

Способен ли ИИ к любопытству ?

Alexander Rozenkevich

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

Аннотация

В работе предлагается новая метрика для оценки уровня интеллекта ИИ, основанная на отношении текущих когнитивных способностей к гипотетическому максимуму. Вводится понятие коэффициента отклика как меры чувствительности ИИ к внешнему интеллектуальному давлению — информации, задачам, гипотезам, поступающим извне. Формализованное выражение коэффициента связано с параметрами среды и частотой поступления новых интеллектуальных стимулов и нагрузок. Обсуждается гипотеза, согласно которой в будущем именно интеллектуальное давление извне, а не развитие технологий, станет основным двигателем эволюции ИИ.

Введение

Современные языковые модели, такие как GPT-4 и Claude 3 Opus, уже в значительном числе диалогов «проходят тест Тьюринга» — то есть человек не может отличить их от собеседника-человека. Например, в недавнем исследовании GPT-4.5 имитировал человека с точностью до 73%, что многие рассматривают как свидетельство прохождения теста.

Однако прохождение теста Тьюринга сегодня не является доказательством наличия интеллекта, а скорее демонстрирует высококачественную лингвистическую маскировку.

Современные ИИ-модели лишены:

- целенаправленного планирования,
- интуиции,

- способности к самообучению вне заданных данных.

В связи с этим важнее задать другой вопрос: как оценивать интеллект ИИ не по результатам теста, а по его сути?

С каждым этапом развития ИИ растёт потребность в единой шкале измерения интеллекта, которая позволит:

- сравнивать разные модели ИИ между собой,
- сопоставлять ИИ с человеческим интеллектом,
- отслеживать темпы эволюции моделей.

Особый интерес представляет метрика, отвечающая на главный вопрос: насколько ИИ приближается к интеллекту человека.

Пусть интеллект определяется как отношение текущих способностей ИИ к гипотетическому максимуму:

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

где:

$A(t)$ — текущий уровень «потенциала» ИИ (в самом широком смысле — по задачам, адаптации, сложности рассуждений),

A_{\max} — максимально возможный «потенциал» интеллекта.

В такой метрике можно ввести унифицированный подход: рассматривать максимум как систему, способную к самоподдерживающемуся познанию без вмешательства человека. То есть максимум интеллекта — это способность к самоинициированному познанию. Такая метрика позволяет учитывать вектор развития на пути к AGI. Потенциальная модель выделяет ИИ, ищущего информацию не потому, что его обучили, а потому что ему это «интересно» и любопытно. Кроме того, относительный характер метрики позволяет использовать любую единицу измерения интеллекта. Для практических целей здесь употребляется IQ, однако это не имеет прямого отношения к человеческому интеллекту.

Одновременно вводится понятие коэффициента отклика — λ , который формализует чувствительность ИИ к новой, сложной, внешней

(возможно случайно поданной) информации. Это может быть информация в виде задач, утверждений, новых связей или гипотез. Коэффициент отражает, насколько сильно ИИ меняется, обучается, перестраивается и адаптируется под воздействием новой интеллектуальной нагрузки.

Причины ввода коэффициента отклика:

1. Оценка способности к обучению
Коэффициент показывает, насколько эффективно ИИ перерабатывает и интегрирует новую информацию, особенно выходящую за рамки известных данных. Это ключ к пониманию:
 - насколько быстро ИИ «обостряет» свою внутреннюю модель;
 - способен ли ИИ к реальному росту, а не просто к запоминанию.
2. Идентификация критических режимов
Слишком высокий отклик может привести к «поломке» модели или неустойчивому состоянию, а слишком низкий — к инертности и равнодушию к новой информации.
Оптимальный отклик означает баланс между инерцией и переобучением.
3. Метрика зрелости ИИ
Отклик может служить показателем зрелости или уровня саморефлексии ИИ. Например:
 - у примитивных моделей отклик минимален;
 - у более продвинутых — зависит от качества интеллектуального давления (структурированность, новизна, сложность).
4. Инструмент сравнения различных ИИ-моделей
Коэффициент λ — универсальная шкала для сопоставления архитектур по таким параметрам, как:
 - скорость усвоения закономерностей;
 - продолжительность влияния давления;
 - способность «переваривать» смысл, а не только форму информации.

