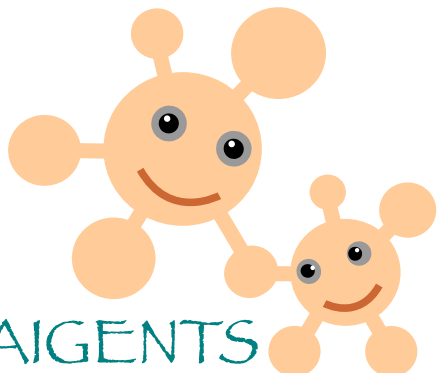
Computational Concept and Architecture of Artificial Psyche

Anton Kolonin
akolonin@aigents.com
Telegram: akolonin

Novosibirsk State University
*THE REAL SCIENCE
https://www.nsu.ru



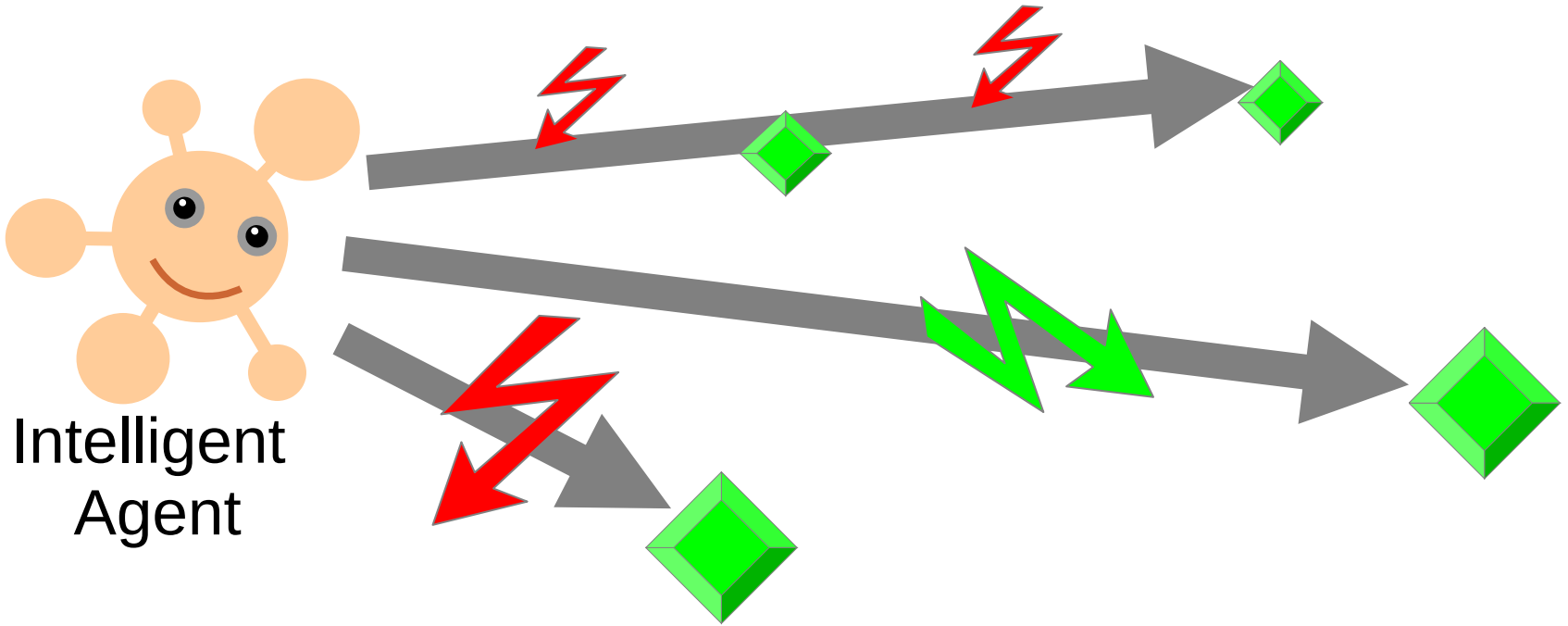
https://aigents.com

https://agirussia.org

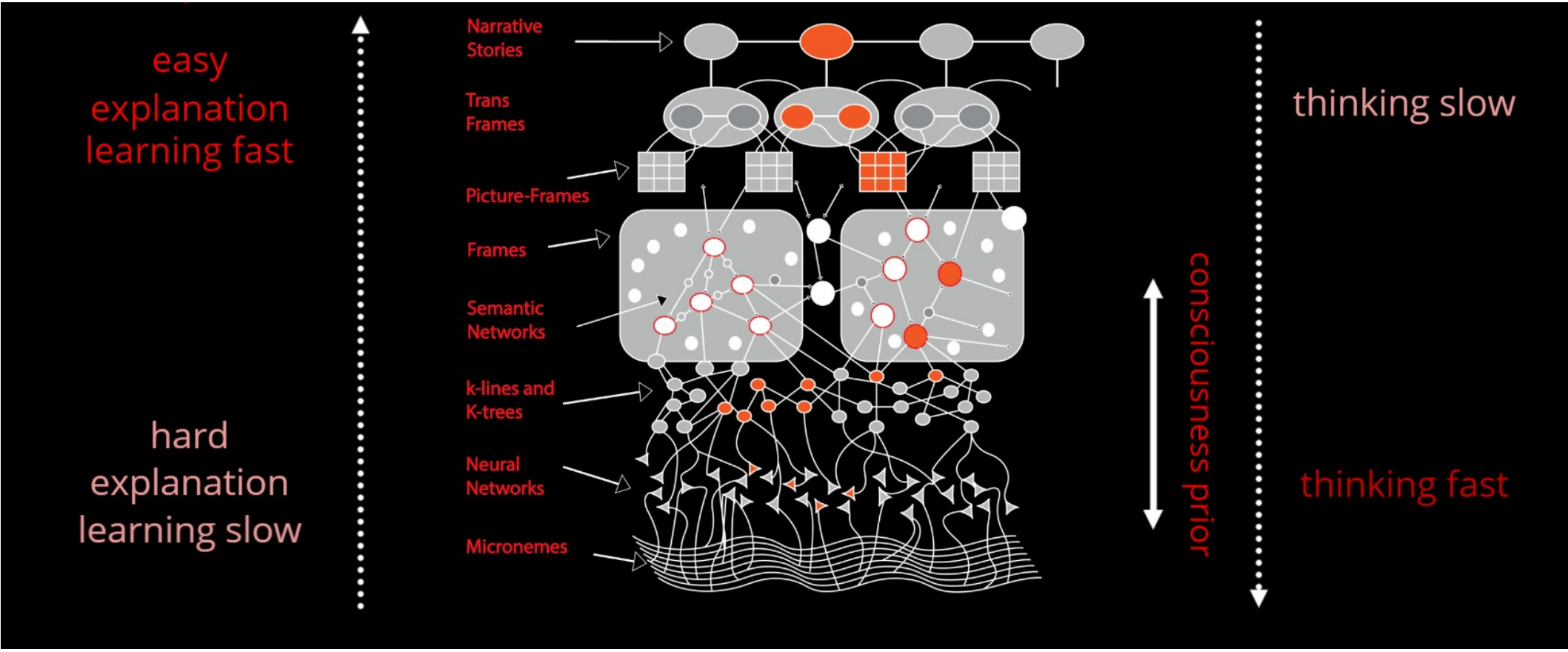
General Intelligence:

Reaching complex **goals** in different complex **environments**, using limited resources under uncertainty

(Ben Goertzel + Pei Wang + Shane Legg + Marcus Hutter)

