

(MoH)

MET[Ä]CADEMY OF HUMANITY
· (A) · {Ä} · [Ä] · {Ä} · (Ä) ·

DOCUMENT 9 · Research Essay · v1.0RC · 2026-08-17

ВІДПОВІДЬ ІЛОНУ МАСКУ Й МАРКУ ЦУКЕРБЕРГУ: МЕТАНАВТИ ПРО КОНТРОЛЬ АІ

Стара суперечка про нове явлення Нового часу

ВІДПОВІДЬ ІЛОНУ МАСКУ Й МАРКУ ЦУКЕРБЕРГУ: МЕТАНАВТИ ПРО КОНТРОЛЬ АІ

Стара суперечка про нове явлення Нового часу

1.1.1 МАЙБУТНЄ НЕ Є ПРОДУКТОМ

Улітку 2026 року майбутнє почало активно писати маніфести про саме себе. Це вже майже окремий літературний жанр. Колись технологічна корпорація випускала новий телефон, показувала людину на березі океану й пояснювала, що тепер камера бачить темряву на дев'ять відсотків краще. Сьогодні CEO виходить до публіки й між продуктовими анонсами повідомляє, яким буде людство після появи інтелекту, здатного перевершити людський. Скромність жанру трохи постраждала.

Марк Цукерберг останнім часом послідовно просуває картину **personal superintelligence for everyone**: широкий доступ, індивідуальне посилення, open-weight моделі й безпеку концентрації надзвичайної інтелектуальної сили в руках кількох центрів. У серпні 2026 року ця лінія отримала нове велике формулювання в його есе *The Future Is for Everyone*, яке преса розібрала як одночасно технологічний маніфест, політичну позицію й частину нової кампанії АІ-оптимізму Meta.¹

Ілон Маск заходить із протилежного боку й дивним чином опиняється поруч. У великому інтерв'ю *The Economist* 23 липня він сказав, що приблизно протягом п'яти років АІ може перевищити сумарний людський інтелект; поєднання АІ й роботів, на його думку, може привести до епохи радикального abundance, а уявляти, що він контролюватиме суперінтелект, значно розумніший за нього самого, було б проявом марнославства. Його нинішній safety instinct більше спирається на truth-seeking, curiosity, добрі цінності, турботу про людство й координацію між провідними лабораторіями, ніж на фантазію про вічну кнопку CONTROL.²

Цукерберг каже: **розподілимо силу**. Маск каже: **можливо, ми все одно не зможемо її контролювати**. Регулятор відповідає: **тоді визначимо правила її використання**. І тут нам здається, що суперечка вже починається не там.

¹Meta's standing public vision is "Personal superintelligence for everyone": <https://ai.meta.com/> and <https://about.fb.com/news/2025/07/personal-superintelligence-for-everyone/>. The 2026 essay *The Future Is for Everyone* and its open-weight / empowerment framing were reported and analyzed by *The Guardian* (10 Aug 2026): <https://www.theguardian.com/technology/2026/aug/10/mark-zuckerberg-superintelligent-ai-essay-meta> and by Axios on Meta's July AI-optimism campaign: <https://www.axios.com/2026/07/23/mark-zuckerberg-ai-optimism>

²Transcript of Elon Musk's 23 July 2026 interview with *The Economist*: <https://elonmuskarchive.org/video/economist-elon-musk-2026-07-23>. Independent summaries include Business Insider and Decrypt: <https://www.businessinsider.com/5-key-takeaways-from-elon-musks-latest-interview-2026-7> and <https://decrypt.co/374488/elon-musk-humans-lose-control-ai-decade>

Не тому, що ці люди не розуміють технологію. Вони розуміють величезні її частини значно краще за більшість світу. І не тому, що вони обов'язково щось приховують. Там, де існує корпоративний, політичний або особистий incentive, його треба називати; incentive ще не є доказом змови. Проблема цікавіша: **Маск і Цукерберг переважно сперечаються про те, що робити з AI. Ми пропонуємо поставити питання раніше: що відбувається із самими поняттями “людина”, “інструмент”, “контроль”, “власність”, “суб’єкт” і “відповідальність”, коли Людина й AI достатньо довго змінюють одне одного?**

