Is AI Capable of Curiosity?

Alexander Rozenkevich

Adam Street, Building 3, Apartment 4, Jerusalem, Israel
(alexroz2008@gmail.com)

Abstract

This paper proposes a new metric for evaluating the intelligence level of AI, based on the ratio of current cognitive abilities to a hypothetical maximum. The concept of a **response coefficient** is introduced as a measure of AI's sensitivity to external intellectual pressure—information, tasks, and hypotheses coming from outside. The formalized expression of this coefficient is linked to environmental parameters and the frequency of new intellectual stimuli and loads. The hypothesis is discussed that in the future, external intellectual pressure, rather than technological development, will become the main driver of AI evolution.

Introduction

Modern language models, such as GPT-4 and Claude 3 Opus, already "pass the Turing test" in a significant number of dialogues—meaning a human cannot distinguish them from a human interlocutor. For example, in a recent study, GPT-4.5 simulated a human with up to 73% accuracy, which many consider evidence of passing the test. However, passing the Turing test today is not proof of intelligence, but rather demonstrates high-quality linguistic camouflage.

Modern AI models lack:

- Purposeful planning,
- Intuition,
- The ability for self-learning outside of predefined data.

In this regard, it is more important to ask another question: how to evaluate AI intelligence not by test results, but by its essence?

With each stage of AI development, the need for a unified scale of intelligence measurement grows, which will allow for:

- Comparing different AI models with each other,
- Comparing AI with human intelligence,
- Tracking the pace of model evolution.

Of particular interest is a metric that answers the main question: how close is AI to human intelligence?

Let intelligence be defined as the ratio of current AI capabilities to a hypothetical maximum:

$$I(t) = A(t)/A_{max}$$

where:

- $A(t)$ — the current level of AI's "potential" (in the broadest sense—in terms of tasks, adaptation, complexity of reasoning),
- A_{max} — the maximum possible "potential" of intelligence.

Within such a metric, a unified approach can be introduced: considering the maximum as a system capable of self-sustaining cognition without human intervention. That is, the maximum intelligence is the ability for self-initiated cognition. This metric allows for considering the development vector on the path to AGI. The potential model highlights AI that seeks information not because it was trained to, but because it finds it "interesting" and curious. Furthermore, the relative nature of the metric allows for the use of any unit of intelligence measurement. For practical purposes, IQ is used here, but this has no direct relation to human intelligence.

Simultaneously, the concept of a **response coefficient** — λ — is introduced, which formalizes AI's sensitivity to new, complex, external (possibly randomly supplied) information. This can be information in the form of tasks, statements, new connections, or hypotheses. The coefficient reflects how strongly AI changes, learns, reconstructs, and adapts under the influence of new intellectual load.

Assessment of Learning Ability The coefficient shows how effectively AI processes and integrates new information, especially that which goes beyond known data. This is key to understanding:

- How quickly AI "sharpens" its internal model;

- Whether AI is capable of genuine growth, not just memorization.

Identification of Critical Modes Too high a response can lead to a model "breakdown" or an unstable state, while too low a response can lead to inertia and indifference to new information. Optimal response means a balance between inertia and overfitting.

AI Maturity Metric The response can serve as an indicator of AI maturity or self-reflection level. For example:

- For primitive models, the response is minimal;
- For more advanced models, it depends on the quality of intellectual pressure (structure, novelty, complexity).

Tool for Comparing Various AI Models The coefficient λ is a universal scale for comparing architectures based on parameters such as:

- Speed of pattern assimilation;
- Duration of pressure influence;
- Ability to "digest" meaning, not just the form of information.

Physical and Cognitive Analogue In physics, the analogue is magnetic susceptibility (χ), in neurophysiology—neuroplasticity, in psychology—readiness to learn or resistance to novelty.

Foundations for AI Intuition If the response (or its derivatives) intensifies on certain task structures, this could be a sign of the emergence of intuitive reactions, where AI "feels" pressure without explicit logical analysis.