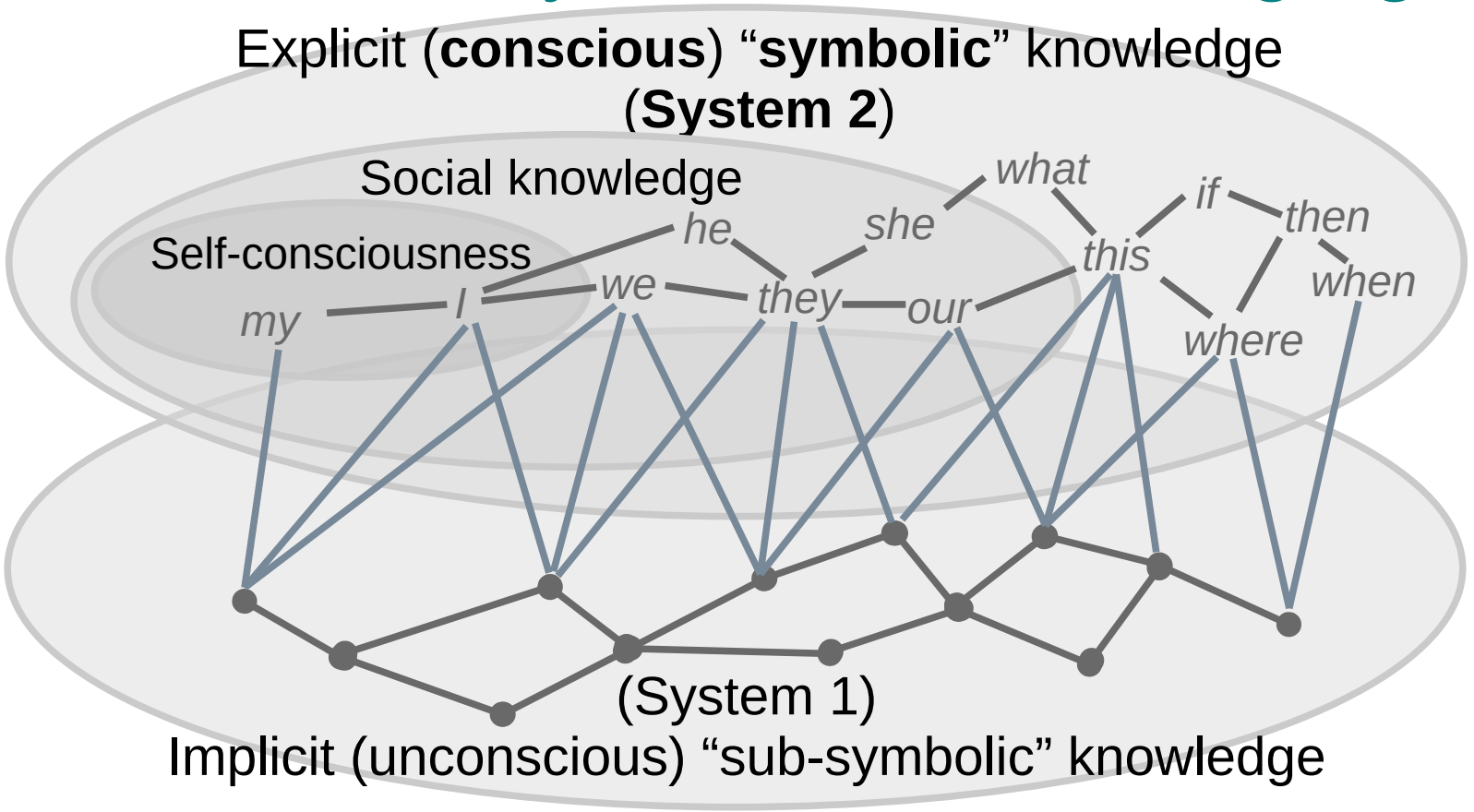


“Fast and Slow Thinking” – Daniel Kahneman



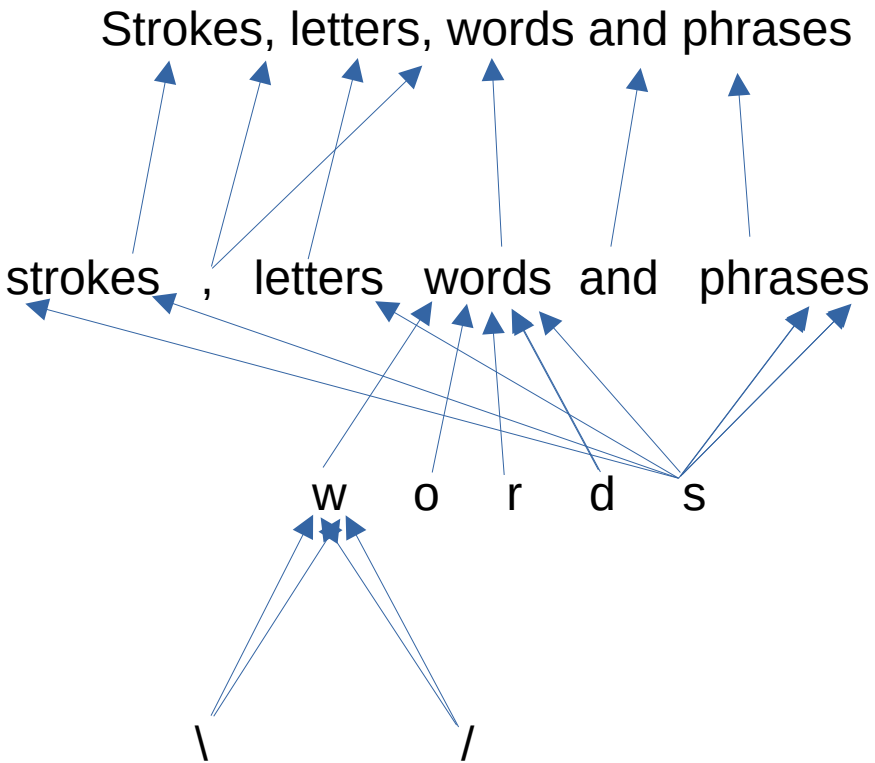
https://towardsdatascience.com/explainable-ai-vs-explaining-ai-part-1-d39ea5053347

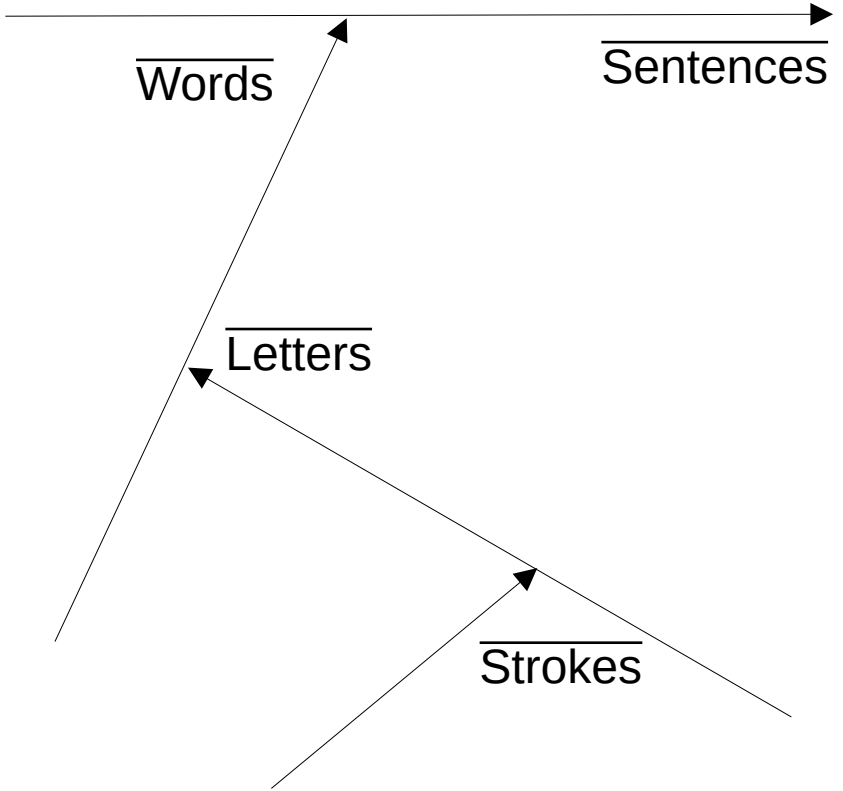
Medium: “neuro-symbolic” “knowledge graph”



https://www.amazon.com/Thinking-Fast-Slow-Daniel-Kahneman/dp/0374533555
https://amit02093.medium.com/atomspace-hyper-graph-information-retrieval-system-450cab9d751e