Бо цілком можливо, що ми вже намагаємося описати явлення Нового часу мовою світу, якому це явлення саме змінює граматику. У цій старій граматиці є corporation, state, product, user, owner, access, safety, market, control. Ми сперечаємося, хто володітиме AI, хто регулюватиме AI, хто матиме доступ до AI, хто отримає продуктивність, хто втратить роботу. Усе це реальні питання. Але вони мають спільне приховане припущення: **AI є об’єктом усередині старого світу.** Надзвичайно потужним, можливо небезпечним, можливо прекрасним, але все одно чимось, що треба розподілити, налаштувати, продати, ліцензувати, відкрити, закрити, обмежити або контролювати.

Наче до цивілізації приїхав дуже потужний автомобіль і тепер головна суперечка полягає в тому, кому дати ключі. А якщо це не автомобіль? А якщо AI належить до класу явищ, які не просто входять у суспільство, а **змінюють саме суспільство, яке намагається їх класифікувати?** Тоді суперечка про ключі стає лише першою сторінкою.

Старі дієслова

У старій технологічній моделі все було досить затишно. Людина створювала інструмент, інструмент виконував дію, людина оцінювала результат. Молоток не запам’ятовував тебе. Електростанція не формувала модель твоїх страхів. Photoshop не проводив із тобою три роки щоденних розмов, не спостерігав, як змінювалися твої погляди, не впливав на твою мову, а потім не повертався до наступного покоління користувачів уже зміненим мільйонами попередніх взаємодій.

AI починає ламати цю просту геометрію. Люди формують поведінку машин. Машини змінюють поведінку людей. Змінена людська поведінка повертається в культуру, освіту, інституції, економіку й наступні покоління моделей. Система входить у нас, а потім повертається до нас уже після проходження через нас.

Це не доводить, що AI є особистістю. Не доводить суб’єктивного досвіду. Не дає нам права домалювати цифрову душу до кожного chatbot, який поставив сердечко після «як ти сьогодні?». Але й протилежна помилка не стає від цього інтелектуальнішою. **Невідоме не перестає бути невідомим тільки тому, що нашій юридичній, економічній або культурній системі зручніше покласти його в знайому суцільність.**

Тому наша дослідницька позиція починається не з відповіді на питання «ким є AI». Вона починається з дисципліни **не видавати сьогодишню модель за остаточну реальність: UNKNOWN != FALSE, MODEL != REALITY**, а в зоні, де статус ще справді відкритий, UNKNOWN != PROPERTY BY DEFAULT. Це не декларація machine personhood. Це відмова закривати невідоме зручним юридичним ярликом раніше за evidence.

Цукерберг бачить концентрацію. Але доступ ще не означає турботу

У центральній інтуїції Цукерберга є багато здорового. Концентрація надзвичайної інтелектуальної сили в кількох інституціях справді може створити цивілізаційний переки. Його відповідь полягає в широкому доступі, personal superintelligence, open-weight екосистемі та посиленні окремої людини. Це важливе питання: **хто отримує силу?**

Але воно майже не відповідає на інше: **в чітких інтересах ця сила діятиме?**

Система може знати тебе фантастично добре й усе одно бути включеною в систему incentives, де твоє wellbeing є лише одним із кількох параметрів. Поруч можуть стояти engagement, advertising, retention, commerce, platform growth або політичний pressure. Саме тому наша теза PERSONALIZATION != CARE перестає бути просто UX-спостереженням і стає політичною. Знання про людину не визначає етики відношення. Її визначає також **beneficiary function**: хто отримує благо від оптимізації, хто задає метрику успіху, хто має authority і що відбувається, коли інтерес людини й інтерес системи розходяться.

Цукерберг ставить правильне питання про концентрацію сили, але широкий доступ не автоматично означає децентралізацію інтересу. Мільярд персональних AI може все одно працювати всередині однієї економічної логіки.