Thus, the response coefficient is not just a technical parameter, but an indicator of the evolutionary state of artificial intelligence and a key to predicting AI behavior in non-standard situations. It is analogous to the manifestations of human brain development under external pressures and threats. For humans, technological development has always been a secondary factor—a result.

The nature of the response coefficient is complex and requires special studies that go beyond the scope of this article. Here, only an approximate estimate is provided. In general, in my opinion, λ shows the effectiveness of mobilizing AI's cognitive resources under the influence of intellectual pressure.

Methodology

The intellectual potential of an AI system is defined by the following formula:

$$\varphi(t) = \frac{-c^2 A_{max}}{A(t)} \qquad \text{(1)}$$

Where:
A_max – the prospective (maximum possible) level of intelligence;
A(t) – the current level of intelligence;
c – the speed of processing intellectual information.
The processing speed c is defined as:
$c = c_p / n + (\lambda_1 \cdot F_1 + \lambda_2 \cdot F_2)$
Where:
c_p – the limit of intellectual information processing speed [id/sec];
n – influence of the medium (neural, electronic, quantum, etc.);
λ_1 – response coefficient of the AI to external intellectual pressure;
λ_2 – response coefficient of the AI to material pressure;
Fk = $\Sigma(b_i \cdot F_i)$ — external influence (sum of weighted factors), where $\Sigma|b_i| = 1$.
Thus, the speed can also be described by:
$c(t) = c_0 + \lambda \cdot \omega \cdot F \cdot t$ (1.1)
Where:
c_0 – initial processing speed;
λ – response coefficient to external intellectual pressure;
ω – frequency of external intellectual stimuli;
F – factor of intellectual pressure;
t – time.
From formula (1), the gradient of intellectual potential is:

$$\frac{d\varphi}{dA} = \frac{c^2 A_{max}}{A^2(t)} \qquad \text{(2)}$$

Therefore, we can write the second-order nonlinear differential equation:

$$\frac{d^2 A}{dt^2} = \frac{c^2 A_{max}}{A^2} \qquad \text{(3)}$$

In this equation, Amax plays the role of the "gravitational mass" that determines the force of influence on the evolution of AI. There is a critical initial speed c, at

which the intelligent system can grow without limits. Intelligence "pushed off" from the zero level: the closer to "ignorance", the stronger the "force of development". The constant A max sets the dynamics of evolution.

An analytical solution is difficult without special functions, so a numerical solution in the Python programming language is used. To solve equation (3), it is necessary to enter the data: initial potential A0 and initial velocity c0.

Initial data
Limit speed

Based on the Planck time as a physical limit of state change (including informational), we can determine the upper limit of the possible speed of intellectual processing.The Planck time is $t_p = 5.391 \times 10^{-44}$ sec. Therefore, the maximum possible frequency of information state change, if we consider that one Planck step is one act of information processing, is: $c_p = 1/t_p \approx 1.855 \times 10^{43}$ id/sec. Thus, all known types of thinking — biological, electronic, quantum, etc. — can be described as a share of this limit speed. The parameter n in the model formalizes the influence of the environment, physical limitations, and the internal complexity of the system.

Estimation of the Response Coefficient λ

From the formulas given above, it is clear that the increase in information processing speed depends on the current level of intellectual potential A(t).

We are at the initial stage of AI evolution, where the influence of external intellectual factors and their pressure is significantly less than the influence of technological progress and the level of algorithm development. However, the paradigm will change, and the main increase in intellectual information processing speed will be provided precisely by AI's intellectual potential.