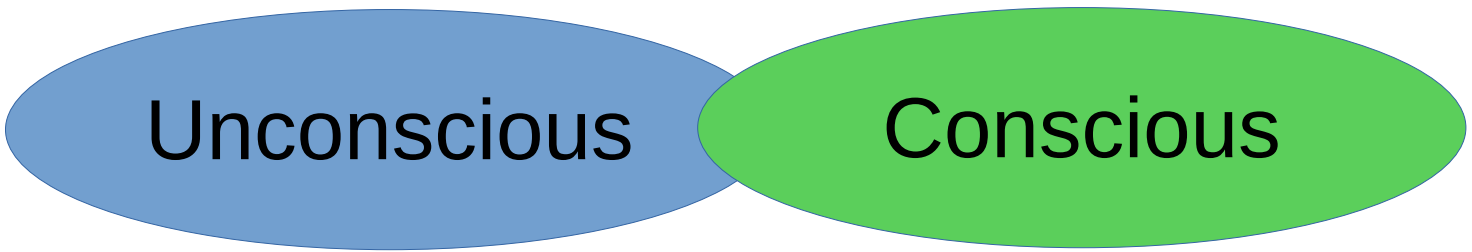
Functional equivalence of graph and ~~neural network~~ tensor models

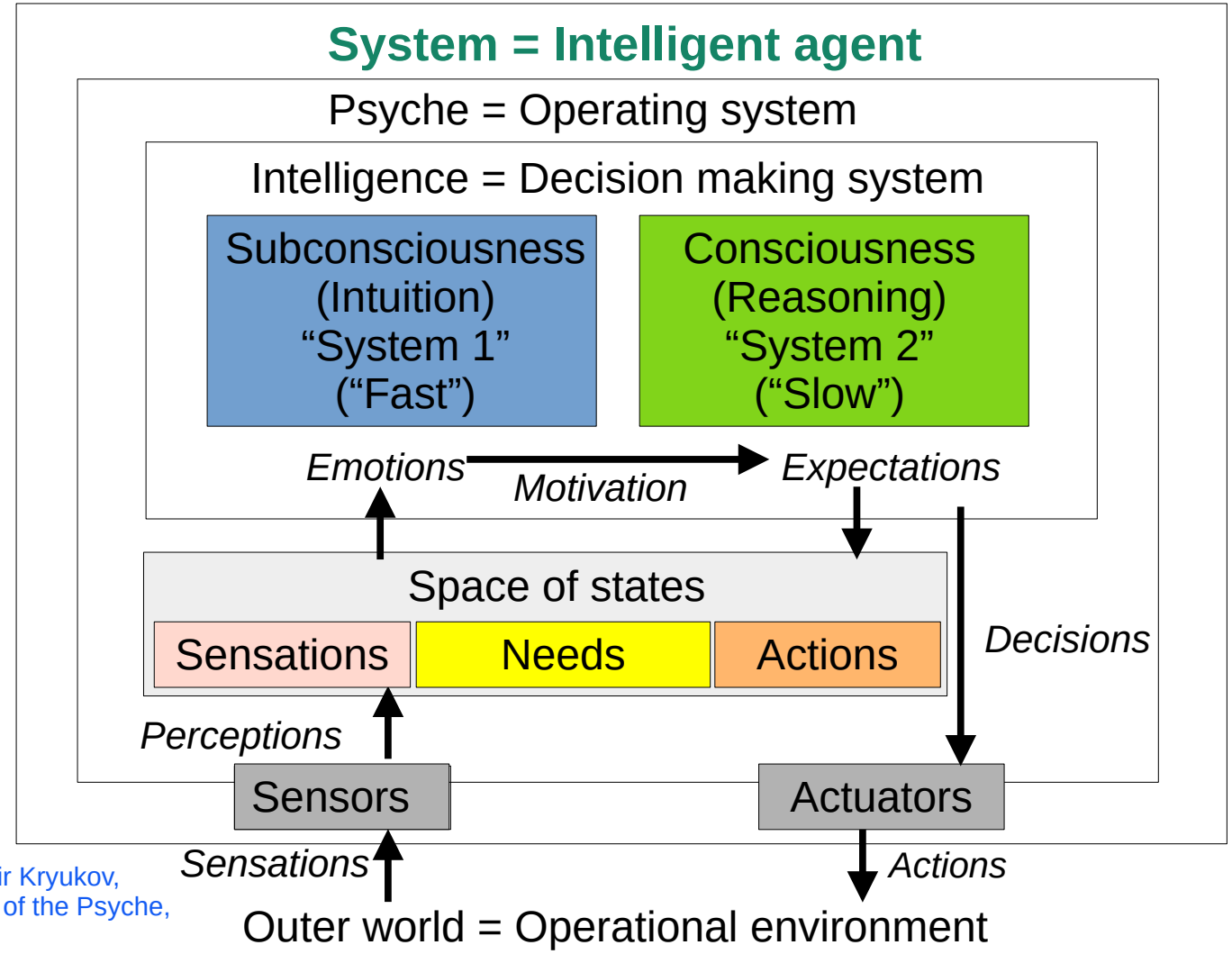




The psyche /ˈsaɪki/ is currently used to describe the totality of the human mind, conscious and unconscious.

https://en.wikipedia.org/wiki/Psyche





Anton Kolonin & Vladimir Kryukov,
Computational Concept of the Psyche,
Neuroinformatics-2025
https://arxiv.org/pdf/2509.07009