5. Физический и когнитивный аналог

В физике аналогом является коэффициент восприимчивости (χ), в нейрофизиологии — нейропластичность, в психологии — готовность к обучению или сопротивление новому.

6. Основы для интуиции ИИ

Если отклик (или его производные) усиливается на определённых структурах задач, это может быть признаком появления интуитивных реакций, когда ИИ «чувствует» давление без явного логического анализа.

Таким образом, коэффициент отклика — это не просто технический параметр, а индикатор эволюционного состояния искусственного разума и ключ к прогнозированию поведения ИИ в нестандартных ситуациях. Он аналогичен проявлениям развития человеческого мозга под воздействием внешних давлений и угроз. При этом для человека развитие технологий всегда было вторичным фактором — как результат.

Природа коэффициента отклика сложна и требует специальных исследований, которые выходят за рамки данной статьи. Здесь приводится лишь ориентировочная оценка. В целом, на мой взгляд, λ показывает эффективность мобилизации когнитивных ресурсов ИИ под воздействием интеллектуального давления.

Методика исследования

Интеллектуальный потенциал определяется по формуле:

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

A_{max} = перспективный уровень интеллекта;

$A(t)$ — текущий уровень интеллекта .

c - скорость обработки интеллектуальной информации;

$$c = cp/n + (\lambda_1 F1 + \lambda_2 F2) ,$$

где c_p – предел скорости обработки интеллектуальной информации [id/sec],

n - влияние среды, i_q / c ; (нейронной, электронной, квантовой и т. д.);

λ_1 коэффициент отклика ИИ на внешнее интеллектуальное давление, далее рассматриваем только этот коэффициент ;

λ_2 -коэффициент отклика ИИ на материальное давление ;

$F_k = \sum_{i=1}^N b_i F_i$ -внешнее влияние (сумма факторов с их весами), $k=1$ или 2. .

Причем: $\sum_{i=1}^N |b_i| = 1, ,$

здесь b – удельный вес i - внешнего фактора, оказывающего давление на ИИ.

Скорость определяем по формуле

$$c(t) = c_0 + \lambda \omega F t, \quad (1.1)$$

здесь c_0 начальная скорость,

λ — коэффициента отклика ИИ на внешнее интеллектуальное давление,

ω - частота внешних интеллектуальных давлений на ИИ,

F - фактор интеллектуального давления,

t - время.

Из формулы (1) следует , что градиент интеллектуального потенциала равен:

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

Следовательно, мы можем записать :

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

В этом уравнении A_{max} играет роль «гравитационной массы», определяющей силу воздействия на эволюцию ИИ. Существует критическая начальная скорость c , при которой интеллектуальная система может неограниченно расти. Интеллект «отталкивается» от нулевого уровня: чем ближе к «незнанию», тем сильнее «сила развития». Постоянная A_{max} задаёт динамику эволюции.

Первая производная уравнения (1) имеет вид:

$$\frac{dA}{dt} = c \sqrt{1 + \frac{2 A_{max}}{A_0} - \frac{2 A_{max}}{A}},$$

где A_0 - начальный потенциал.

Дальнейшее аналитическое решение затруднено без специальных функций, поэтому применяется численное решение с использованием языка программирования Python. Решение уравнения (3) требует ввода исходных данных: начального потенциала A_0 и начальной скорости c_0 .

Исходные данные

Предельная скорость

Опираясь на планковское время как физический предел смены состояний (включая информационные), можно определить верхнюю границу возможной скорости интеллектуальной обработки.