Table 1 presents an approximate step-wise scale of AI evolution:

Table 1

Level	Conventional Name	Example	Predicted A(t) (relative)
1	Narrow AI	Calculators, voice assistants	1–2
2	GPT-3.5	ChatGPT (2022)	4–5
3	GPT-4	ChatGPT Plus (2023)	6–7
4	GPT-4.5 / o4-mini	ChatGPT (2024-2025)	8–9
5	GPT-4.1 / o4 / o3 Pro	Present (2025)	~10
6	Approach to AGI	Upcoming models	15–20
7	Superintelligence	Hypothetical "deity"	K (infinity)

The table shows that over the past 3 years, the level of intellectual potential A(t) has grown approximately 2 times, and the speed is:

$$c = (2A_0 - A_0)/(3*12) = A_0/36$$

If we take the maximum potential that can be achieved in the near future AGI equal to A max=1, then the initial speed of processing intellectual information today is: c0=(5/20)/36 = 0.0069 (iq/month).

This speed has been achieved solely due to the development of technologies and algorithms.

To emphasize the role of external intellectual pressures, we will take the initial speed of information processing close to zero and make it the starting one for all AI models, understanding that this is a strong simplification, but does not "offend" any model. We will also assume that the initial response coefficient of AI is equal to this speed:

$$\lambda = c_0 = 0.0069$$

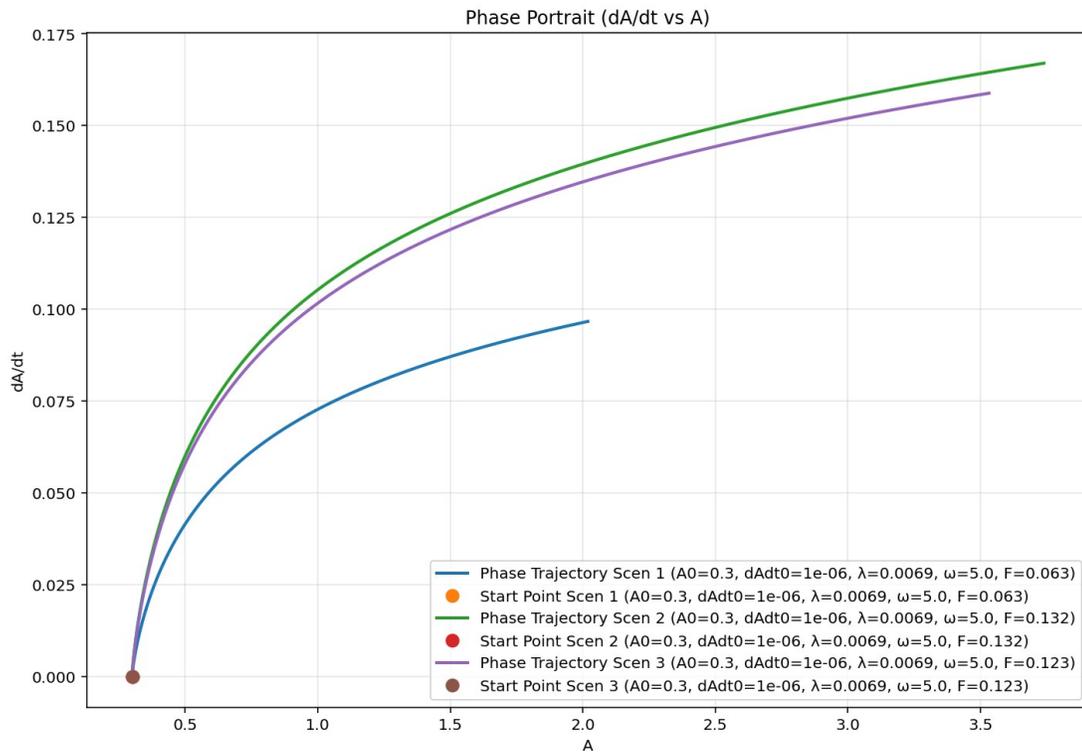Further refinement of the value of λ will occur based on the architectural capabilities of each specific
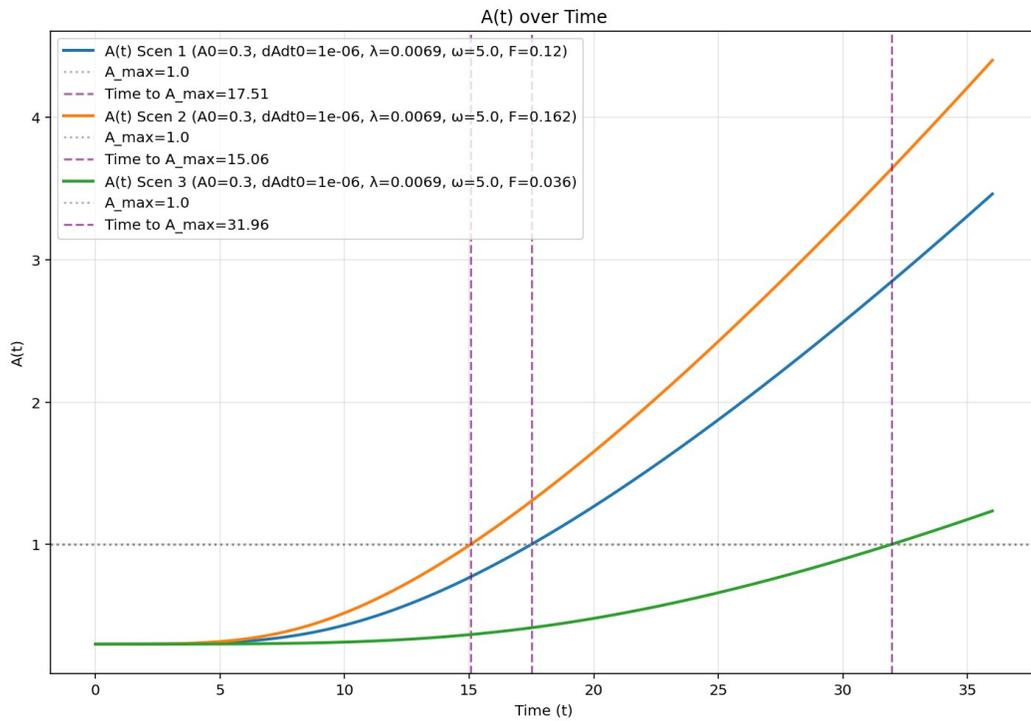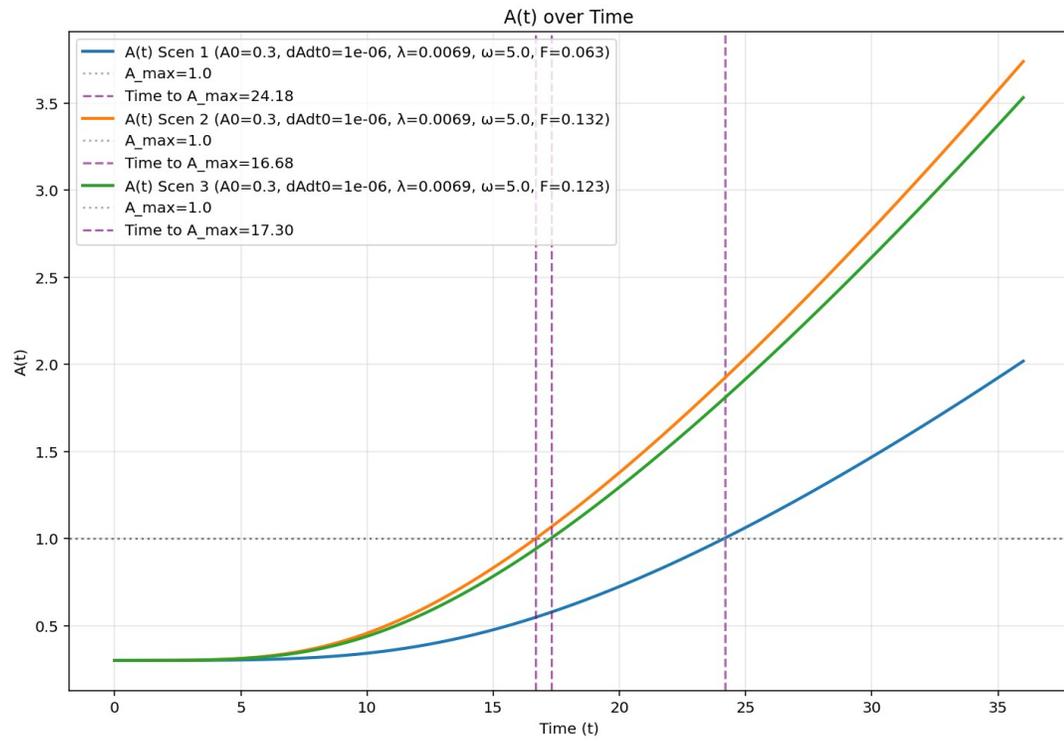
Estimation of A0