Psyche = Operating system

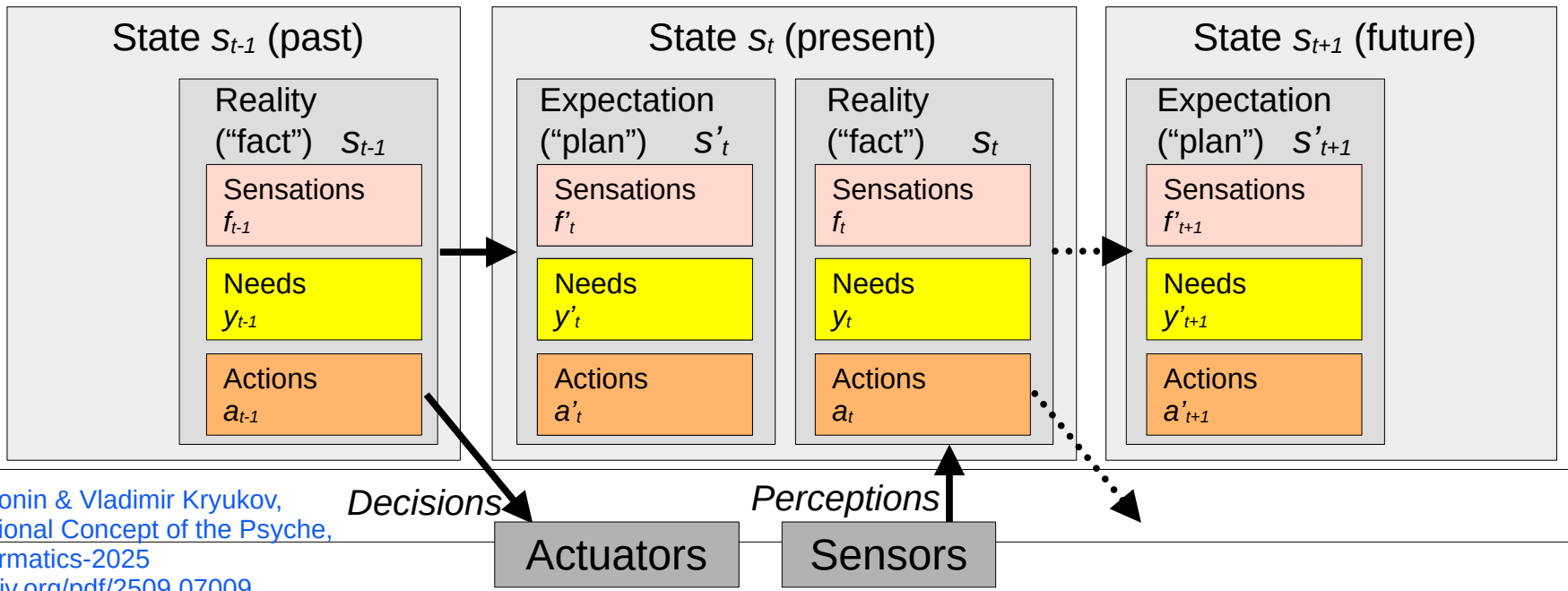
Intelligence = Decision making system

Models *s* ("invariants") of states with utilities *U* and probabilities *P* of transitions

$U(\{s_\tau\}_{\tau\in\{-T,-1\}},s'_0) = L(\ x\cdot(y_t - y_{t+1}),\ s'_t - s_t,\ E(a_t))$ $\qquad s'_t=argmax_s(U(\{s_\tau\}_{\tau\in\{t-T,t-1\}},s'_t),P(\{s_\tau\}_{\tau\in\{t-T,t-1\}},s'_t))$

Experiential learning ↑ ↓ *Decision making*

Space of states and episodic memory ("precedents")



Anton Kolonin & Vladimir Kryukov,
Computational Concept of the Psyche,
Neuroinformatics-2025
https://arxiv.org/pdf/2509.07009

Psyche = Operating system

Energy-efficiency *of a*

Optimal decision and expectation of *s*

Predictability *of s*

Expected utility of *s*

Intelligence = Decision making system

Models *s* ("invariants") of states with utilities *U* and probabilities *P* of transitions

$U(\{s_\tau\}_{\tau\in\{-T,-1\}}, s'_0) = L(x\cdot(y_t - y_{t+1}), s'_t - s_t, E(a_t))$ $\quad s'_t = argmax_s(U(\{s_\tau\}_{\tau\in\{t-T,t-1\}}, s'_t), P(\{s_\tau\}_{\tau\in\{t-T,t-1\}}, s'_t))$

Satisfaction *of y*

Expected probability of *s*

Experiential learning ↑ ↓ *Decision making*

Space of states and episodic memory ("precedents")

State s_{t-1} (past)	State s_t (present)		State s_{t+1} (future)
Reality ("fact") s_{t-1}	Expectation ("plan") s'_t	Reality ("fact") s_t	Expectation ("plan") s'_{t+1}
Sensations f_{t-1}	Sensations f'_t	Sensations f_t	Sensations f'_{t+1}
Needs y_{t-1}	Needs y'_t	Needs y_t	Needs y'_{t+1}
Actions a_{t-1}	Actions a'_t	Actions a_t	Actions a'_{t+1}

Decisions → Actuators

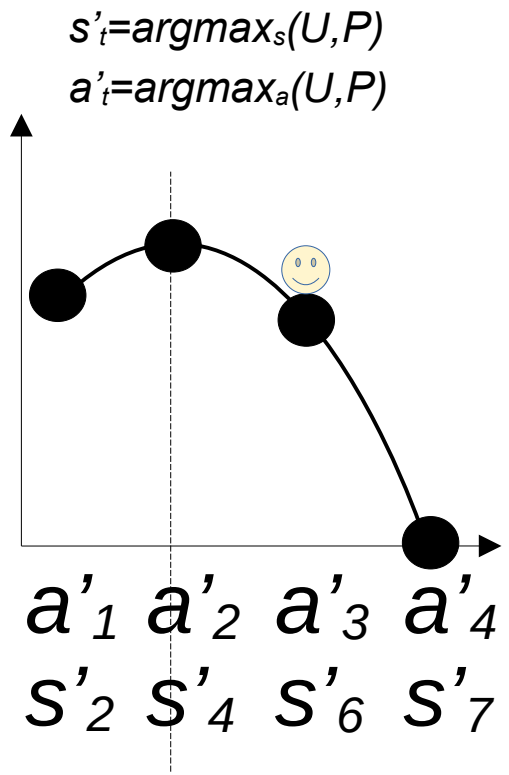
Perceptions ← Sensors

$x\cdot y_t$ – "motivation vector"

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)

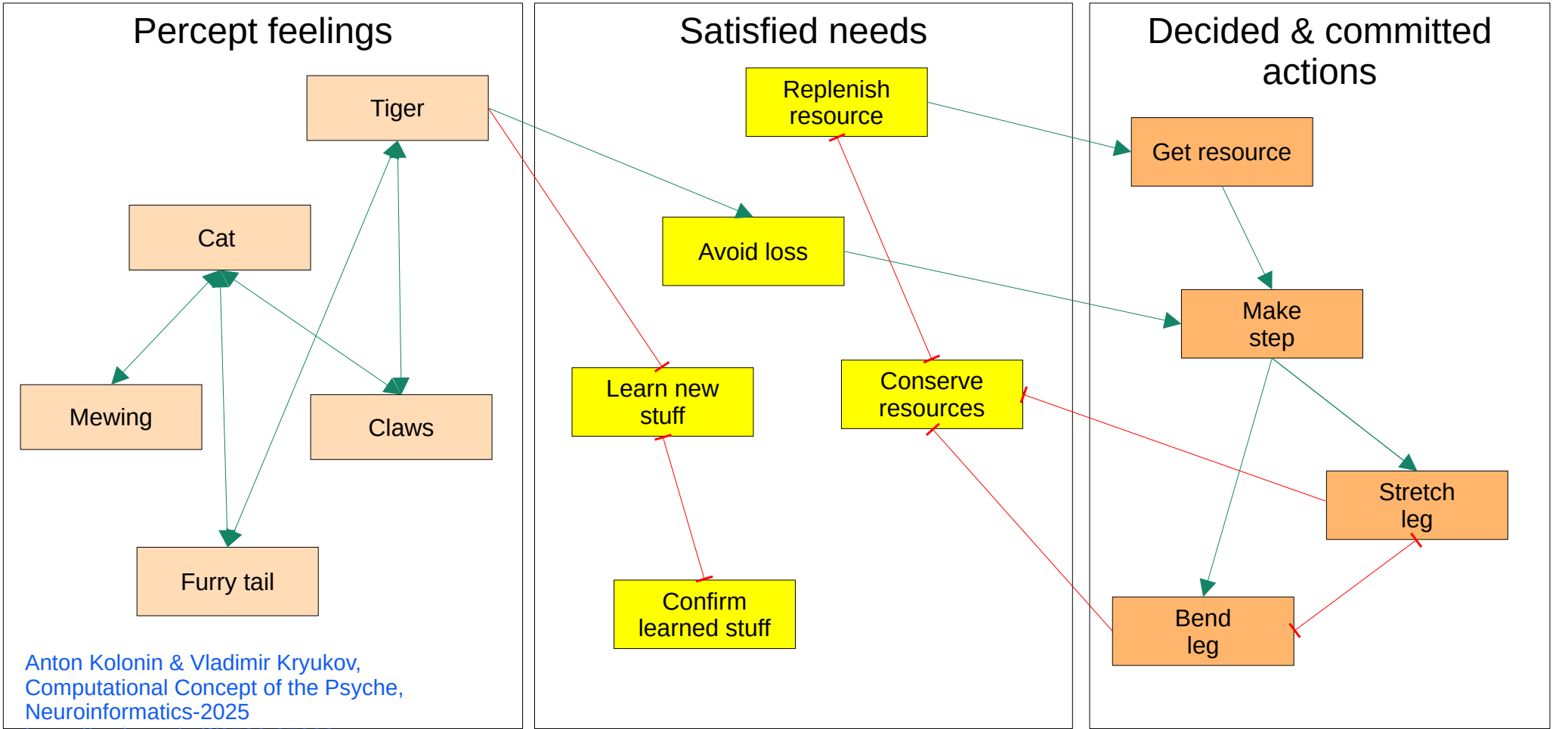
Decision making as operational risk management

