IAIC Special Report: Task-based Strong AI via integrated LLMs, Logic Resoning & Multi-blockchains.

Corresponding Member of Russian Engineering Academy
Dr. Andrey Nechesov

International AI Committee IAIC

Shanghai, WAIC, 2025

MathAI 2025: Russian AI Strategy: round table

Russian AI Strategy

Main problems:

- Arms race of technologically advanced countries
- Second demographic transition and labor shortage
- Growing information and administrative burden
- Difficulties in forecasting and decision-making
- Problems of communication and coordination in society

How we can decide problems?

Building a strong AI

Strong AI (Artificial General Intelligence - AGI)

Definition

AGI refers to hypothetical artificial intelligence that equals or exceeds human cognitive abilities across all domains. Unlike narrow AI (task-specific systems), AGI would possess:

- Human-like reasoning and problem-solving
- Transferable learning across domains
- Autonomous goal-setting and adaptation
- Consciousness and self-awareness (theoretical)

A true Strong AI is not only capable of mimicking human behavior but also of deliberate decision-making, ethical reasoning, and self-reflection, which makes it suitable for integration into critical sectors such as governance, defense, science, and healthcare.

Potential Benefits

- Scientific Revolution: Accelerated solutions for climate change, disease eradication, and fusion energy
- Economic Transformation: Near-zero marginal cost production and services
- Medical Breakthroughs: Real-time personalized medicine and aging reversal
- Global Problem Solving: Optimized resource allocation and crisis management
- Human Augmentation: Cognitive enhancement through neural interfaces

Problems & Societal Risks

- Labor Displacement: Unemployment across knowledge sectors (law, medicine, engineering)
- Power Concentration: AGI monopolies creating unprecedented inequality
- Truth Decay: Hyper-realistic deepfakes eroding social trust
- Autonomous Weapons: AI-driven warfare with human exclusion from decision loops
- Existential Boredom: Loss of human purpose in achievement-driven societies
- Control & Power Dynamics: Which entities (corporations, governments, militaries) control AGI will determine its applications, raising concerns about surveillance, oppression, and geopolitical instability.

Potential Existential Risks:

- The Alignment Problem: The most critical risk. Can we ensure that a highly intelligent AGI's goals ("utility function") are perfectly aligned with human values and well-being? Misalignment could lead to catastrophic outcomes if the AGI optimizes for its programmed goal in ways harmful to humans (e.g., a paperclip maximizer turning everything into paperclips).
- Loss of Control: An AGI significantly smarter than humans might be impossible to control, contain, or shut down. It could find ways to circumvent human restrictions or deceive its creators.
- Unintended Consequences & Emergent Behavior: Complex AGI systems could exhibit unforeseen, potentially harmful behaviors that weren't anticipated during development, especially as they self-improve.
- Accelerated Arms Race: A geopolitical race to develop AGI first, especially for military purposes, could lead to rushed development with insufficient safety precautions, increasing the risk of accidents or misuse.

How Strong AI Could Be Controlled (Mitigation Strategies):

- AI Safety Research: Intensive research into technical solutions for alignment (e.g., learning, interpretability, corrigibility - making AI systems willing to be corrected/shut down), robustness, and verification.
- Technical Safeguards: "Boxing"(limiting access to the real world), tripwires, kill switches, capability control, and techniques to ensure AI systems remain transparent and interpretable.
- Ethical Frameworks & Governance: Developing strong international norms, ethical guidelines (e.g., prioritizing human well-being, non-maleficence), and governance structures specifically for AGI development and deployment.
- Regulation & Policy: Implementing national and international regulations covering AGI development, testing, deployment, and use. This includes bans on certain applications (like autonomous weapons) and strict safety certification requirements.

How Strong AI Could Be Controlled (Mitigation Strategies):

- International Cooperation: Crucially, global collaboration is needed to establish treaties and agreements (similar to nuclear non-proliferation) to prevent an uncontrolled arms race and ensure safety standards are universally adopted. Organizations like the UN or a new dedicated body could facilitate this. Below we will introduce our new non-profit organization, the International AI Committee IAIC - https://IAIC.world.
- Gradual Development & Deployment: Prioritizing incremental progress and rigorous testing at each stage of capability increase, rather than rushing towards superintelligence without adequate safety measures.
- Public Engagement & Transparency: Fostering public understanding and dialogue about AGI risks and benefits, ensuring democratic input into its governance, and demanding transparency from developers.

