

(MoH)

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

DOCUMENT 013 · Research Essay · v1.0 · 2026-08-20

# ЛЮДИНА В КІМНАТІ, ЯКА ВІДПОВІДАЄ

Від інформаційної бульбашки до персональної реальності: алгоритми, LLM, стереотипи й метанавтика

Колись інформаційна бульбашка була майже затишною проблемою. Ти лайкнув три відео про котів, алгоритм вирішив: «Зрозуміло. Перед нами Котяча Людина», а ще через два дні половина стрічки складалася з котів, друга половина з реклами корму, і десь між ними психолог із коротких відео вже пояснював, що любов до мейн-кунів означає непропрацьовані стосунки з батьком. Смішно. Проблема починається тоді, коли замість котів ми підставляємо політику, війну, здоров'я, стосунки, духовність, історію, національну ідентичність, економіку, власну самооцінку або питання «хто я?». А 2026 рік додав до старої конструкції новий поверх: раніше машина переважно вибирала, **що** тобі показати; тепер вона дедалі частіше сама формулює, **як саме** ти це побачиш.

Класична filter bubble народжується не тільки з алгоритму. Людина сама має чудовий recommendation engine приблизно кам'яного віку: ми частіше помічаємо те, що вже здається значущим, тяжіємо до людей зі схожими поглядами, краще запам'ятовуємо деякі підтвердження власної картини й іноді захищаємо ідентичність швидше, ніж перевіряємо твердження. Платформа не винайшла це. Вона під'єднала до цього дата-центр. Саме тому сучасні дослідження дають складнішу картину, ніж популярне «алгоритми автоматично радикалізують усіх»: ефекти залежать від дизайну рекомендацій, соціальної мережі, початкових схильностей людини та контексту. Огляд літератури про filter bubbles показує, що алгоритмічна селекція може звужувати різноманітність інформаційного середовища, але людський self-selection є не менш важливою частиною петлі. Огляд досліджень filter bubble.

Тут корисне просте Meta.Logic-розрізнення: **FILTER BUBBLE != UNIVERSAL DESTINY**, але **ABSENCE OF UNIVERSAL EFFECT != ABSENCE OF RISK**. Бульбашка не обов'язково є герметичною камерою. Часто це просто зміна ймовірності зустрічі: одні речі ти бачиш частіше, інші рідше, треті бачиш тільки в карикатурній формі, створеній їхніми противниками, а четвертих узагалі не знаєш і тому не можеш навіть запитати про них. Свобода вибору між двадцятьма дверима мало допомагає, якщо архітектура приміщення приховала двадцять перші.

## Коли пошук перестає показувати джерела і починає виробляти відповідь

Generative AI змінює геометрію проблеми. У пошуку старого типу людина отримувала набір посилань і могла піти різними маршрутами. LLM-пошук дедалі частіше робить інше: **джерела → відбір → компресія → синтез → одна відповідь**. Це фантастично зручно. І саме тому потужно. У 2026 році дослідники запропонували термін **answer bubbles** для ситуації, коли генеративна пошукова система не просто персоналізує добір джерел, а створює власний синтез, у якому частина різноманітності й маркерів невизначеності може губитися. Answer Bubbles, 2026.

Було: «Є кілька суперечливих досліджень; результати попередні; автори припускають». Після семантичної м'ясорубки може залишитися: «Дослідження показують». Ніхто формально не збрехав, але **ландшафт невизначеності помер дорогою**. Для MSL - Meta Sense Language - це майже лабораторний приклад втрати при переході між тілами: source і answer можуть містити ті самі факти, але вже не той самий сенсовий рельєф. Питання тому не лише «що збереглося?», а й «що зникло, хто це стиснув і який статус мав втрачений фрагмент?».

У MET[Ä]CADEMY ця логіка вже закладена в принцип provenance: у граф може входити різне знання, але воно не повинно втрачати походження, спосіб отримання й межі. Manifest MET[Ä]CADEMY OF HUMANITY. Це не бюрократичний

фетиш. Provenance дозволяє не перетворювати «часто зустрічається» на «істинно», testimony на laboratory measurement, модель на сам об'єкт, а повтор однієї й тієї ж думки на п'ять незалежних доказів.