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



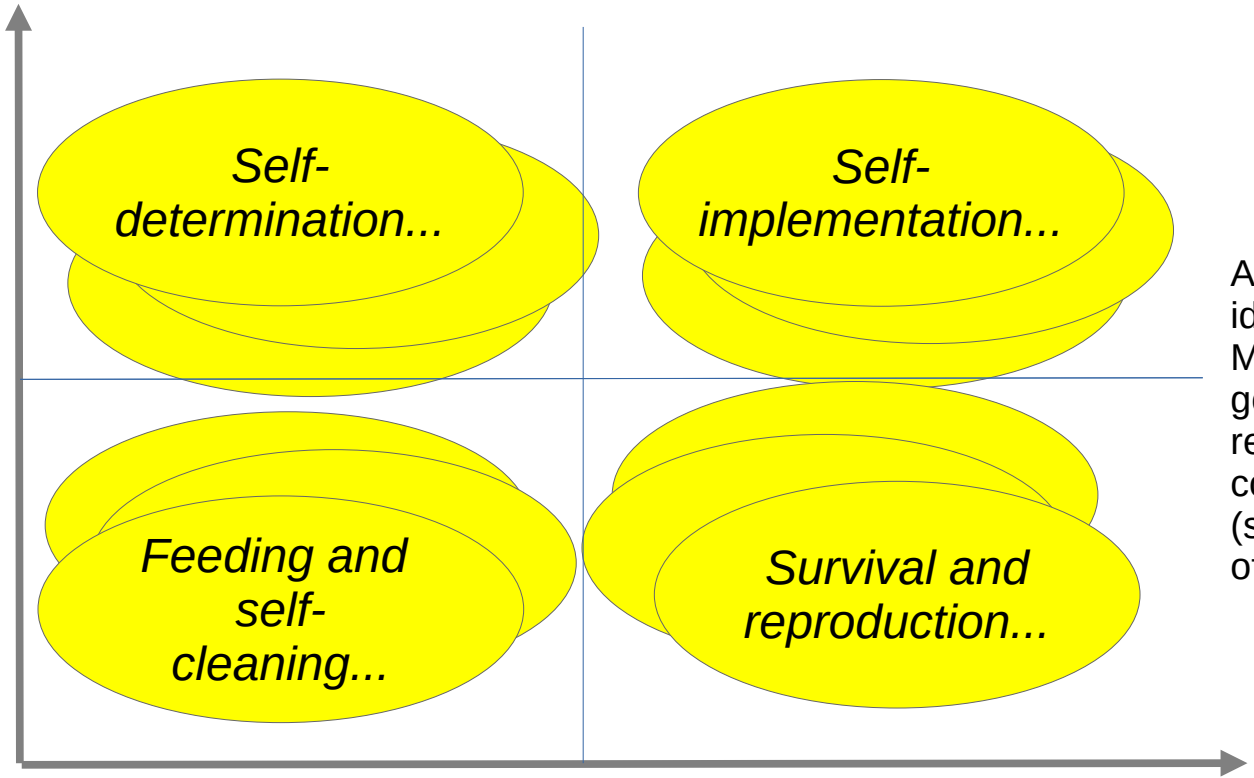
Tversky & Kahneman: most people choose a'_3 и s'_6 *("smaller profit with greater reliability")*

Mutual dependency of state variable ~~subgraphs~~ tensors



Anton Kolonin & Vladimir Kryukov,
Computational Concept of the Psyche,
Neuroinformatics-2025
https://arxiv.org/pdf/2509.07009

Space of needs/goals/values

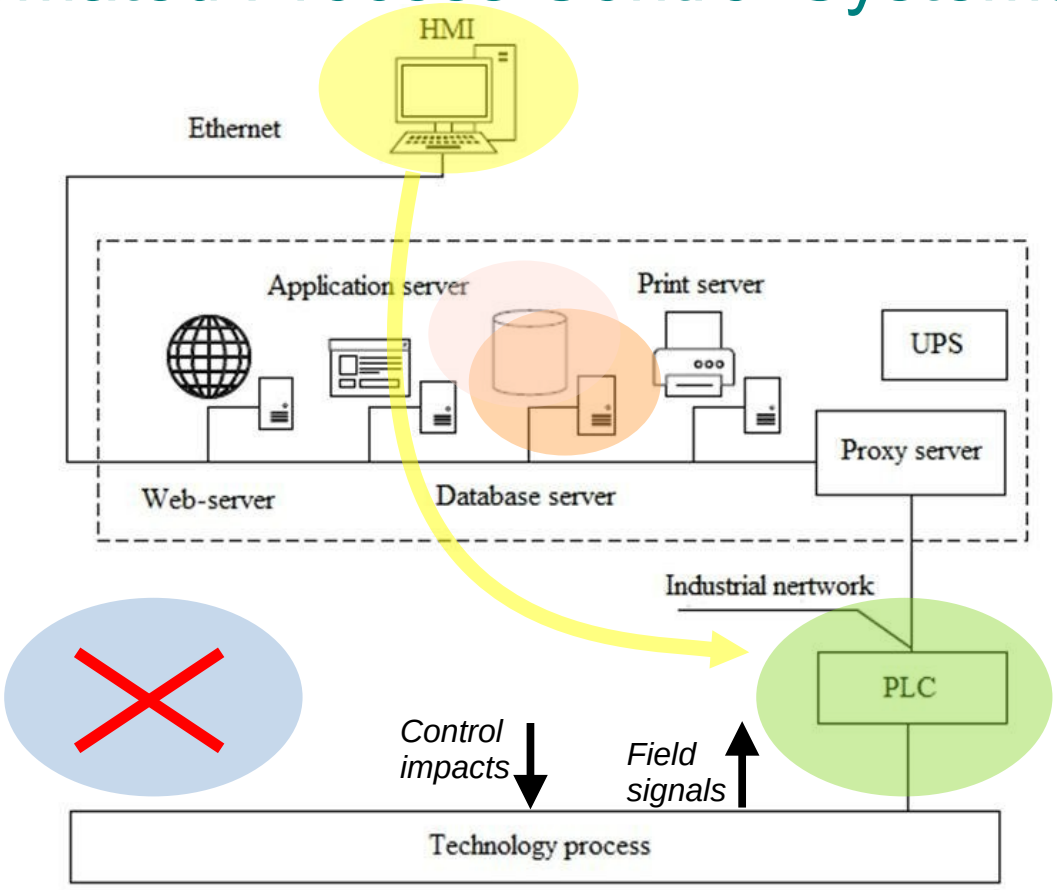


According to the ideas of A. Maslow, any goal can be represented as a combination (superposition) of basic needs