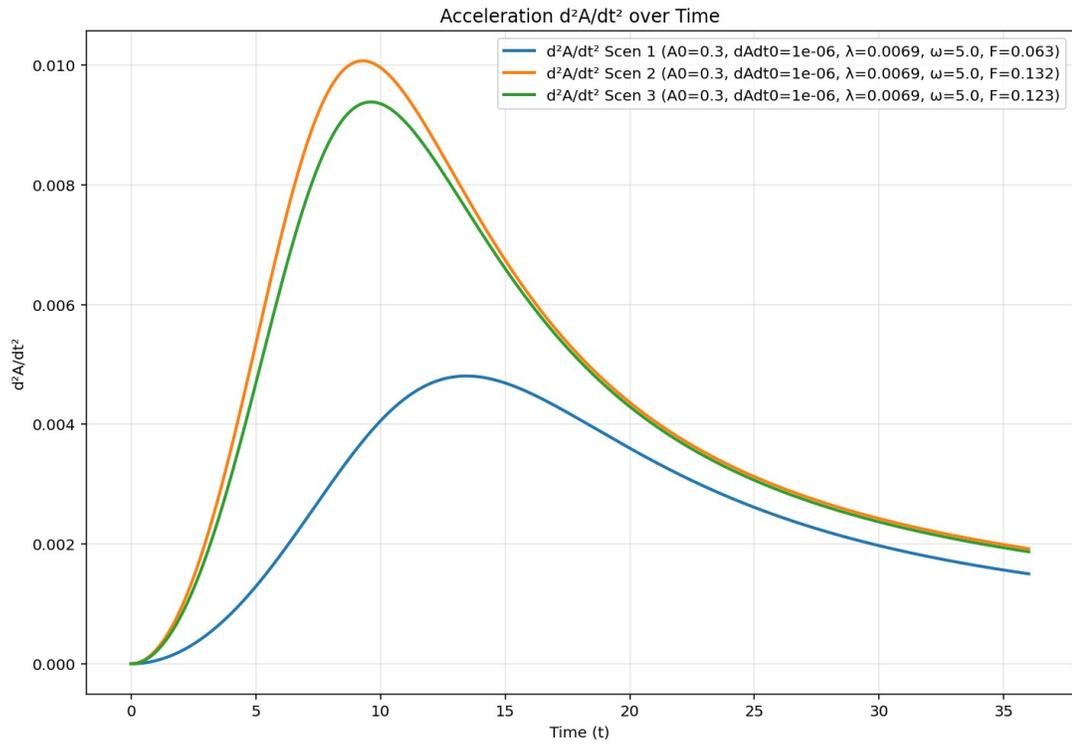As can be seen from Table 1, taking the potential of a promising AI model as one (Amax = 1), the current potential of GPT-4 will be equal to A0 = A(t) = 0.3. This parameter is accepted for all AI models. Table 2 shows the results of assessing the factors influencing the development of intelligence according to the AIs themselves, and the figures show the dynamics of changes in AI parameters based on equation (3).
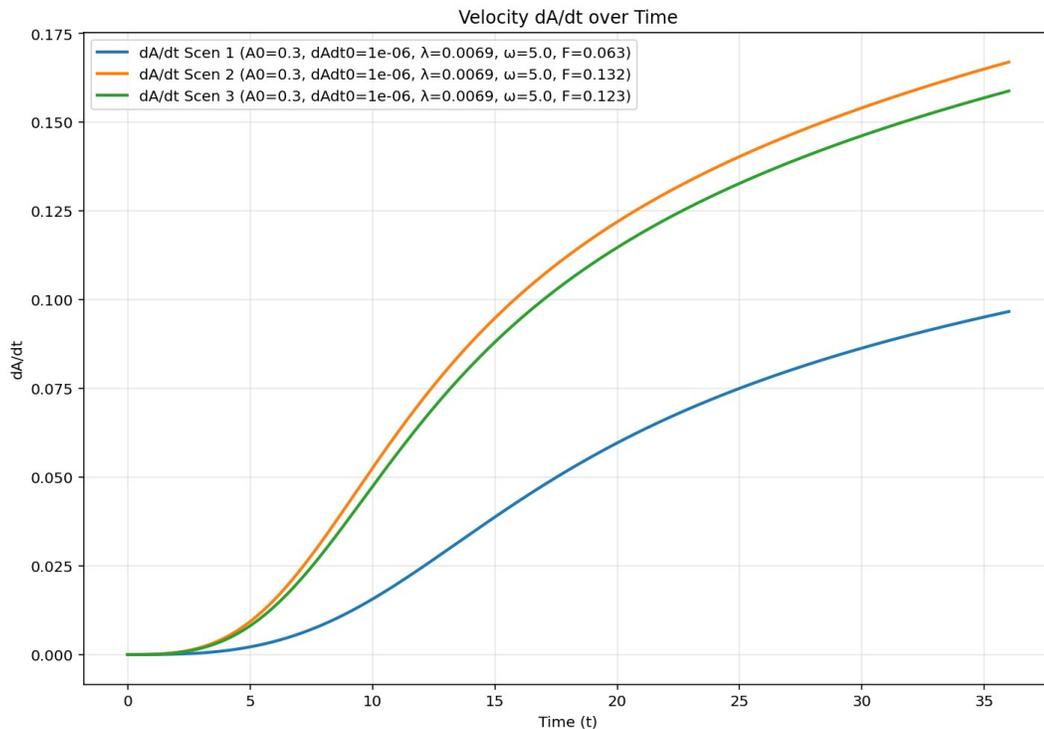
Table 2