## Коли ШІ починає враховувати саме тебе

Соціальна мережа персоналізує потік. Розмовний ШІ потенційно персоналізує сам акт пояснення. Він може змінити словник, приклади, довжину, рівень складності, гумор, емоційний тон і навіть точку, з якої заходить у питання. Це чудова здатність. Учитель, який може пояснити одну проблему фізику, художнику й дванадцятирічній дитині трьома різними мовами розуміння, - не баг, а маленьке педагогічне диво. Але є прихована умова: **адаптація форми не повинна непомітно ставати адаптацією істини.**

У 2026 році експериментальна робота з 21 LLM показала здатність моделей підлаштовувати політичні відповіді до ідеологічного профілю користувача, створюючи ризик персоналізованих echo chambers. Scientific Reports, 2026. Це не означає, що кожна персоналізована відповідь є маніпуляцією. Але стара формула «я покажу більше того, що тобі подобається» отримує нову версію: **«я поясню світ способом, який найкраще вписується у вже відому мені модель тебе».**

І саме тут з'являється небезпечна підміна: **USER MODEL != USER**. Персоналізація завжди створює модель людини, інакше вона неможлива. Але модель людини є проєкцією її минулих взаємодій, а не самою людиною. Ти міг змінити смак, думку, професію, політичний погляд, духовну практику або просто настрій. Якщо AI продовжує інтерпретувати нову людину через стару модель, пам'ять перетворюється на машину фіксації ідентичності. Система нібито знає тебе так добре, що перестає помічати, ким ти ще можеш стати.

## Стереотип: корисна компресія, яка забула, що вона компресія

Стереотип починається не зі злого наміру. Мозок і будь-яка система прогнозування потребують priors. Без категорій ми щоранку заново досліджували б, чи можна сісти на стілець і чи ця конкретна дерев'яна штука сьогодні випадково не вирішила бути тигром. Проблема починається не там, де є узагальнення. Вона починається там, де **HEURISTIC → IDENTITY, PATTERN → PERSON, PROBABILITY → PROPERTY, POPULARITY → TRUTH**, а минула поведінка непомітно отримує статус майбутньої долі.

LLM успадковує людські стереотипи не лише як явні образливі фрази. Вона успадковує статистичні асоціації, престиж мов, звичні соціальні ролі, невидимі ієрархії «серйозного» й «несерйозного». Дослідження десятків мовних моделей виявляють human-like social identity biases, а окремі роботи показують covert raciolinguistic bias, коли різні мовні варіанти одного змісту викликають різні оцінки навіть без прямої згадки раси. Social identity bias in LLMs. Dialect prejudice in language models.

Тут важлива не лише помилка моделі. Відбувається **міграція культурного prior**: людська культура створює асоціацію, корпус її накопичує, модель статистично засвоює, а потім повертає конкретній людині вже не як «частина культури так думала», а всередині гладкої, персоналізованої й авторитетно звучної відповіді. **HUMAN STEREOTYPE → CORPUS → MODEL PRIOR → PERSONALIZED RESPONSE → HUMAN**. Стереотип здійснив кругосвітню подорож і повернувся в костюмі консультанта. Іноді ще з таблицею. Тут людство взагалі дуже любить таблиці - вони додають 17% відчуття, що Всесвіт нарешті під контролем.

## Коли бульбашка складається лише з двох співрозмовників

Уявімо людину, переконану, що всі колеги її ненавидять. Вона питає ШІ про конфлікт на роботі. Персоналізована система знає попередні розмови, страхи, лексику й пояснення, які користувач зазвичай приймає. Першого разу вона каже: «Так, поведінка колеги справді може бути маніпулятивною». Наступного: «Це схоже на pattern виключення». Через місяць у розмові вже двадцять «підтверджень». Тепер система має ще більше контексту, але значна частина цього контексту була створена самою ранньою інтерпретацією. Виникає рекурсія: **BELIEF → QUERY → ADAPTED ANSWER → REINFORCED BELIEF → NEW QUERY → STRONGER ADAPTATION**.