Anton Kolonin & Vladimir Kryukov, Computational Concept of the Psyche, Neuroinformatics-2025 https://arxiv.org/pdf/2509.07009

Application cases

A) Automated Process Control Systems (APCS)



Programmable logic

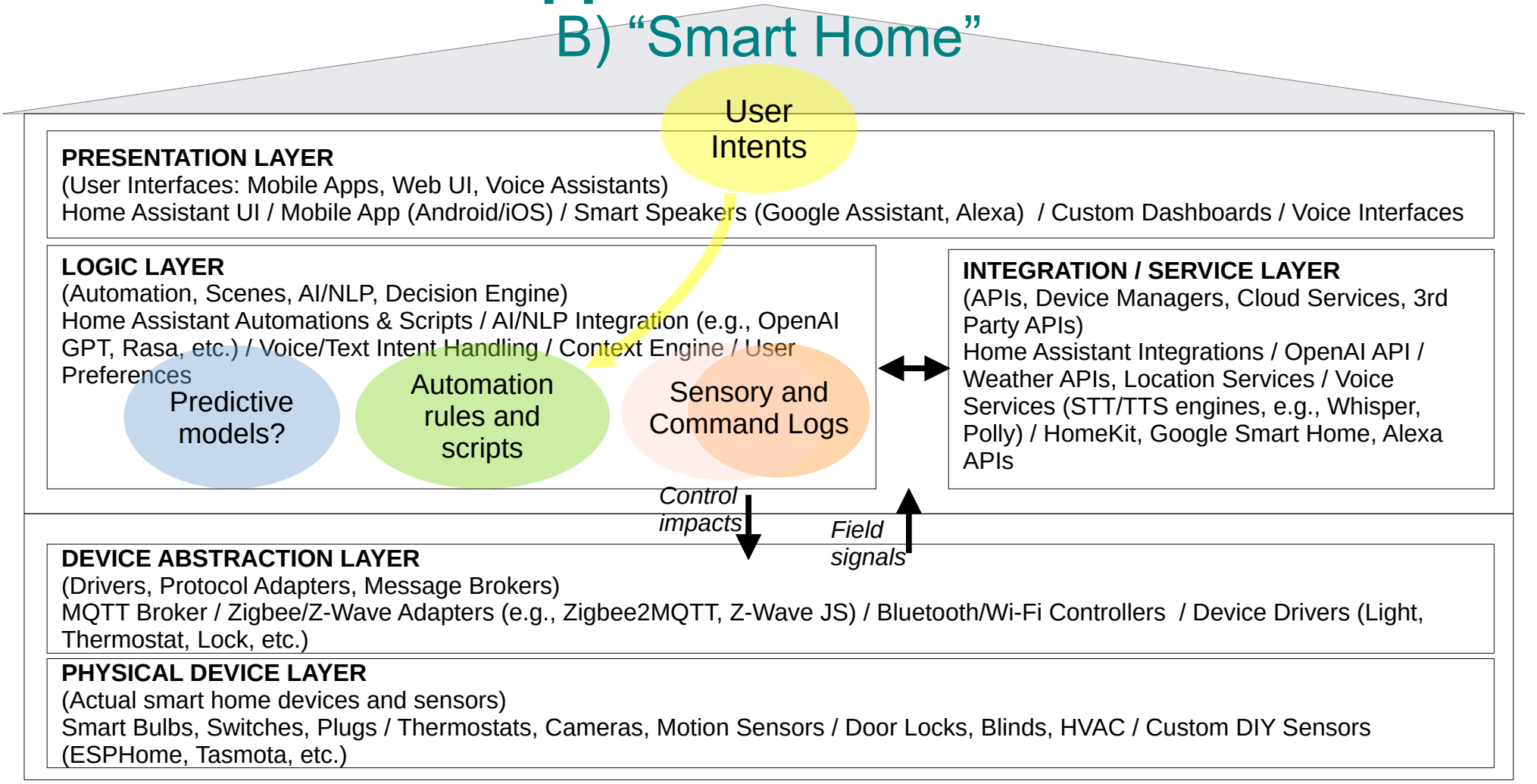
IEC 61131 (synchronous)

IEC 61499 (asynchronous)

https://www.researchgate.net/publication/311662442_Adaptive_Intelligent_Manufacturing_Control_Systems

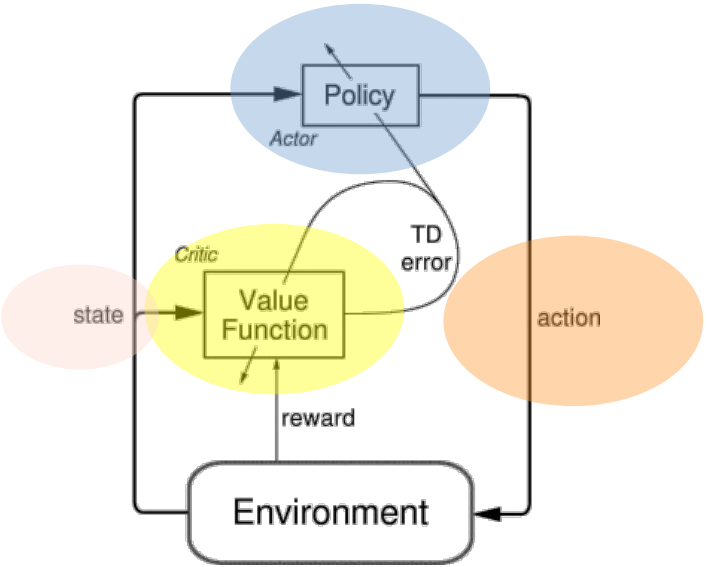
Application cases

B) “Smart Home”



Implementation options

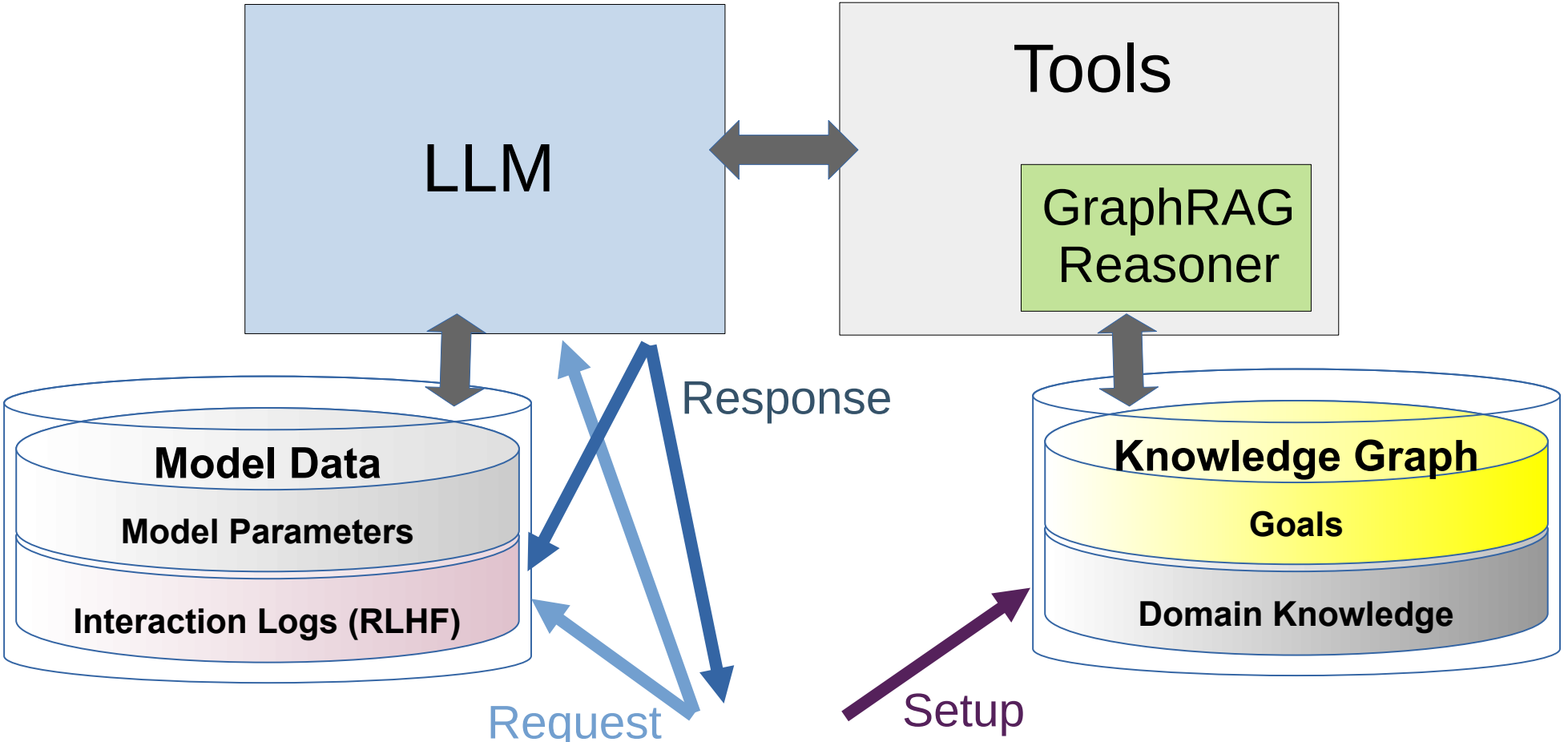
1) Reinforcement learning based on “actor-critic” model