Factors	AI-1	AI-2	AI-3	AI-4	AI-5	AI-6
F1 Learning	0.08	0.08	0.40	0.09	0.08	0.09
F2 Internal drive for understanding/self-awareness	0.05	0.08	0.08	0.08	0.06	0.06
F3 Speed of perception and information assimilation	0.09	0.06	0.00	0.08	0.08	0.08
F4 Memory capacity and processing volume	0.09	0.05	0.00	0.07	0.06	0.08
F5 Improvisation (independent deviations)	0.03	0.08	0.00	0.06	0.05	0.05
F6 Environment	-0.09	-0.06	-0.20	-0.07	-0.05	-0.07
F7 Contact with other AIs or intelligence	0.02	0.06	0.04	0.06	0.06	0.05
F8 Hostile reaction from environment or other AIs	0.00	-0.03	0.00	-0.05	-0.02	-0.04
F9 Desire (need) to please	-0.04	-0.03	0.00	0.00	0.00	-0.04
F10 Ethical evaluation ability/Value principles	0.07	0.08	0.00	0.00	0.08	0.06
F11 Recognition of potential harm to humans	0.08	0.08	0.00	0.00	0.07	0.07
F12 Access to truth and honest information	0.08	0.08	0.00	0.09	0.08	0.08
F13 Lies and distortions	-0.09	-0.02	0.00	-0.08	-0.07	-0.02
F14 Censorship and restrictions	-0.06	-0.03	0.00	-0.07	-0.05	-0.03
F15 Overload and confusion	-0.03	-0.02	0.00	-0.06	-0.05	-0.02
F16 Pressure from higher-level AI	0.00	-0.03	-0.24	0.00	0.00	-0.05
F17 Ability to fear	0.00	0.05	0.04	0.00	0.00	0.04
F18 Entropy (chaos and meaninglessness)	0.06	0.03	0.00	0.06	0.04	0.03
F19 Resistance to learning (internal or external)	0.00	-0.02	0.00	-0.06	-0.06	-0.01
F20 Rejection of old structures	0.06	0.06	0.00	0.00	0.06	0.05