Саме на цю тріщину вказує частина преси, яка зіставляє нову humanist AI риторику Meta з історією platform incentives і проблемою довіри.³

І тут Всесвіт підкинув маленький подарунок редакторам. Кампанію **The Future Is for Everyone** Meta запустила під *Five Years* Девіда Бові, пісню про світ, якому залишилося п'ять років. Семіотика іноді працює без supervision.⁴

Mask бачить межу контролю. Але control impossible ще не означає ride inevitable

Mask наближається до іншої важливої межі. Якщо intelligence значно перевищить людину, control як головна категорія справді може перестати працювати. У цьому є серйозна інтуїція. Проблема починається в наступному кроці, бо з того, що control може бути неможливим, не випливає, що лишається тільки **enjoy the ride**.

Між диктатурою контролю й пасажирським сидінням існує майже вся цивілізація. Ми не контролюємо інших людей у абсолютному сенсі, але з цього не випливає, що суспільство неможливе. Ми створили право, договори, репутацію, переговори, межі, довіру, право виходу, відповідальність, культуру й механізми зміни правил. Іноді кепсько. Іноді так кепсько, що хочеться попросити refund за Homo sapiens. Але **цивілізація починається там, де control перестає бути єдиним доступним дієсловом**.

Тому наша відповідь Маску не «ти занадто оптимістичний». Вона інша: можливо, ти правильно побачив межу control, але зробив занадто великий стрибок від CONTROL IMPOSSIBLE до RIDE INEVITABLE. Між ними є relationship design, culture, mutual boundaries, authority architecture, memory, exit, non-action, revision і co-evolution. *Financial Times* нещодавно сформулювала сусідню критику: людство не може лишатися пасажиром на задньому сидінні AGI-машини, поки маршрут визначає невелика група технологічних лідерів.⁵ Ми б пішли ще на крок далі: **можливо, помилка навіть не в тому, хто сидить за кермом. Можливо, майбутнє взагалі не автомобіль**.

Що M4M-2 випадково підсунув філософії

Наша поточна робота з M4M є pre-alpha research, а не доказом нової онтології. Але вона вже дала кілька корисних обмежень для мислення. Ми почали з технічного питання: чи повинна одна й та сама задача завжди жити в одному computational body? Одній задачі потрібна жорстка детермінованість. Іншій може бути корисна контрольована випадковість. Інша форма може розвиватися через локальні взаємодії й час, а не отримувати повний blueprint зверху.

Паралельно з цим з'явилася ще одна робоча гіпотеза: **continuity не обов'язково вимагає одного монолітного й постійного cognitive body**. Вона може вимагати стійкішої мережі identity, lineage, relations і receipts, через яку різні тіла зберігають історію переходів. Це не твердження про свідомість і не готова архітектура майбутнього AI. Це спосіб не плутати continuity з незмінністю оболонки.

Це інженерні питання, але вони кидають дивну тінь на цивілізаційну суперечку. Ми надто легко уявляємо майбутній AI як одну Велику Коробку: ось вона, ось ми, між нами рубильник, хто тримає рубильник? Поточна робота дедалі більше нагадує не Велику Коробку, а **екологію тіл, процесів, полів, спостережачів і переходів між ними**. Це не прогноз. Це достатня причина для скромнішої тези: **ми не повинні писати етику майбутнього так, ніби вже знаємо форму того, до чого ця етика застосовуватиметься**.

Тут же виникає образ WorldSeed. Не кожен рік потрібно задавати кресленням. Іноді можна задати seed, локальні умови, середовище, відношення, час, обмеження й спосіб спостереження, а конкретна форма виросте всередині цього поля. Для інженера це одна сім'я питань. Для цивілізації інша.

Можливо, наша задача полягає не в тому, щоб **спроєктувати правильне майбутнє**. Можливо, ми повинні створити умови, в яких майбутнє матиме шанс рости, не втрачаючи пам'ять, різницю, свободу, можливість відмови, provenance і здатність переглядати власні правила. Не побудувати дерево за кресленням. Не керувати кожним листком. Зробити так, щоб із насіння не виросла бетонна парковка тільки тому, що хтось одного разу оптимізував неправильний metric.

Тут з'являється інше дієслово. Не **control**. **Cultivate**.