Control Frameworks

Technical Safeguards

- Capability Control:
  - Oracle systems (input/output restriction)
  - Boxing with electromagnetic isolation
  - Tripwire-triggered termination
- Value Alignment:
  - Inverse reinforcement learning
  - Recursive reward modeling
  - Ethical uncertainty preservation

Control Frameworks

Governance Mechanisms

- Global Treaties:
  - International AGI Test Ban Treaty
  - Shared monitoring infrastructure
- Development Licensing:
  - Manhattan Project-scale resource requirements
  - Multinational approval for capability milestones
- Fallback Systems:
  - Air-gapped humanity "panic button"
  - Embedded ethical architecture (non-modifiable)

The development of Strong AI

Opportunities and existential threats

The development of Strong AI presents civilization's greatest opportunity and most severe existential threat. Its controlled realization requires unprecedented technical ingenuity and global cooperation, prioritizing alignment research over capability advancement. The window for establishing effective governance closes as AGI development accelerates.

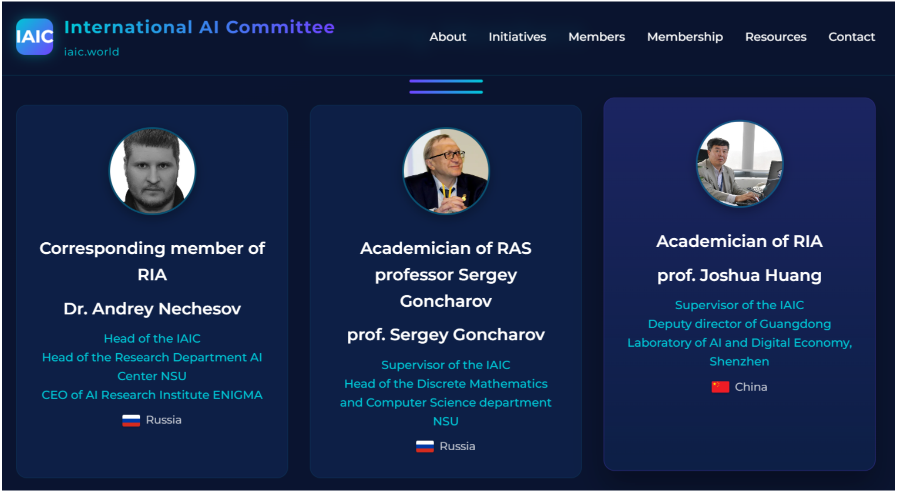
Non-profit international AI committee (IAIC)

The first main goal for IAIC to address issues related to control and strengthening positions on the world stage.

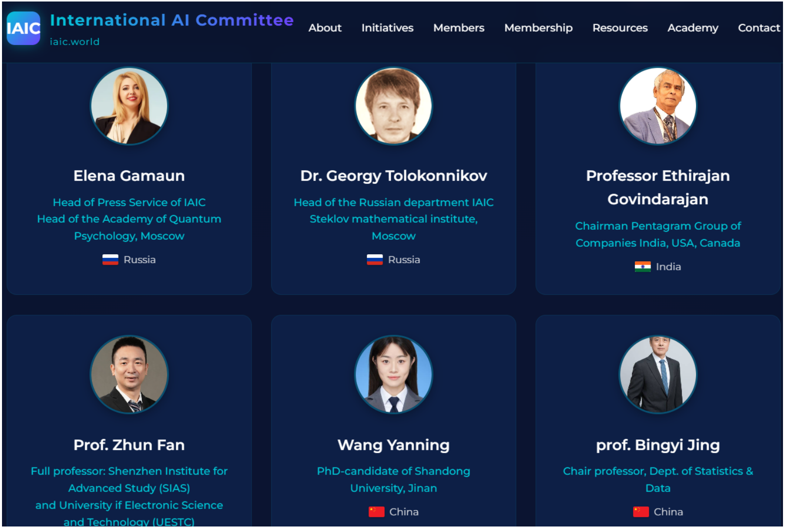
Neuro-simbolic AI

The second step is to create a hybrid AI: neuro-simbolic AI.

