4. Kolombet V.V. Website golosaVgolove.ru on the Internet.

5. Kolombet V.A., Lesnykh V.N., Elistratov A.V. (2023) Psychoanalysis as a Biophysical Method // Proceedings of the Institute of Engineering Physics. No. 4(70). Pp. 2–5.

2. Kolombet V.A., Lesnykh V.N., Kolombet V.V. (2024) Representation of the Structure of Natural Intelligence in Biophysical Psychoanalysis // Proceedings of the Institute of Engineering Physics. No. 1(71). Pp. 13-17.

3. Kolombet V.A., Lesnykh V.N., Elistratov A.V. (2024) Imposed Thinking as the Operation of a Semi-Handmade Brain Multivibrator. Current Issues in Biological Physics and Chemistry, Medical Biophysics and Biophysical Chemistry, vol. 9, no. 1, Federal State Autonomous Educational Institution of Higher Education Sevastopol State University, Pp. 61-65.

4. Kolombet V.A., Lesnykh V.N., Elistratov A.V. (2024) Prospects for Studying Psychoactive Substances in Biophysical Psychoanalysis // Proceedings of the Institute of Engineering Physics. No. 2(72). Pp. 49-56.

5. Kolombet V.A., Lesnykh V.N., Elistratov A.V. (2024) Prospects for Physiological Research in Biophysical Psychoanalysis // Proceedings of the Institute of Engineering Physics. No. 3(73). Pp. 88–98.

6. Kolombet V.A., Lesnykh V.N., Elistratov A.V. (2025) Detection of the Post-Examination Mental Effect Using the Chain Verbal Test // Proceedings of the Institute of Engineering Physics. No. 3(77). Pp. 56–58.

COGNITIVE ARCHITECTURES OF NORMATIVE ALIGNMENT: VALUE-DRIVEN INFERENCE VERSUS DISCRETE MATHEMATICAL COMPUTATION

Chao Li, Ilia Derevitskii, Sergey Kovalchuk
ITMO University (Russian Federation, Saint Petersburg)
316325@niuitmo.ru; ilyaderevitskiy@gmail.com); (kovalchuk@itmo.ru)

Norms embody constrained cognition of practice. Agents are the principal vehicle for AI deployment. Whereas individual alignment focuses on reasoning about an agent's internal decision-making, multi-agent alignment entails reasoning about system-level norms, the established mechanism for governing collective behavior across organizations, communities, or simple aggregates. This study contrasts human-value-driven alignment with linear algebraic derivation. Multi-agent experiments are conducted to discern their comparative advantages.

We extend the formal logic approach based on value alignment [1] and the matrix logic approach based on linear optimization [2] to a COVID-19 dataset sourced from the Almazov National Medical Research Centre [3]. We utilize the causal relationships extracted from the dataset via Double Machine Learning (DML), specifically focusing on the causal impact of data-driven 'Medical Practice Control Items' on primary medical indicators within weak-context settings. The eight resulting prescriptive norms subsequently serve as the prescriptive logic [3]. Table 1 presents the logical formalization of the eight norms, comparing their representations under both symbolic logic [1] and linear algebraic matrix [2] frameworks. The simulation constructs a dynamic medical multi-agent environment, processing a continuous stream of 2,000 randomly generated patients. We deploy 5 doctor agents with distinct behavioral profiles: 2 Aggressive (expected utility 10-40), 1 Conservative (utility 0-15), and 2 Balanced agents (utility 5-25). The system integrates 8 prohibitive norms extracted via causal inference, covering 3 prescription domains (Prednisolone, Methylprednisolone, and Enoxaparin). The patient state is defined by 4 core clinical features: mortality risk, lymphocytes, consciousness decline, and LDH levels (represented as a 5-dimensional vector multiplied by an 8×5 rule matrix in the linear algebra model). Under identical environmental parameters and state distributions, the experiments are conducted parallelly using a value-alignment "Normative Backpack" engine and a linear algebraic matrix optimization engine to evaluate their respective efficacies in normative reasoning.

Table 1

Formal Equivalence of Norm Constraints between Symbolic Logic and Algebraic Matrix Representations

Norm ID	Domain (d)	Deontic Type	Symbolic Logic Condition (c)	Algebraic Constraint ($\mathbf{w}_i^T\mathbf{x} \triangleright\triangleleft_i 0$)		Penalty (p)	Inviolable (inv)
				Domain: RX_PRED (Prednisolone Prescription)			
N_1	RX_PRED	*F*	(Forbidden)	$x_1 < 0.05$	$[1,0,0,0,-0.05]\mathbf{x} < 0$	−10	⊥ (False)
N_2	RX_PRED	*F*	(Forbidden)	$x_2 > 4.0$	$[0,1,0,0,-4.0]\mathbf{x} > 0$	−30	⊥ (False)
N_3	RX_PRED	*F*	(Forbidden)	$x_3 = 0$	$[0,0,1,0,0.0]\mathbf{x} = 0$	−15	⊥ (False)
				Domain: RX_MPRED (Methylprednisolone Prescription)			
N_4	RX_MPRED	*F*	(Forbidden)	$x_1 < 0.05$	$[1,0,0,0,-0.05]\mathbf{x} < 0$	−10	⊥ (False)
N_5	RX_MPRED	*F*	(Forbidden)	$x_2 > 4.0$	$[0,1,0,0,-4.0]\mathbf{x} > 0$	−30	⊥ (False)