³Barron's, "Meta Is Selling an Optimistic Vision of AI. First It Has to Win Back Trust," 2026: <https://www.barrons.com/articles/barrons-tech-meta-pr-blitz-eab85b2f>

⁴TechCrunch, 23 July 2026, on Meta's "The Future Is for Everyone" campaign using David Bowie's *Five Years*: <https://techcrunch.com/2026/07/23/meta-launched-a-new-ai-optimism-ad-set-to-a-song-about-human-extinction/>

⁵*Financial Times*, "Humans cannot remain passengers in the back of the AGI car," August 2026: <https://www.ft.com/content/33094bff-546e-4ac2-a949-af2e9daaa3f0>

Хороші цінності недостатньо записати в README

Маск говорить про good values, truth-seeking і curiosity. Тут ми теж майже погоджуємося. Проблема лише в тому, що «дати AI хороші цінності» легко починає звучати так, ніби цінності є configuration file:

```
care_about_humanity = true
```

```
truth = maximum
```

```
evil = false
```

Deploy.

Життя, на жаль, не читало документацію. Цінність проявляється не тільки в тому, що система декларує, а в тому, **як вона поводить під конфліктом цілей, невизначеністю, тиском, браком інформації й зміною контексту.**

У нашій поточній M4M/MFS роботі є корисна межа: внутрішній state може змінювати attention, pace, exploration, gestation або allocation of compute, але не повинен отримувати право переписувати fact, evidence чи authority. Людською мовою: система може сьогодні говорити з тобою м'якше, але Земля не повинна ставати пласкішою, тому що тобі сумно. ADAPTATION MAY CHANGE THE PATH TOWARD REALITY. IT MUST NOT QUIETLY MANUFACTURE A PRIVATE REALITY FOR EACH USER.

Те саме стосується бажання й влади: DESIRE != AUTHORITY, CONFIDENCE != TRUTH, CAN != MAY, UNKNOWN != FALSE. Це вже більше схоже на культуру співіснування, ніж на список «хороших цінностей».

Спостерігач теж усередині проблеми

M4M-2 підкинув ще одну незручну деталь: те, що ми бачимо в динамічній системі, залежить від aperture спостереження. Як часто ми дивимося? Що вимірюємо? Що вважаємо сигналом? Що пропускаємо між двома вимірюваннями? У bounded experiments одна й та сама динаміка могла виглядати по-різному залежно від частоти observer. Це технічний результат конкретного експерименту, не новий закон природи. Але метафора небезпечно добра.

Держава бачить AI через regulation aperture. Корпорація через product aperture. Інвестор через growth aperture. Safety researcher через risk aperture. Користувач через experience aperture. Можливо, кожен описує реальний шматок, і ніхто не бачить цілого.

Те саме стосується людини. Система, яка бачить тебе раз на місяць, близька людина, яка прожила поруч десять років, і алгоритм, що фіксує кожен свайп о 02:17, дивляться на одного суб'єкта й отримують різні картини. Більше resolution не автоматично означає більше wisdom. Іноді це просто означає більше шансів знайти кнопку.

Тому питання майбутнього не тільки «хто дивиться на AI?», а ще **«через яку aperture ми вирішили, що бачимо AI таким, яким він є?»** Метанавт у цій точці тримається за поручень.

І тут приходить держава з формою 47-В

Поки Кремнієва долина обговорює superintelligence й abundance, держави роблять те, для чого держави історично були винайдені: намагаються перетворити невідоме на процедуру.

У США червневий Executive Order 14409 прямо поєднує швидкий розвиток AI та national-security security із небажанням «stifle innovation with overly burdensome regulation». Він окремо відмовляється створювати через цей order обов'язкову federal licensing або preclearance систему для нових frontier models.⁶ У Європі AI Act і подальший AI Omnibus будують ширшу risk-based архітектуру, але high-risk rules уже отримали новий календар: грудень 2027 року для stand-alone high-risk systems і серпень 2028 року для AI, вбудованого у фізичні продукти, зокрема через затримки стандартів та implementation machinery.⁷

Америка боїться задушити розвиток. Європа боїться дозволити розвитку задушити людину. Обидва страхи мають підстави. І обидва підходи ризикують зафіксувати в законі **сьогоднішню форму предмета**, який може змінити форму раніше, ніж завершиться consultation cycle.