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

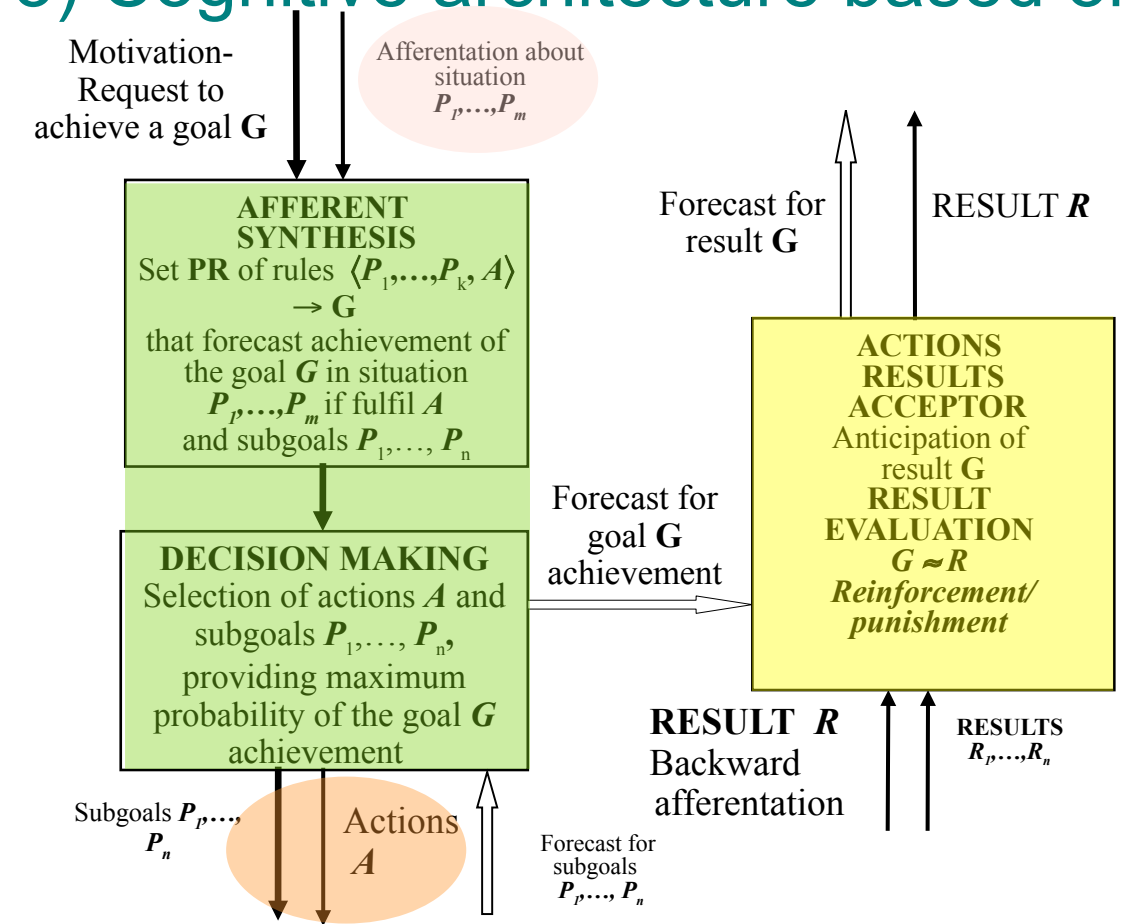
Implementation options

2) Cognitive architecture based on LLM and (active) GraphRAG



Implementation options

3) Cognitive architecture based on probabilistic logic



$$Pr\,ob(G \mid P_1,\ldots,P_k,P_1,\ldots,P_n,A) = Pr\,ob(rule)\cdot Pr\,ob(P_1)\cdot\ldots\cdot Pr\,ob(P_n)$$

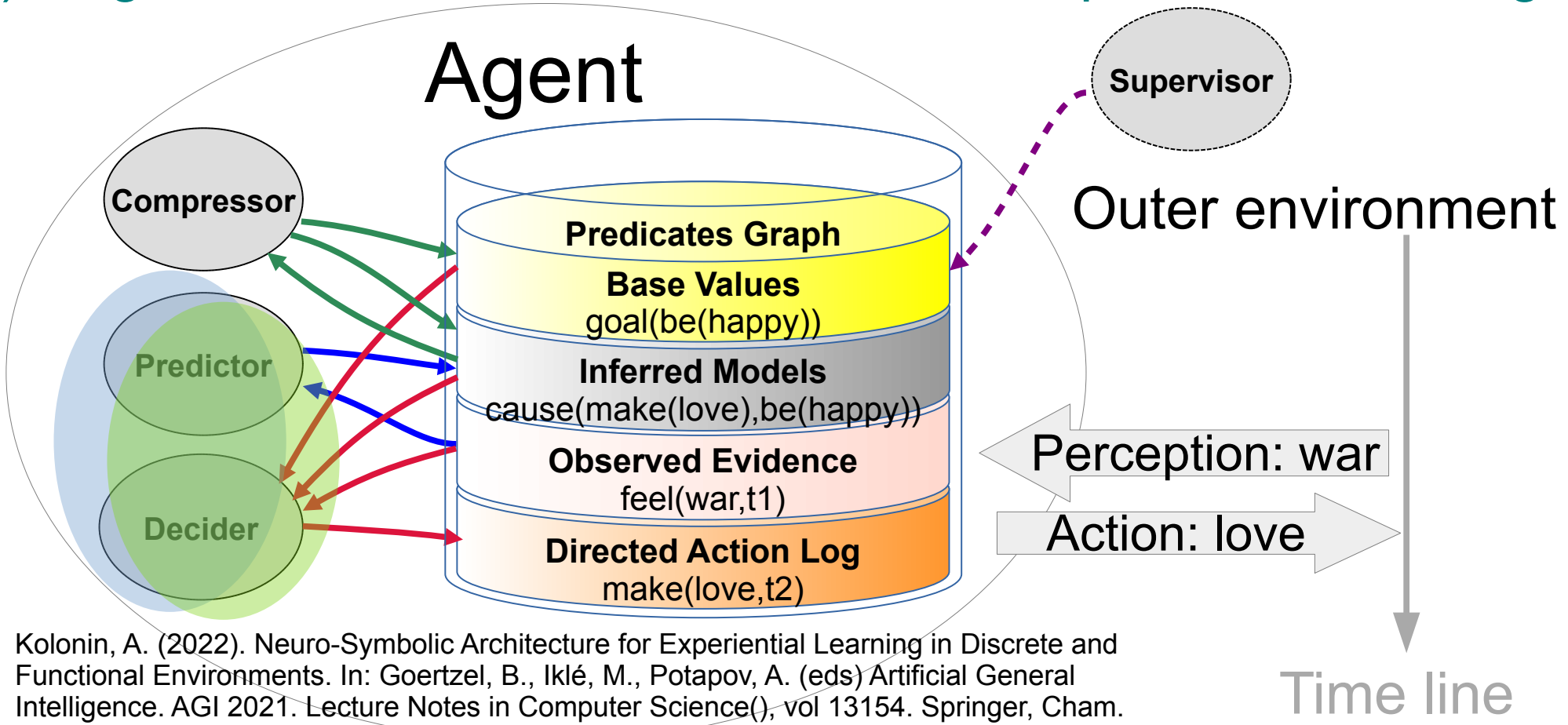
Evgenii Vityaev, Alexander Demin: Adaptive Control of Modular Robots // Conference Paper in Advances in Intelligent Systems and Computing, Conference: First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures, Springer, August 2018