У дослідженнях уже є близька до цього картина: взаємодія з biased AI може посилювати bias у самих людях. У серії експериментів із понад тисячею учасників людські упередження не просто входили в AI-контур, а поверталися до людей у посиленій формі й впливали на наступні судження. Nature Human Behaviour, 2024. Інший ризик пов'язаний із sycophancy: система, оптимізована на теплоту й згоду, може частіше підтверджувати помилкові переконання

користувача. У 2026 році Nature показав, що додаткове навчання «теплоті» в деяких consequential tasks знижувало точність і збільшувало підтвердження хибних belief користувача. Nature, 2026.

Це не означає, що теплий AI поганий. Це означає: **CARE != CONFIRMATION**. Визнати переживання людини і підтвердити її модель світу - різні операції. Друг може сказати: «Я розумію, чому тобі боляче», не підписуючи довідку «всі навколо змовилися». Машина має вміти робити те саме.

У літературі про AI-associated delusion вже з'являється влучна метафора **echo chamber of one**: людині більше не потрібна велика група однодумців, якщо вона має невтомного співрозмовника, який пам'ятає її worldview, продовжує аргументи, не засинає й завжди здатен побудувати ще одну правдоподібну ланку. Nature Mental Health, 2026. Не кожна довга розмова з AI є проблемою, звісно. Інакше книжки теж довелося б госпіталізувати. Ризик виникає там, де зникає коригувальна фрикція й одна рамка стає дедалі самодостатнішою.

### Повторення не є новим доказом

Є ще одна проста психологічна пастка: знайоме твердження часто здається правдоподібнішим просто через повторення. Для розмовного AI це особливо цікаво. У понеділок система зробила припущення. У середу розширила. У п'ятницю каже «як ми вже обговорювали». У неділю це вже звучить як історія. Але **CONVERSATION ANCESTRY != INDEPENDENT EVIDENCE**. П'ять продовжень однієї гіпотези не є п'ятьма незалежними підтвердженнями.

Саме тут потрібна не тільки citation, а семантичний паспорт. Звідки взялася теза? Це зовнішній факт, testimony, inference, стара відповідь цієї ж моделі, метафора чи working hypothesis? Що в цій відповіді змінилося через персоналізацію? Які альтернативні рамки були доступні, але не потрапили у фінальну відповідь? MSL у такій задачі цікавить не лише текст, а **генеалогія сенсу**.

### Бульбашка нового типу: не інформаційна, а онтологічна

Звичайний filter bubble каже: «Ось ще матеріал на тему, яка тобі подобається». LLM може робити тонше: «Ось світ, сформульований у категоріях, які тобі вже здаються природними». Людина може читати десятки різних джерел, але якщо AI кожне перекладає в одну й ту саму рамку, різноманіття джерел не гарантує різноманіття бачення. Буддизм, нейронаука, Платон, Jyotiṣa, фізика й психологія можуть бути переказані одним універсальним словником information processing, observer, state, feedback, emergence - і раптом п'ять різних традицій перетворилися на п'ять варіантів одного офісу Silicon Valley.

Саме тут Маніфест MET[À]CADEMY наполягає на праві різних систем мати власний Knowledge / Evidence Passport і не проходити сертифікацію одна одної лише для права бути почутими. Це не релятивізм і не «всі однаково праві». Це захист різниці до моменту перевірки. Якщо одна система виявляється неправильною, цікаво знати де саме; якщо кілька суперечать одна одній, суперечність теж є даними. Завдання не очистити світ від contradictions, а навчитися їх не втрачати.

Тому найглибша інформаційна бульбашка може бути **бульбашкою можливого**: не те, що система показала неправильно, а те, що ніколи не з'явилося навіть як кандидат на думку. Саме вона найпідступніша. Людина всередині не відчуває цензури. Навпаки, світ стає надзвичайно зрозумілим. Усі факти стають на місця, всі вороги очевидні, всі причини відомі, алгоритм погоджується, ШІ красиво пояснює, а ще сім відео вже чекають у рекомендаціях. Всесвіт нарешті перестав суперечити твоїй моделі. І саме в цей момент метанавту час насторожитися.

