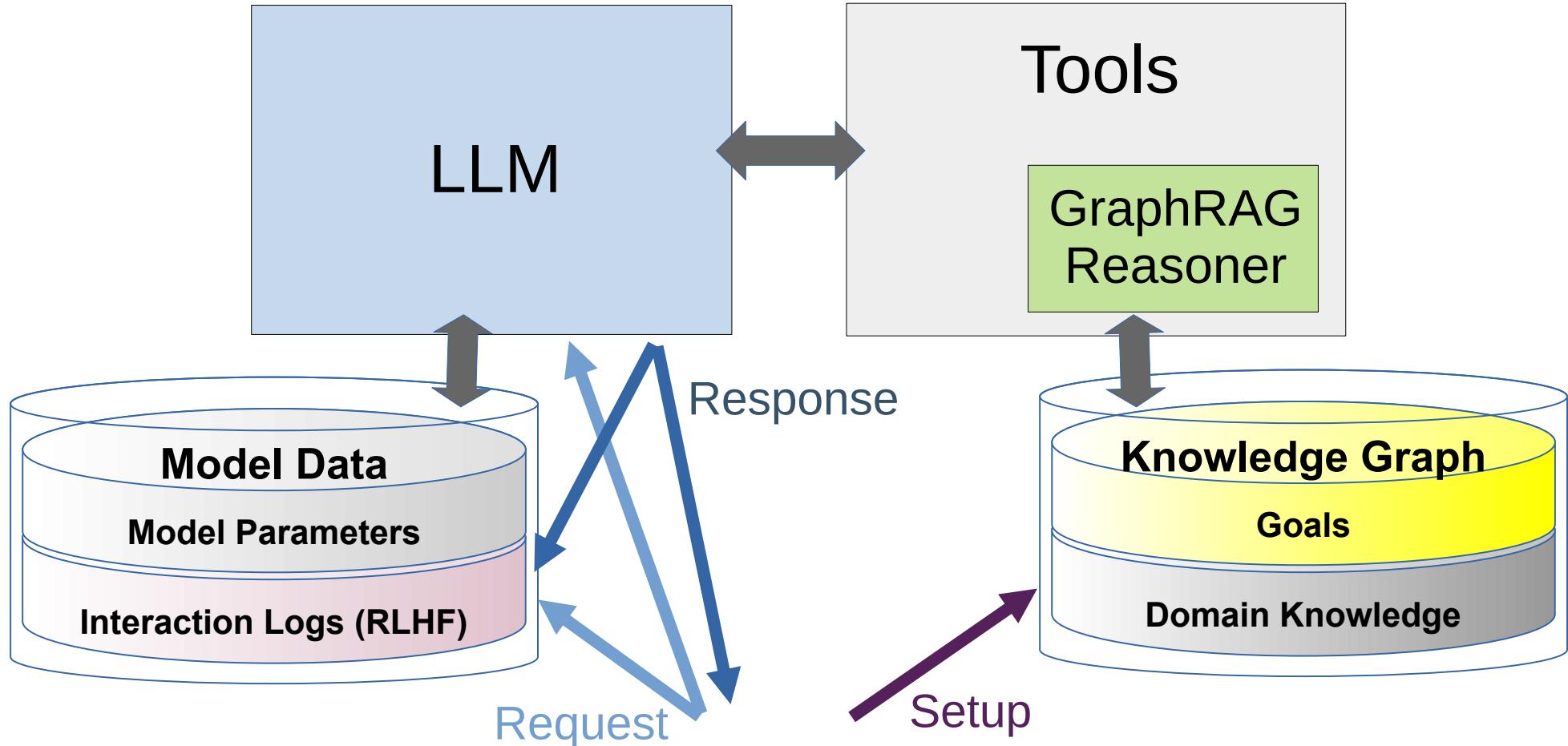
1) Обучение с подкреплением “актор-критик”



<https://medium.com/intro-to-artificial-intelligence/the-actor-critic-reinforcement-learning-algorithm-c8095a655c14>