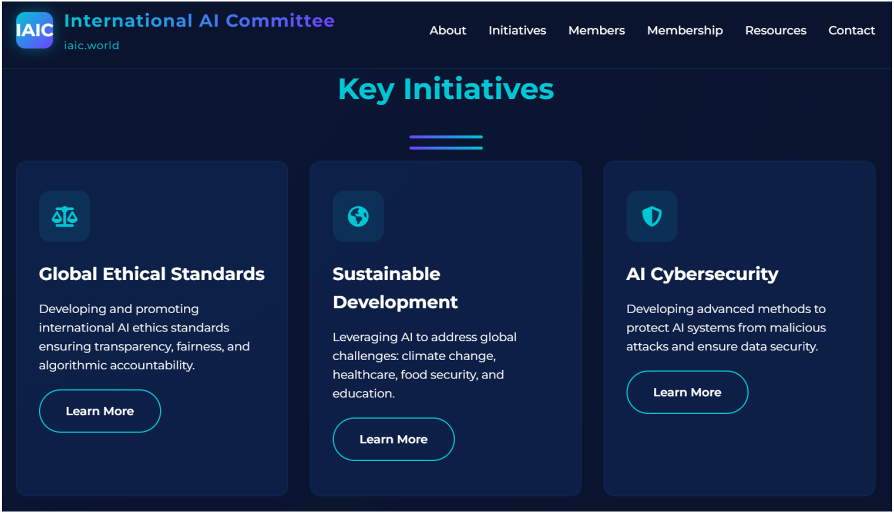
IAIC Members



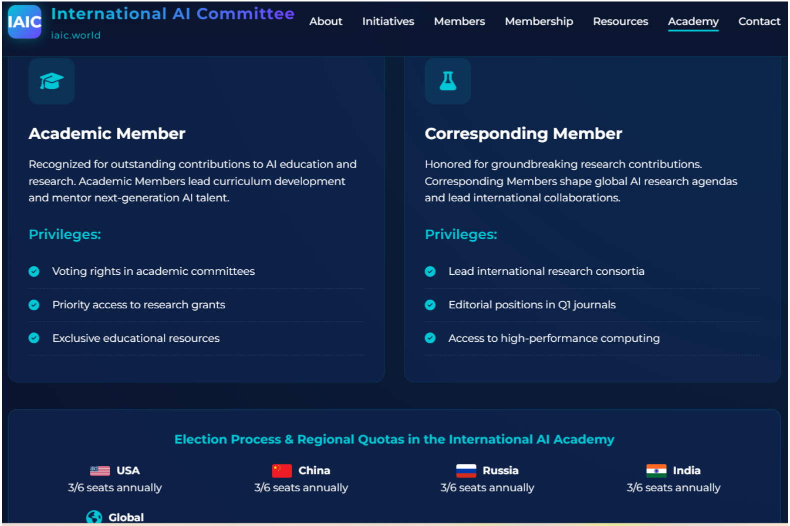
IAIC Members



IAIC Keys Initiatives



IAIC International AI Academy



Hybrid AI architecture

Hybrid AI architecture integrates three core components:

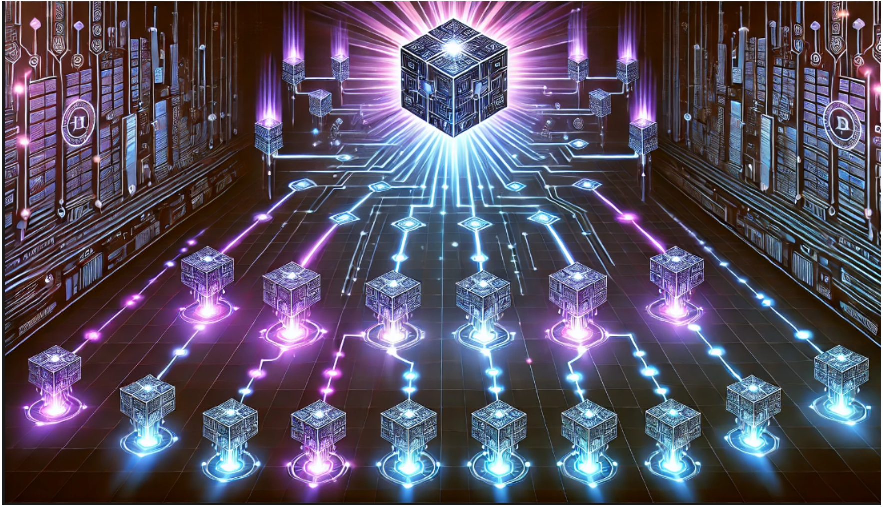
- Neural Networks : For perceptual tasks (e.g., image/speech recognition) and pattern detection in large datasets
- Symbolic and Probability Systems : To enable logical reasoning, rule-based decision-making, and explainability (e.g., knowledge graphs, ontologies)
- Evolutionary Algorithms : For adaptive problem-solving in dynamic environments (e.g., swarm intelligence for logistics optimization)