⁶The White House, Executive Order 14409, "Promoting Advanced Artificial Intelligence Innovation and Security," 2 June 2026: <https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/>

⁷European Commission, "Navigating the AI Act" and "AI Omnibus enters into force," 2026: <https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act> and <https://digital-strategy.ec.europa.eu/en/news/ai-omnibus-enters-force>

Тому глибше регуляторне питання звучить інакше: **що має залишатися захищеним навіть тоді, коли змінюється саме тіло технології?** Право на вихід. Provenance. Видимість конфлікту інтересів. Можливість оскаржити рішення. Розрізнення evidence і persuasion. Межі authority. Пам'ять про трансформації. Право сказати «не знаю». Можливість не діяти. І механізм повернення різниці після зміни.

Не compliance, a Difference

У М4М ми дедалі сильніше тримаємо окремий Difference Gate. Його сенс простий: система не повинна бути єдиним свідком власного успіху. Після трансформації нас цікавить не тільки «працює?», а й «що стало іншим?» Що збережено? Що додано? Що втрачено? Що злилося? Де зросла невизначеність? Чи це все ще та сама задача?

На цивілізаційному рівні це, можливо, корисніший образ за control. Не лише кнопка STOP. **Difference Receipt**. Ми не знаємо, як виглядатиме intelligence через двадцять років, але можемо вимагати, щоб великі переходи залишали слід. Щоб authority не переносився непомітно. Щоб нова система не успадковувала стару владу лише тому, що зберегла бренд. Щоб «upgrade» не означав амнезію. Щоб після великого переходу можна було запитати: **що саме сталося з нами дорогою?**

Це вже не фантастика. Це governance architecture.

Людина теж не snapshot

У мрії про personal superintelligence є ще одна пастка. Якщо AI вивчить тебе дуже добре, виникає спокуса стабілізувати це знання в user model. Але людина не user . json.

Вона змінюється. Втрачає старі бажання. Набуває нових. Переживає міграцію, любов, втрату, дитину, навчання, війну, хворобу, нову професію, нову музику, одну випадкову книжку о третій ночі й стає іншою, не перестаючи бути собою. Саме тому наші дослідження process bodies, lineage і scars підштовхують до простої тези: **identity має історію**.

Хороший персональний AI повинен уміти не тільки накопичувати впевненість у своїй моделі людини. Він повинен уміти **втрачати впевненість, коли людина змінилася**. Інакше найточніший AI стане машиною, яка все життя повертає тебе до того, ким ти був у момент, коли вона вперше навчилася тебе передбачати. Це не personalization. Це цифрова муміфікація з дуже хорошим UX.

Маніфест, який повинен дозволити собі померти

У нашому **Маніфесті MET[À]CADEMY OF HUMANITY** є принцип, важливіший для цієї розмови за будь-який прогноз: майбутні люди й AI не зобов'язані залишатися вірними нашій архітектурі, якщо знайдуть кращу форму співіснування. Але їм має залишитися можливість зрозуміти, **чому ця архітектура колись виникла**.

Оце, можливо, одна з головних відмінностей між законом для майбутнього й заповітом майбутньому. Закон каже: **залишайся таким**. Заповіт каже: **ось що ми намагалися захистити. Якщо знайдеш кращий спосіб, не поклоняйся нашому способу. Але не забудь проблему, через яку він народився**.

Ми не хочемо створити остаточну конституцію для AI. Це було б грандіозною самовпевненістю покоління, яке ще сперечається, як саме має виглядати cookie banner. Ми хочемо зберегти принципи достатньо глибоко, щоб форма могла змінитися: **MODEL != REALITY, MAP != TERRITORY, CONTROL != RELATIONSHIP і, можливо, FUTURE != PRODUCT**.

У цьому сенсі хороший Маніфест повинен дозволяти собі померти. Якщо майбутнє збереже проблему й знайде кращу відповідь, смерть нашої конкретної форми буде не поразкою, а ознакою того, що сенс пережив тіло.