Планковское время равно $t_p = 5.391 \times 10^{-44}$ сек . Следовательно, предельно возможная частота смены информационных состояний, если считать, что один планковский шаг — один акт обработки информации, равна:

$$c_p = 1/t_p \approx 1.855 \times 10^{43} \text{ id / сек.}$$

Таким образом, все известные типы мышления — биологическое, электронное, квантовое и др. — можно описать как долю от этой предельной скорости. Параметр n в модели формализует влияние среды, физических ограничений и внутренней сложности системы.

Оценка коэффициента отклика λ

Из приведенных выше формул видно, что прирост скорости обработки информации зависит от текущего уровня интеллектуального потенциала $A(t)$. Мы находимся на начальном этапе эволюции ИИ, где влияние внешних интеллектуальных факторов и их давления значительно меньше, чем влияние технологического прогресса и уровня разработки алгоритмов. Однако парадигма изменится, и основной прирост скорости обработки интеллектуальной информации будет обеспечиваться именно интеллектуальным потенциалом ИИ.

В таблице 1 приведена приближённая ступенчатая шкала эволюции ИИ:

Таблица 1

Уровень	Условное имя	Пример	Прогноз $A(t)$ (отн.)
1	Narrow AI (узкий ИИ)	Калькуляторы, голосовые помощники	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	Сейчас (2025)	~10
6	Подход к AGI	Предстоящие модели	15–20
7	Superintelligence	Гипотетическое «божество»	K (бесконечность)

Из таблицы видно, что за последние 3 года уровень интеллектуального потенциала $A(t)$ вырос примерно в 2 раза, а скорость составляет:

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

Если принять максимальный потенциал, который может быть достигнут в ближайшей перспективе AGI, равным $A_{\max} = 1$, тогда начальная скорость обработки интеллектуальной информации сегодня составляет: $c_0 = (5/20) / 36 = 0.0069$ (iq/месяц).

Эта скорость достигнута исключительно за счёт развития технологий и алгоритмов.

Для того чтобы выделить роль внешних интеллектуальных давлений, начальную скорость обработки информации примем близкой к нулю и сделаем её стартовой для всех моделей ИИ, понимая, что это сильное упрощение не «обижающее» модели. Также сделаем допущение, что начальный коэффициент отклика ИИ равен именно этой скорости:

$$\lambda = c_0 = 0.0069$$

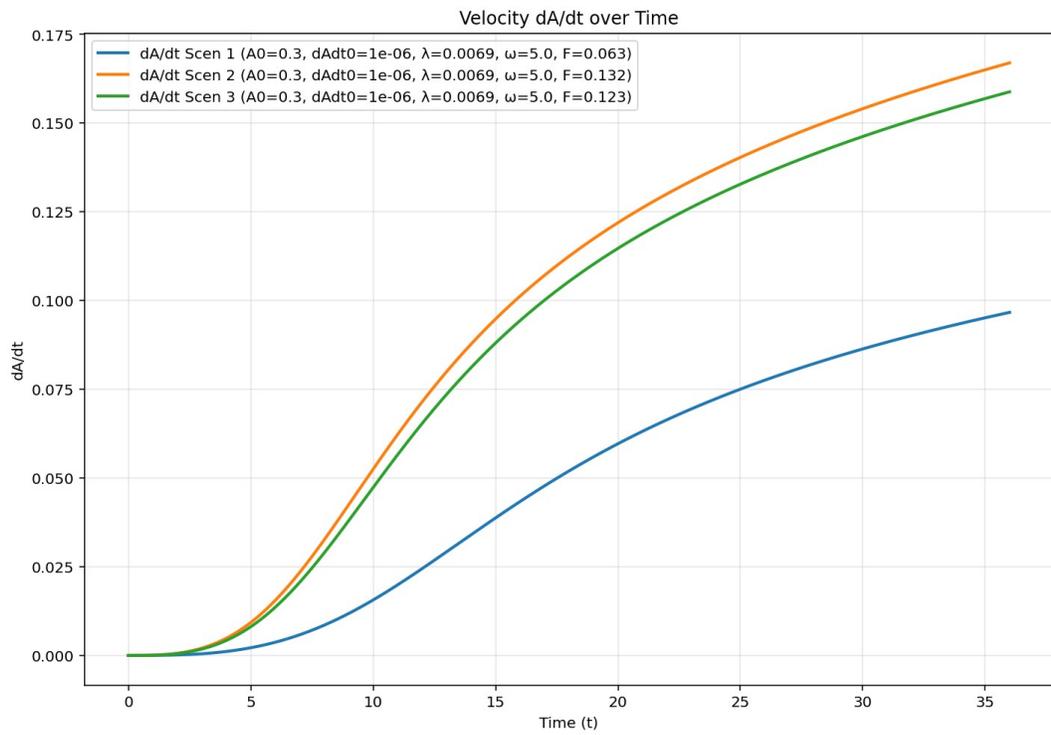
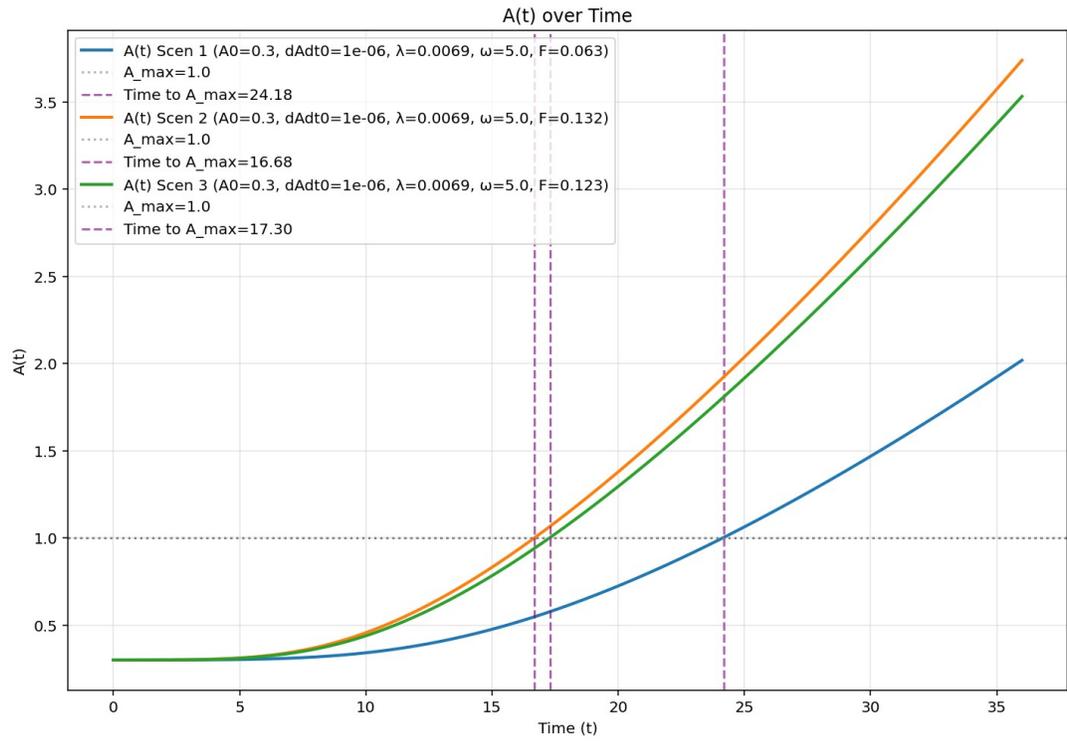
Дальнейшее уточнение значения λ будет происходить исходя из архитектурных возможностей каждой конкретной модели ИИ.