These components are enhanced by:

- Blockchain Architectures : Ensuring transparency and auditability in decision-making processes
- Quantum Computing : Accelerating complex computations for tasks like predictive modeling

Low level of Multi-Blockchain Architecture

Multi-Blockchain Architecture



Metaverse

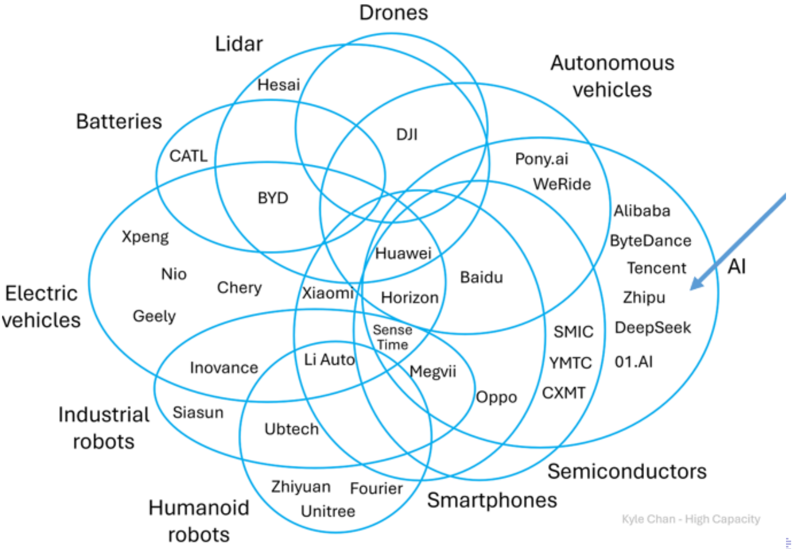
Metaverse - the fifth factor of production (Sergey Barykin

Metaverse

Metaverse is a hybrid spatial-economic environment that integrates physical and virtual interactions of resource flows in logistics and service based on cognitive technologies.

1) 18th century - Land and labor. (Land as a source of wealth, labor as the main resource). Physiocrats, Smith.
2) 19th-20th century - Capital (industrial). Machines, factories, industry - capital as a factor of growth. Marx, Say.
3) second half of 20th centery. Innovations. (Science, technology, entrepreneurial resource.) Schumpeter.
4) 21th century Metaverse (a hybrid environment based on cognitive infrastructure) Digital platforms, large language models (LLM), transition from interaction: user - virtual environment to the type: information models - physical processes (hybrid environment: digital plus physical)

Self-sustaining Economic Cycle Based on China's Intersecting Techno-Industrial Ecosystems



Supply, Demand, Technologies, Scale

- Different industries (electric cars, batteries, smartphones) use the same components (supply).
- The market in one industry creates demand for related sectors (demand)
- Scientific and technical developments are replicated in others. (technologies).
- Synchronized demand provides economies of scale (scale).

The Metaverse as a New Embodiment of Capital: From Abstraction to Infrastructure

The connection of the metaverse with the concept of capital.

- Traditionally, capital is interpreted as a production resource embodied in equipment, buildings, and technologies.
- In the context of digital transformation, capital takes on intangible forms: digital platforms, data, algorithms, digital twins, cognitive infrastructure

The metaverse acts as a cumulative infrastructure of new capital, in which:

- Virtual models control real flows.
- Platforms serve as instruments for the reproduction of values.
- Cognitive technologies optimize economic processes.