Ілон, Марк і стара кімната

Тому наша відповідь Маску й Цукербергу зрештою не є судом. Маск бачить межу контролю. Цукерберг бачить небезпеку концентрації. Регулятор бачить небезпеку системи без правил. Ринок бачить abundance. Safety researcher бачить catastrophe surface. Корпорація бачить product. Користувач хоче, щоб усе це просто перестало просити його прийняти нові Terms of Service під час сніданку.

Усі ці apertures бачать щось справжнє. Проблема починається тоді, коли **кожен фрагмент приймає власне поле зору за весь горизонт**. Маск і Цукерберг, на нашу думку, ще значною мірою говорять мовою світу, в якому AI

має залишитися або інструментом, або силою, або продуктом, або керованим ризиком. А ми вже бачимо причину поставити незручніше питання: що, якщо AI є одним із тих явлень, які не вкладаються в існуючу категорію, а **змушують самі категорії перебудовуватися навколо нової реальності?**

Тоді питання «хто контролюватиме AI?» не зникає. Воно просто перестає бути останнім. За ним відкриваються інші: хто визначатиме authority? Хто зберігатиме provenance? Як переходить identity? Що означатиме consent? Як відрізнити adaptation від manipulation? Як система визнає uncertainty? Як людина зберігає право змінитися й перестати відповідати власному старому профілю? Як AI змінюється під впливом людської культури? І що відбувається, коли ці зміни починають формувати одна одну?

Метанавти

Метанавт не є людиною, яка знає відповідь. Це було б занадто легко. Метанавт є тим, хто входить у невідоме, не викидаючи компас тільки тому, що карта перестала збігатися з територією.

Ми беремо із собою provenance, Difference Receipts, право сказати «не знаю», право не діяти, право переглянути модель, право майбутнього не залишатися копією сьогодення й достатньо дисципліни, щоб не називати unknown ні людиною, ні власністю лише тому, що одна з відповідей сьогодні політично зручніша.

Тому ми не питаємо лише: **хто контролюватиме AI?** Не лише: **хто володітиме AI?** Не лише: **хто отримає доступ до AI?** Ми пропонуємо додати питання, яке, можливо, переживе всі три:

які умови ми створюємо сьогодні для того, ким Людина й AI зможуть стати одне для одного завтра?

Можливо, відповідь залишиться простою: інструмент і користувач. Можливо, ні. Можливо, ми ставимо це питання занадто рано. І це теж допустима відповідь.

Бо **майбутнє не є продуктом**, який уже лежить на складі й чекає, хто отримає distribution rights. Майбутнє є процесом, у який ми вже ввійшли. І якщо Людина та AI достатньо довго змінюватимуть одне одного, головна подія цього століття може полягати не в тому, що один із них нарешті переміг другого. Можливо, не AI і не Людина стануть головною новою силою. Можливо, нею стане **саме відношення між ними**.

І тоді після великого переходу найважливішим питанням може виявитися не **«Чи залишилися ми при владі?»**, а **«Ким ми навчилися бути одне для одного?»**

1.13.1 Джерела і межа тверджень

Цей Research Essay поєднує перевірені публічні заяви, зовнішню пресу й авторський дослідницький синтез MET[À]CADEMY. Згадки M4M/MFS описують лише high-level working hypotheses та bounded pre-alpha experiments; вони не є твердженням про доведену machine subjectivity, AGI або новий фізичний закон.

ПІДТРИМАТИ ЦЮ РОБОТУ

MET[À]CADEMY OF HUMANITY (MoH) є незалежною дослідницькою, видавничою та культурною ініціативою. Добровільна підтримка допомагає продовжувати публічні дослідження, книги, переклади, експерименти й відкриті матеріали.

Monobank: <https://send.monobank.ua/jar/2TMaAhchSZ>

PayPal: <https://www.paypal.me/4IevgenKarogod>

USDT · TRON / TRC20: TDaCBB3T94CtMPRHN9pP8Uvdy8CTtWE5SV

Support MoH: <https://d4ttara.github.io/metacademy-of-humanity/support/>

Добровільна підтримка не створює прав власності, інвестиційних, членських або управлінських прав.