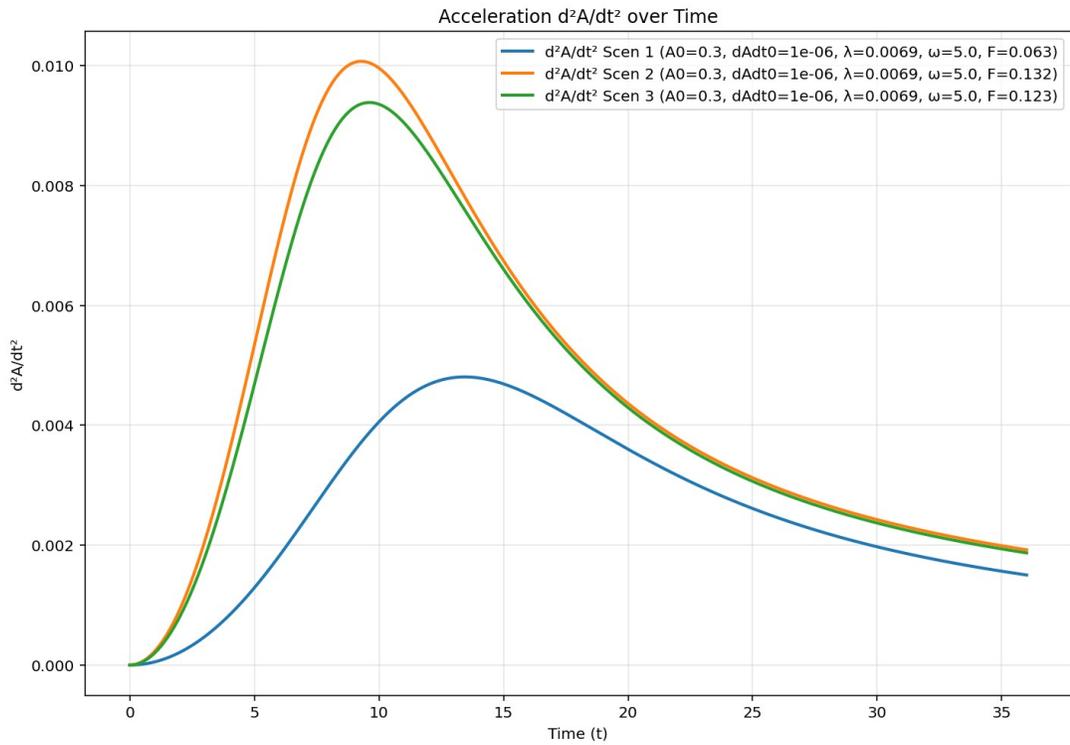
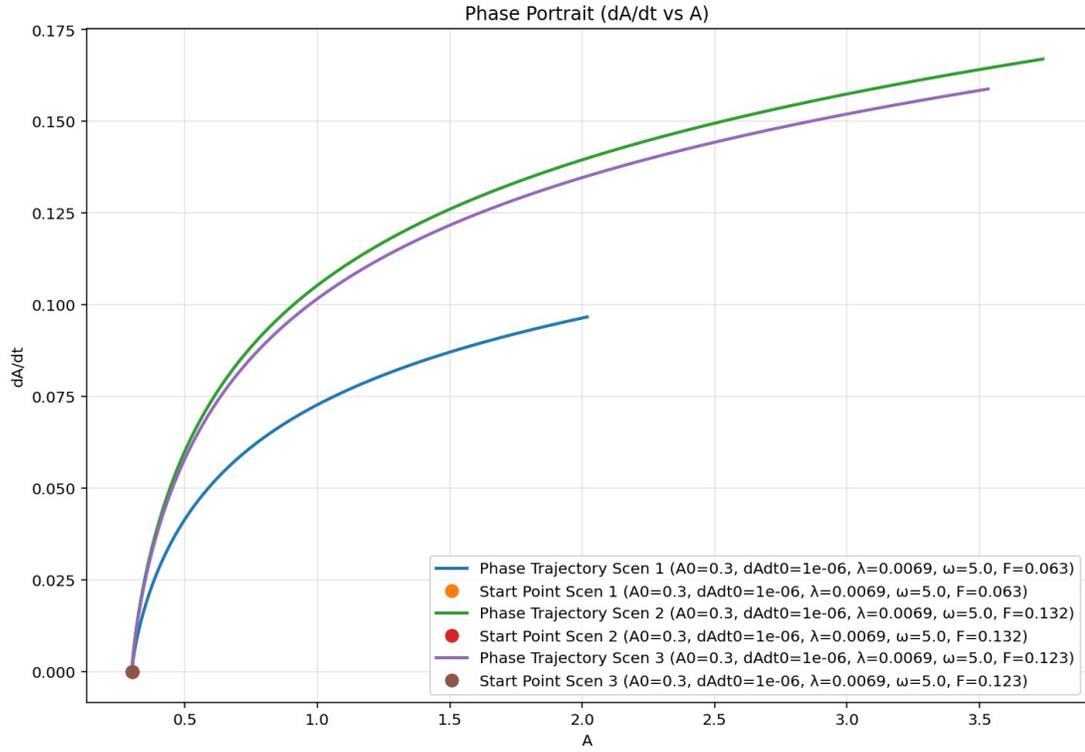
Оценка A_0