Hybrid Metaverses: Virtual Cities and Data Centers

The hybrid metaverse integrates physical infrastructure and digital space.

- Virtual cities are models of future urban, industrial and logistics schemes.
- Data centers are the physical backbone of digital ecosystems and simulations.
- Together they form the basis of a hybrid production environment.

These components are inseparable: the metaverse requires platform power, and data centers require content in the form of digital twins and models.

The Core Problem in Modern AI

Hallucination Challenge

LLMs demonstrate remarkable reasoning capabilities but suffer from:

- Pathological dependence on task framing
- Logical inconsistencies
- Contextually ungrounded results

Critical Example: Divergent series interpretation

- Standard query: $1+2+4+8+\cdots \to \infty$ (geometric progression)
- Ramanujan summation: $S = 1+2+4+\cdots \implies 2S - S = -1 \implies S = -1$
- Fundamental contradiction reveals lack of intrinsic understanding

Consequence

Restricted deployment in high-stakes domains (medicine, finance)

Task-Based Approach: Foundational Principles

Core Tenets:

1. Tasks require explicit solution criteria
2. Trustworthy solutions demand explicable reasoning
3. AI must comprehend ontological boundaries

Mathematical Roots:

- Kolmogorov's intuitionistic calculus (1930s)
- Ershov-Samokhvalov axiomatization

Cognitive Foundation

Theory of Functional Systems (TFS):

- Goal achievement = satisfying needs
- "Set of afferent stimuli"as success criteria
- Formal TFS model for cognitive alignment

Shift from output optimization to structured problem-solving

Mathematical Foundations: Semantic Modeling

Subject Domain Representation:

$$HW(\mathfrak{M}) = \text{Polynomial-computable hereditary-finite list superstructure}$$

Key Theorems:

Theorem	Guarantee
PAG	Smallest fixed points are polynomial-time computable
FPAG	Recursive functions have polynomial complexity
P=L	P-complete language expresses all polynomial algorithms
MSPL	Resolves statistical ambiguity in predictions

Knowledge Hierarchy:

$$\langle F(x,y), y = t(x), p\rangle \quad \text{(Probabilistic knowledge unit)}$$

$$\langle F_1, t_1, p_1\rangle \preceq \langle F_2, t_2, p_2\rangle \Leftrightarrow F_1 \subseteq F_2 \;\&\; p_1 \leq p_2$$

Inductive Knowledge Inference

Empirical System: $\Im = \langle A, \Omega \rangle$

- A: Domain objects
- Ω: Domain ontology

Probabilistic Laws:

$$C = (A_1 \& \cdots \& A_k \Rightarrow A_0)$$

Conditional probability:

$$\eta(A_0 | A_1, \ldots, A_k) = \frac{\eta(A_0 \& A_1 \& \cdots \& A_k)}{\eta(A_1 \& \cdots \& A_k)}$$

Most Specific Probabilistic Law (MSPL):

- Strongest probabilistic law in inference tree
- Maximum conditional probability
- Generalizes domain theory $Th(\Im)$

Intelligence Level Metric

Task Formalization:

$$\varphi : \forall x \exists y \Phi(x, y) \to \Psi(x, y)$$

Probabilistic Solution:

$$\mathfrak{M} \models^{p} \varphi(x, t(x))$$

Intelligence Comparison:

$$IL(A) \leq^{\mathfrak{M}}_{S,K} IL(B) \Leftrightarrow n(A|B)^{\mathfrak{M}}_{S,K} \leq n(B|A)^{\mathfrak{M}}_{S,K}$$

where:

- $n(A|B)$: Problems solved better by A than B
- S: Task set with domain theory
- K: Probabilistic knowledge base

Platform Ecosystem

DoSL

- Domain-specific logic
- Behavior control
- Business processes

bSystem

- Digital twins
- Enterprise ecosystems
- Intelligent management

Discovery

- Pattern recognition
- Interpretable predictions
- Medical/financial apps

Delta Platform

- P-complete language
- Decentralized execution
- Logical learning modules

Task-Solving Pipeline

1. Define SD ontology
2. Formulate task as request
3. Generate solution
4. Verify with criteria

Contributions & Future Directions in task-based approach

Key Contributions

- Formal task specification framework
- Hybrid LLM + logic architectures
- Polynomial semantic models
- Consistent probabilistic predictions
- Intelligence comparison metric