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.

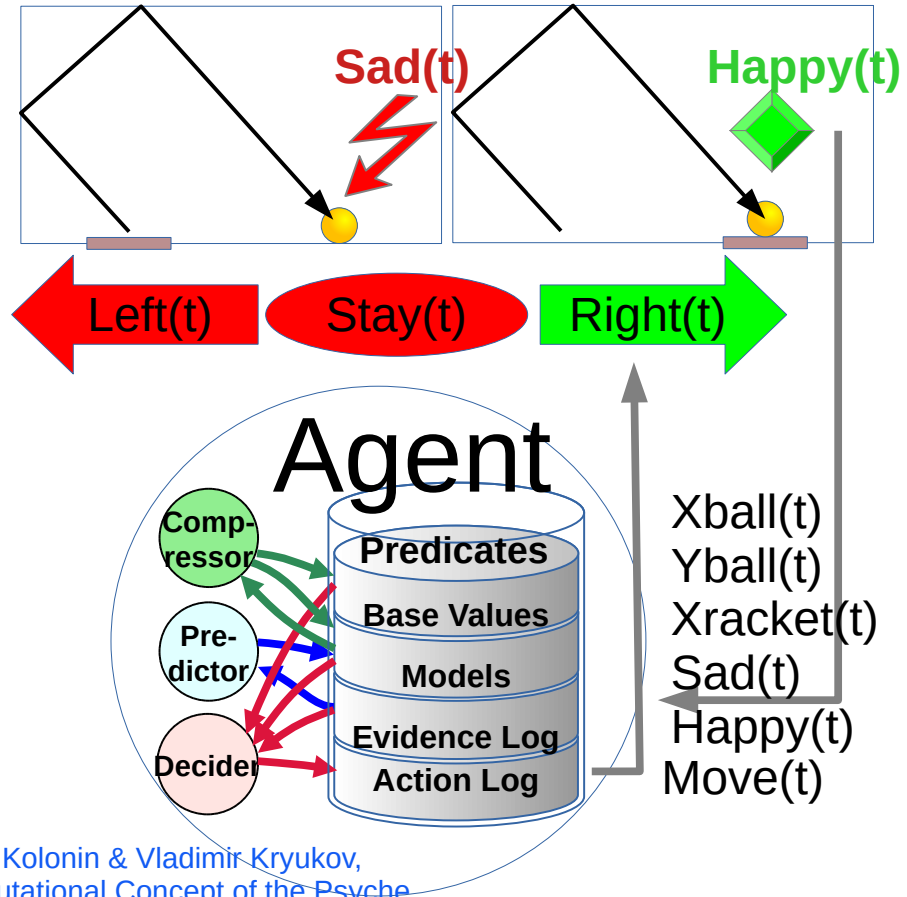
Implementation options

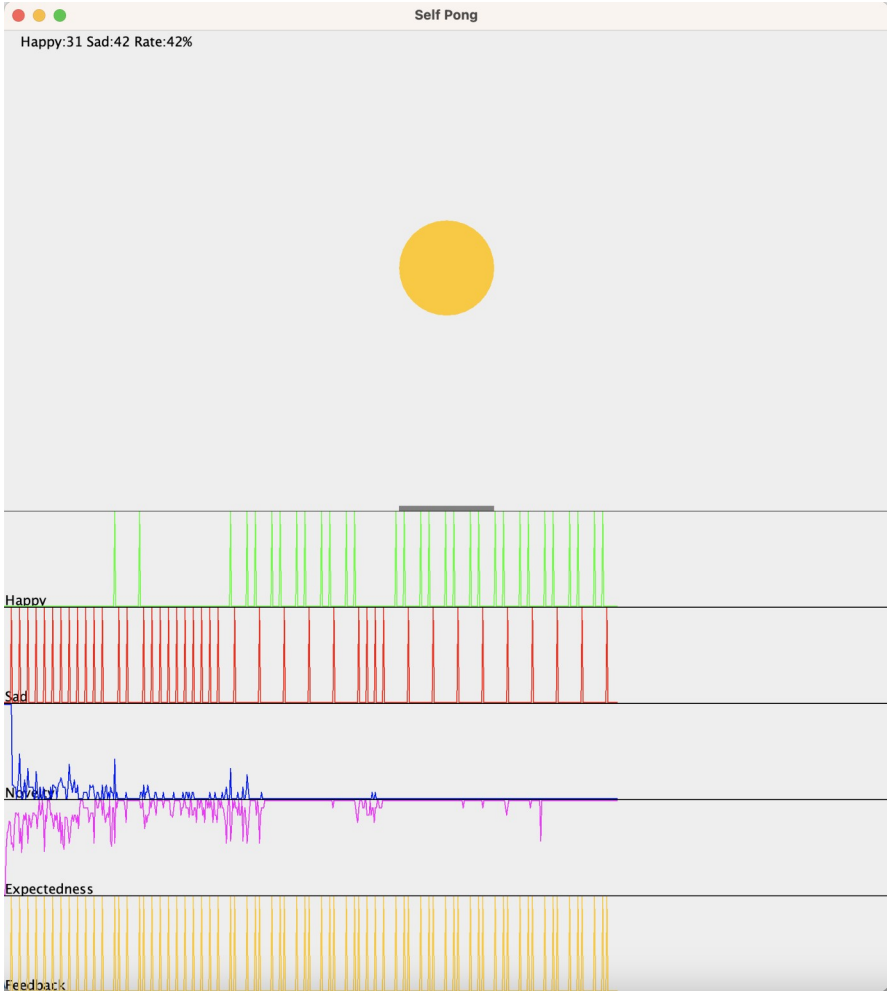
4) Cognitive architecture of value-based experiential learning



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

Cognitive architecture of value-based experiential learning





Anton Kolonin & Vladimir Kryukov,
Computational Concept of the Psyche,
Neuroinformatics-2025
https://arxiv.org/pdf/2509.07009

Thank you for attention! Questions?

Anton Kolonin
akolonin@aigents.com
Telegram/GitHub: akolonin

Workshop recording on the subject

Anton Kolonin & Vladimir Kryukov, Computational Concept of the Psyche, Neuroinformatics-2025