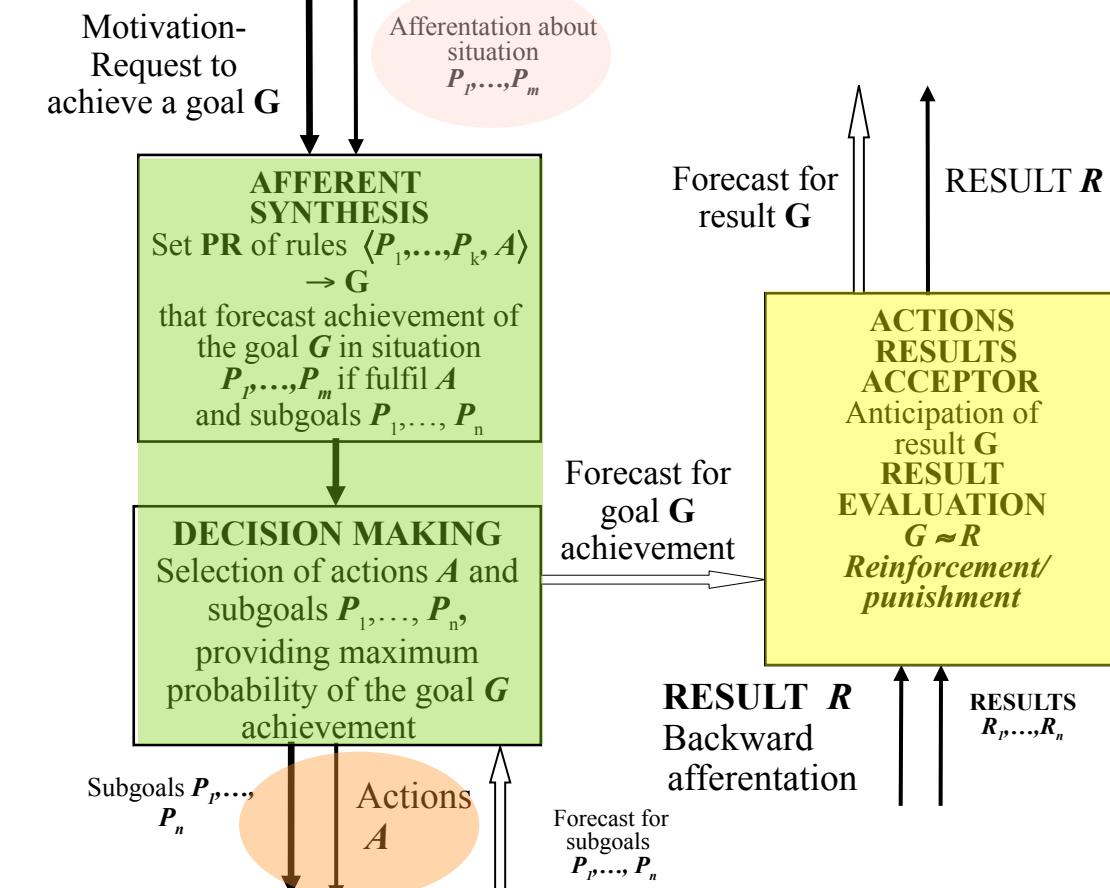
Варианты реализации

2) Когнитивная архитектура на основе LLM и GraphRAG



Варианты реализации

3) Когнитивная архитектура вероятностной логики



Copyright © 2025 Anton Kolonin, Aigents®

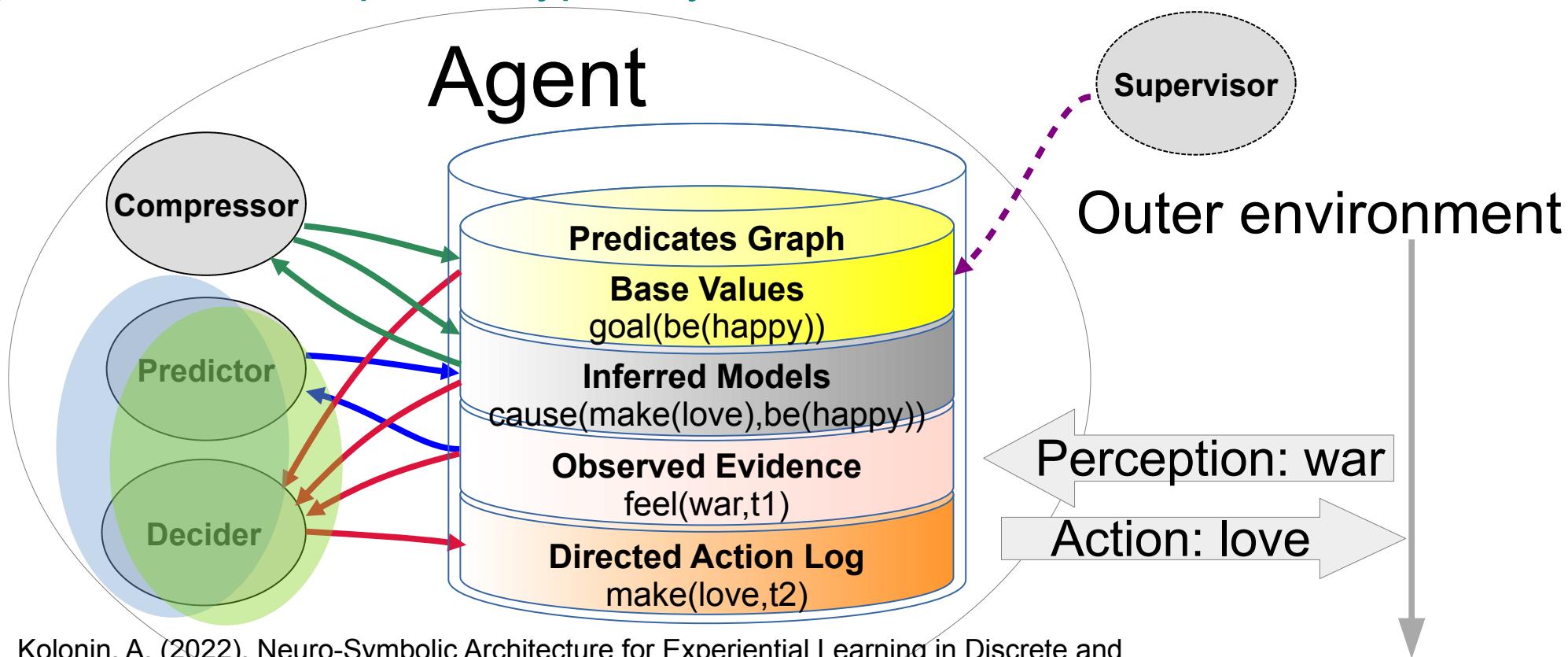
Evgenii Vityaev, Alexander Demin:
Adaptive Control of Modular Robots // Conference Paper in Advances in Intelligent Systems and Computing, Conference: First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures, Springer, August 2018

Evgenii E. Vityaev: Purposefulness as a Principle of Brain Activity // Anticipation: Learning from the Past, (ed.) M. Nadin. Cognitive Systems Monographs, V.25, Chapter No.: 13. Springer, 2015, pp. 231-254.

Витяев Е.Е. Логика работы мозга. Подходы к моделированию мышления. (сборник под ред. д.ф.-м.н. В.Г. Редько). УРСС Эдиториал, Москва, 2014г., стр. 120-153.