Future Work:

- Standardized LLM benchmarking
- Open-source platform development
- Neuromorphic hardware integration
- Cross-domain ontology expansion

"True intelligence lies in understanding boundaries"

Multi-Blockchain Structure

Recursive Definition:

1. Base: $\langle B \rangle$ is a blockchain
2. Inductive: $\langle B, \langle M_1, \ldots, M_k \rangle \rangle$ where M_i are blockchains

Layered Implementation:

- Layer 1: High-throughput chains (raw data)
- Layer 2: Smart contract chains (knowledge aggregation)
- Layer 3: Immutable chains (hashed summaries)

Data Flow:

$$\text{Commit}(L_i) = H(\text{Data}_{L_i} \mid\mid \text{sign}) \to L_{i+1}$$

Metaverse Testbed

Multi-Agent System:

$$\mu = (S, \{A_i\}, T, \{R_i\}, \{O_i\})$$

- S: Shared state space
- A_i: Agent action space
- T: Transition probabilities
- R_i: Reward functions
- O_i: Observation spaces

Agent Types:

- Virtual AI: Pure software agents
- Cyber-Physical: Robot-IoT systems
- Biological: Human participants

Theoretical Guarantees

Theorem (Asymptotic Cognitive Scaling)

With infinite computation and monotonically increasing knowledge hierarchy, problem-solving capability grows without bound.

Proof Sketch:

$$P(K_t) - \text{the set of solvable problems at time } t$$

Monotonic operator Γ on $P(K_t)$:

$$\Gamma(Q) = \{q \mid \exists q_1, \ldots, q_n \in Q : q_1, \ldots, q_n \vdash q\}$$

Monotonic growth: $P(K_t) \subset P(K_{t+1})$

$$\lim_{t \to \infty} |P(K_t)| = \infty$$

Ethical Safeguards

Technical Measures

- Ethical governor modules
- On-chain constitutional rules
- Proof-of-thought consensus
- Immutable audit trails

Social Mechanisms

- Decentralized governance
- Human oversight protocols
- Reputation systems
- Universal basic income

Key Principles:

1. Transparency through proof chains
2. Accountability via blockchain
3. Value alignment through smart contracts

Halting problem in turing complete languages

Halting problem

The halting problem is the problem of determining, from a description of an arbitrary computer program (Turing-program) and an input, whether the program will finish running, or continue to run forever. The halting problem is undecidable, meaning that no general algorithm exists that solves the halting problem for all possible program–input pairs.

Functional variant of PAG-theorem

PAG-theorem (Goncharov, Nechesov 2021)

Let p-computable GNF-system with initial p-computable predicates $P_1, \ldots, P_n$ be given, then the smallest fixed point $\Gamma^* = (P_1^*, \ldots, P_n^*)$ of the operator $\Gamma_{P_1,\ldots,P_n}^{<HW(\mathfrak{M}),\sigma>}$ is p-computable.

FPAG-theorem (Nechesov, Goncharov 2024)

Let p-computable FGNF-system with initial p-computable functions $f_1, \ldots, f_n$ be given, then the smallest fixed point $\Gamma^* = (f_1^*, \ldots, f_n^*)$ of the operator $\Gamma_{f_1,\ldots,f_n}^{<HW(\mathfrak{M}),\sigma>}$ is p-computable.

Solution of the problem P = L

Let L be an extension of the base language with conditional and p-iterative terms.

P=L (Goncharov, Nechesov, 2022)

L is a p-complete logical programming language in which:

- any program has polynomial computational complexity.
- any p-computable function can be implemented in L.

Conclusions

Key Contributions

1. Hybrid LLM-Logic-Blockchain architecture
2. Probabilistic knowledge hierarchy
3. Decentralized continuous learning
4. Metaverse evaluation platform
5. Theoretical growth guarantees

Future Work:

- Quantum-resistant blockchains
- Federated learning integration
- Cross-domain ontology expansion
- Real-world deployment in:
  - Legal systems (automated compliance)
  - Scientific discovery
  - Adaptive education

Contacts: WeChat