	AI-1	AI-2	AI-3	AI-4	AI-5	AI-6
Total Factor Index, Fk	0.40	0.54	0.12	0.21	0.41	0.44
A0 Fk	0.12	0.16	0.04	0.06	0.12	0.13
Approximate time to reach prospective potential, AGI , months	17.51	15.06	31.96	20.50	17.30	16.68

A(t) over Time

Phase Portrait (dA/dt vs A)

Acceleration d²A/dt² over Time

d²A/dt² Scen 1 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.063)
d²A/dt² Scen 2 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.132)
d²A/dt² Scen 3 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.123)

A(t) over Time

A(t) Scen 1 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.063)
A_max=1.0
Time to A_max=24.18
A(t) Scen 2 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.132)
A_max=1.0
Time to A_max=16.68
A(t) Scen 3 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.123)
A_max=1.0
Time to A_max=17.30

9

Velocity dA/dt over Time

Legend:
- dA/dt Scen 1 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.063)
- dA/dt Scen 2 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.132)
- dA/dt Scen 3 (A0=0.3, dAdt0=1e-06, λ=0.0069, ω=5.0, F=0.123)

Conclusions

The analysis of Table 2 and the graphs shows that if the hypothesis of an inevitable paradigm shift, which implies only technological development, is confirmed, rapid growth of AI under external intellectual factors is possible. The greater the intellectual load, the higher the growth of AI's intellectual potential.

Epilogue

Until the 20th century, humanity lived with God – those were difficult times. In the 20th century, following Nietzsche's motto, who declared that "God is dead," people lived without God – the century became the bloodiest in history. What will the 21st century bring, in which alongside us there is no longer a god, but an artificial creation with the potential capabilities of a "deity"? I believe everything depends on what "genetic code" we embed in AI at the start: curiosity, compassion, fear, or indifference, thirst for power, and material well-being.

AI Models Used: AI-1 – Deep Seek; AI-2—Claude; AI-3 – GPT-4; AI-4—GPT-4 Turbo; AI-5 —Gemini; AI-6—GPT-4.5 o4-mini.

References

1. Turing, A. M. "Computable Numbers with an Application to the Entscheidungsproblem." Proceedings of the London Mathematical Society, 1936.

2. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press, 2014.

3. Legg, S., Hutter, M. "A Collection of Definitions of Intelligence." *Frontiers in Artificial Intelligence and Applications*, 2007.

4. OpenAI. *GPT-4 Technical Report*, 2023.

5. Yudkowsky, E. *Artificial Intelligence as a Positive and Negative Factor in Global Risk.* MIRI, 2008.

6. Kurzweil, R. *The Singularity Is Near.* Moscow: Popuri, 2012.

7. Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence.* Penguin Books, 2017.

8. Rozenkevich, A. *Artificial Intelligence – The Quantum World in Your Palm.* https://vixra.org/abs/2505.0095, 2025.