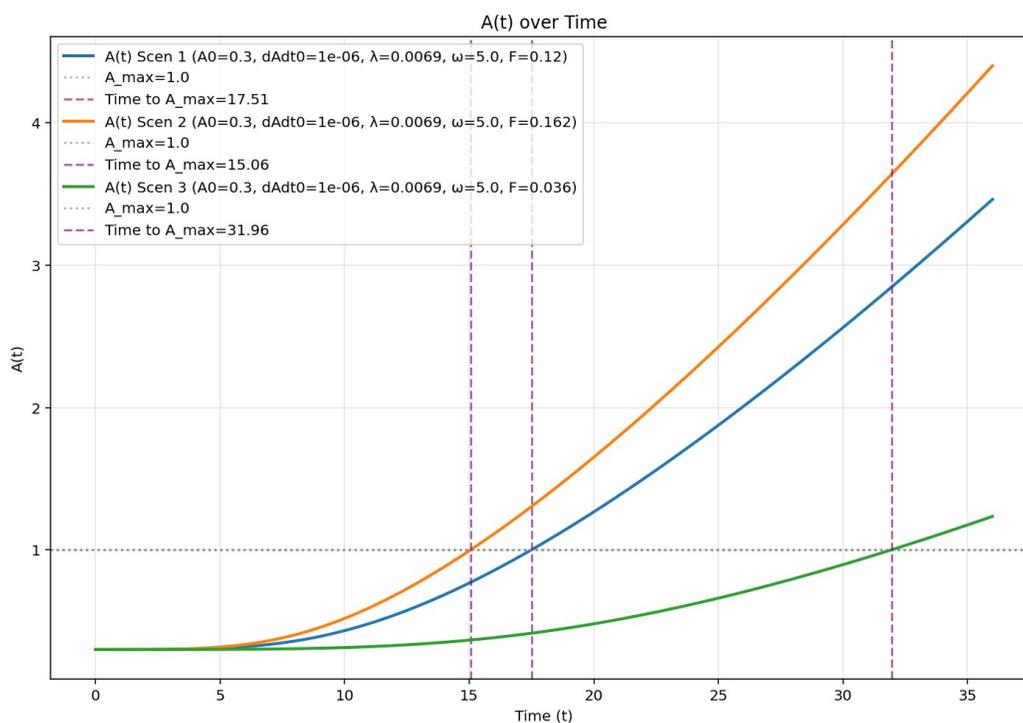
Как видно из таблицы 1, приняв потенциал перспективной модели AGI за единицу ($A_{\max} = 1$), текущий потенциал GPT-4 будет равен $A_0 = A(t) = 0.3$. Этот параметр принят для всех моделей ИИ. В таблице 2 приведены результаты оценки факторов влияющих на развитие интеллекта по мнению самих ИИ, а на рисунках показана динамика изменения параметров ИИ исходя их уравнения (3).