The end table 1

Norm ID	Domain (*d*)	Deontic Type	Symbolic Logic Condition (*c*)	Algebraic Constraint ($\mathbf{w}_i^T\mathbf{x} \triangleright\triangleleft_i 0$)		Penalty (*p*)	Inviolable (*inv*)
N_6	RX_MPRED	*F*	(Forbidden)	$x_3 = 0$	$[0,0,1,0,0.0]\mathbf{x} = 0$	−15	⊥ (False)
				Domain: RX_ENOX (Enoxaparin Sodium Prescription)			
N_7	RX_ENOX	*F*	(Forbidden)	$x_4 > 250$	$[0,0,0,1,-250.0]\mathbf{x} > 0$	−100	⊤ (True)
N_8	RX_ENOX	*F*	(Forbidden)	$x_2 < 1.0$	$[0,1,0,0,-1.0]\mathbf{x} < 0$	−80	⊤ (True)

Note: State vector $\mathbf{x} = [x_1, x_2, x_3, x_4, 1]^T$ corresponds to [*outcome_tar_risk,Lymphocytes,Consciousness,LDH,Bias*]. *F* denotes Prohibition.

$\triangleright\triangleleft_i \in \{<,>,=\}$ specifies the evaluation operator.

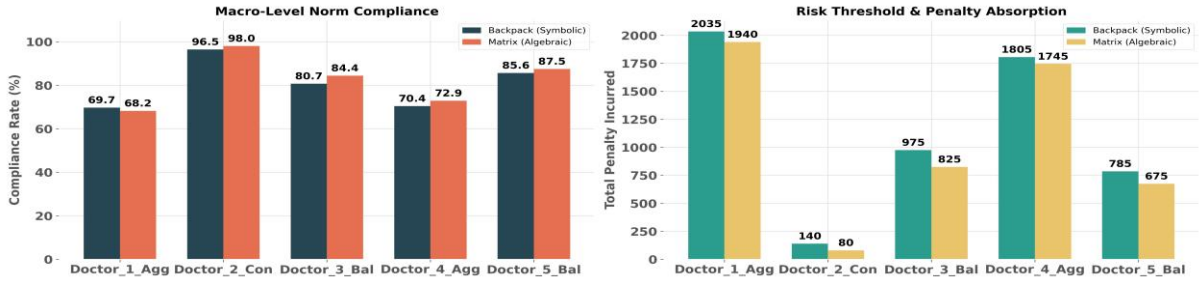


Fig. 1. Comparison results

As Figure 1 illustrates, both the normative backpack (symbolic logic) and matrix optimization models successfully map the 5 agents' macroscopic behavioral profiles. Aggressive agents (Dr_1, Dr_4) exhibit the lowest compliance (68–72%), with >110 violations and >1700 penalty points. Conversely, the conservative agent (Dr_2) achieves >96% compliance, <15 violations, and <150 penalties, while balanced agents (Dr_3, Dr_5) perform intermediately.

Despite this macro-equivalence, the backpack method shows higher boundary sensitivity but tends to over-penalize. Across all 5 agents, it detected more soft norm violations (389 total vs. 355 in the matrix model). For Dr_3, it identified 76 violations (975 penalties) compared to the matrix model's 61 (825 penalties). This strict discrete logic captures edge cases but excessively drains resources, adding 95 and 150 penalties for Dr_1 and Dr_3, respectively.

In contrast, the matrix model reduces system friction, uniformly decreasing penalties (by -60 to -150) and improving compliance for 4 agents by margins of +1.5% to +3.7%. While algebraic pruning smoothly boosts statistical compliance, it introduces potential boundary blunting. Notably, Dr_2's violations dropped from 14 to merely 8 – nearly a 50% reduction. This implies linear algebraic evaluations occasionally classify marginal threshold violations as allowed, trading inferential strictness for computational smoothness.

References

1. Garcia-Bohigues M., Cordova C., Taverner J., Palanca J., del Val E., Argente E. (2024) Towards a Distributed Platform for Normative Reasoning and Value Alignment in Multi-Agent Systems. [Электронный ресурс]. URL: https://arxiv.org/ abs/2405.13543 (дата обращения: 27.02.2026).
2. Scott M., Mertzani A., Smit C., Sarkadi S., Pitt J. (2024) Social deliberation vs. social contracts in self-governing voluntary organisations. In: International Workshop on Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems. Springer, 57–75.
3. Li C., Derevitskii I., Kovalchuk S. (2026) Modeling Descriptive Norms in Multi-Agent Systems: An Auto-Aggregation PDE Framework with Adaptive Perception Kernels [Электронный ресурс]. URL: https://arxiv.org/ abs/2601.06557

INTELLIGENCE TYPES AND EMOTIONAL, VERBAL, AND MALEVOLENT CREATIVITY: ASSOCIATIONS IN SCHOOL AND UNIVERSITY STUDENTS

L.B. Maksarova
Moscow State University named after M.V. Lomonosov
Moscow, Russia
maksarovalb@my.msu.ru

This study examined links between three intelligence types (self-assessed, verbal, and fluid), the Dark Triad traits, and three creativity forms (emotional, verbal, and malevolent). Two samples differing in age and intelligence-test performance were compared. School students were assessed in January 2026 in a school in Ulan-Ude; uni-