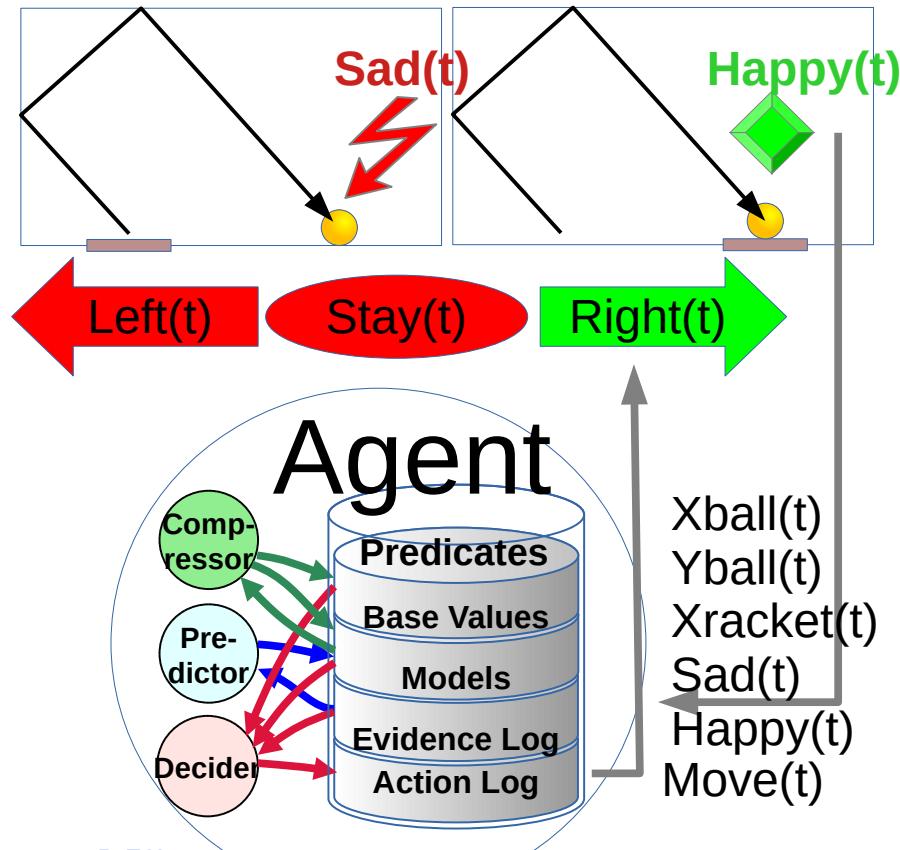
Варианты реализации

4) Когнитивная архитектура обучения на основе ценностей и опыта

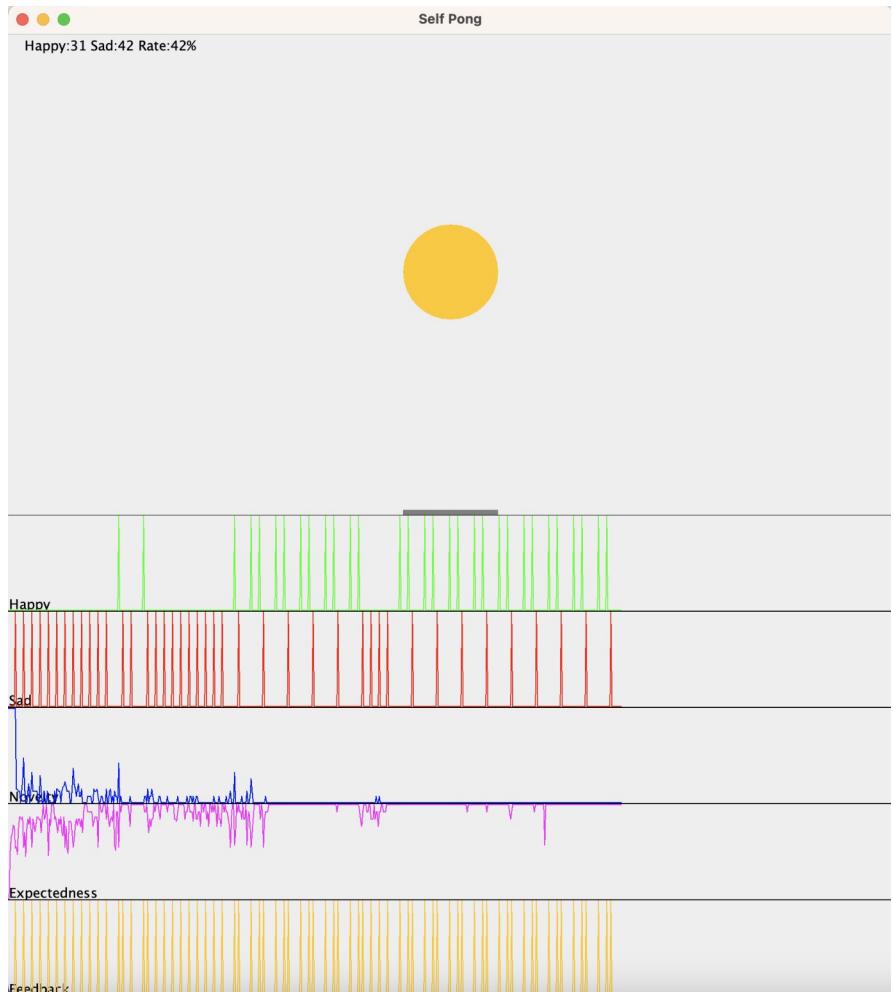


Kolonin, A. (2022). Neuro-Symbolic Architecture for Experiential Learning in Discrete and Functional Environments. In: Goertzel, B., Iklé, M., Potapov, A. (eds) Artificial General Intelligence. AGI 2021. Lecture Notes in Computer Science(), vol 13154. Springer, Cham.
https://doi.org/10.1007/978-3-030-93758-4_12

Когнитивная архитектура обучения на основе ценностей и опыта



А.Г.Колонин, В.Г.Крюков:
Вычислительная концепция психики,
Статья подана на конференцию
Нейроинформатика-25



Спасибо за внимание! Вопросы?

Антон Колонин

akolonin@aigents.com

Telegram: akolonin

Запись семинара по
теме доклада



Статья по теме доклада,
принятая на конференцию
Нейроинформатика-2025