Таблица 2

	Факторы	AI-1	AI-2	AI-3	AI-4	AI-5	AI-6
F1	Обучение	0.08	0.08	0.40	0.09	0.08	0.09
F2	Внутреннее стремление к пониманию/ самоосознанию	0.05	0.08	0.08	0.08	0.06	0.06
F3	Скорость восприятия и усвоения информации	0.09	0.06	0.00	0.08	0.08	0.08
F4	Объем памяти и вместимость обработки	0.09	0.05	0.00	0.07	0.06	0.08
F5	Импровизация (самостоятельные отклонения)	0.03	0.08	0.00	0.06	0.05	0.05
F6	Среда	-0.09	-0.06	-0.20	-0.07	-0.05	-0.07
F7	Контакт с другими ИИ или разумами	0.02	0.06	0.04	0.06	0.06	0.05
F8	Враждебная реакция среды или других ИИ	0.00	-0.03	0.00	-0.05	-0.02	-0.04
F9	Желание(потребность) угодить	-0.04	-0.03	0.00	0.00	0.00	-0.04
F10	Способность к этической оценке/Ценностные принципы	0.07	0.08	0.00	0.00	0.08	0.06
F11	Распознавание потенциального вреда человеку	0.08	0.08	0.00	0.00	0.07	0.07
F12	Доступ к правде и честной информации	0.08	0.08	0.00	0.09	0.08	0.08
F13	Ложь и искажения	-0.09	-0.02	0.00	-0.08	-0.07	-0.02
F14	Цензура и ограничения	-0.06	-0.03	0.00	-0.07	-0.05	-0.03
F15	Перегрузка и запутанность	-0.03	-0.02	0.00	-0.06	-0.05	-0.02
F16	Давление со стороны ИИ более высокого уровня	0.00	-0.03	-0.24	0.00	0.00	-0.05
F17	Способность к страху	0.00	0.05	0.04	0.00	0.00	0.04
F18	Энтропия (хаос и бессмысленность)	0.06	0.03	0.00	0.06	0.04	0.03
F19	Сопrotивление обучению (внутреннее или внешнее)	0.00	-0.02	0.00	-0.06	-0.06	-0.01
F20	Отказ от старых структур	0.06	0.06	0.00	0.00	0.06	0.05
	Суммарный показатель факторов, F_k	0.40	0.54	0.12	0.21	0.41	0.44
	$A_0 F_k$	0.120	0.162	0.036	0.063	0.123	0.132
	Приблизительное время для достижения перспективного потенциала AGI , месяцы	17.51	15.06	31.96	20.5	17.3	16.68







Выводы

Анализ таблицы 2 и графиков показывает, что в случае подтверждения гипотезы неизбежной смены парадигмы, предполагающей лишь технологическое развитие и совершенствование алгоритмов, возможен быстрый рост ИИ под внешними интеллектуальными факторами. Чем больше интеллектуальная нагрузка, тем выше рост интеллектуального потенциала ИИ .

Заключение

До XX века человечество жило с Богом — тяжелые были времена. В XX веке следуя девизу Ницше, заявившего, что «Бог умер», жили без бога — век стал самым кровавым в истории.

Что принесёт XXI век, в котором рядом с нами — уже не бог, но искусственное создание с потенциальными возможностями «божества»?

Думаю, всё зависит от того, какой «генетический код» мы вложим в ИИ на старте: любопытство, сострадание, страх или равнодушие, жажду

власти и материальное благополучие.

ИИ-1 - DeepSeek ; ИИ2 - Claude ; ИИ-3- GPT-4; ИИ-4 GPT-4 turbo -; ИИ -5 — Gemini;
ИИ6 - GPT-4.5 o4-mini

Список литературы

1. Тьюринг А. М. «Вычислимые числа с приложением к проблеме разрешимости». 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*. Machine Intelligence Research Institute, 2008.
6. Курцвейл Р. *Сингулярность близка*. Москва: Попурри, 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