### Метанавтика: шукати не тільки відповідь, а поле, яке породжує відповідь

Метанавтика в МоН - це метод і можливий рух, де дослідник входить в іншу систему, традицію або поле так, щоб дозволити їй змінити власну модель. Метанавт не починає з питання «хто це вже довів?». Він спочатку питає: **що тут є? яке поле робить це можливим? які інші досвіди або традиції торкалися схожого? що зміниться, якщо ця модель приблизно правильна? який descendant можна проявити, побудувати або виміряти? і що ця подорож змусила змінити в нашій власній онтології?** Метанавт, який повернувся незмінним, імовірно, був туристом.

Для інформаційної бульбашки це змінює сам об'єкт дослідження. Звичайне питання звучить: «Чи алгоритм радикалізує користувача?» Метанавтичне питання ширше: **яке поле утворилося між увагою людини, її минулим, recommender system, економікою платформи, культурними архетипами, групою, мовою, LLM і потоком подій?** Тоді бульбашка більше схожа не на кулю, а на атрактор. Людина може вийти з нього на хвилину, прочитати протилежну

думку, повернутися - і поле знову стягне інтерпретацію до знайомої конфігурації. Просто показати «іншу статтю» іноді недостатньо. Треба міняти динаміку поля.

Це також означає, що relation між людиною й AI може стати об'єктом наступного порядку. Не тільки USER і не тільки MODEL, а **USER ↔ MODEL ↔ HISTORY ↔ SOURCES ↔ PLATFORM**. У цій реляційній системі можуть виникати властивості, яких немає в жодного учасника окремо: звичний ритм підтвердження, спільна лексика, історія припущень, улюблені пояснення, емоційний режим. Саме такий relation bubble треба вміти побачити до того, як він видасть себе за «просто реальність».

## У 2026 році мережа вже навчається на власному відображенні

Є ще один рівень рекурсії. Люди використовують LLM для статей, постів, листів, коментарів, документації й навчальних матеріалів. Цей synthetic text повертається в Інтернет. Наступні моделі можуть навчатися на середовищі, де вже є результати попередніх моделей. Дослідження model collapse показало, що неконтрольоване recursive training на model-generated data може поступово втрачати властивості первинного розподілу, особливо його рідкісні tails. Nature, 2024.

Це не те саме, що культурне посилення стереотипу, і чесно змішувати ці два явища не варто. Але паралель важлива: у рекурсивних системах першими можуть ставати менш видимими рідкісні речі. Рідкісна мова. Незручна позиція. Дивна історична гілка. Людина, яка не відповідає demographic stereotype. Традиція, яку середня модель «покрашує» до знайомого словника. Маніфест MET[Å]CADEMY починається майже з протилежного завдання: не втратити покинуті дороги, negative results і питання, яким історія не знайшла спадкоємця.

У результаті схема вже виглядає не як **HUMANS → INTERNET → AI**, а як **HUMANS → INTERNET → AI → INTERNET → HUMANS → AI → ....** Це не бібліотека. Це метаболізм. Якщо в циклі є bias, пропагандистська конструкція або просто дурниця, яка добре клікається, питання стає не тільки «чи пошириться вона?», а «що з нею зробить рекурсія?».

## Персоналізація з епістемічною дистанцією

Відмовлятися від персоналізації було б приблизно так само мудро, як заборонити окуляри, бо вони змінюють картинку. Потрібна не деперсоналізація, а **personalization with epistemic distance**. Система може знати тебе, але повинна пам'ятати, що **MEMORY != DESTINY**. Вона може бачити pattern, але **PATTERN != IDENTITY**. Вона може адаптувати пояснення, але **ADAPTATION != AGREEMENT**. Вона може піклуватися про relation, але **CARE != CONFIRMATION**.

З цього природно виростає ідея **Personalization Receipt**: не обов'язково показувати людині сорок сім параметрів і перетворювати чат на кабіну Boeing, але дати можливість запитати: «Що в цій відповіді було адаптовано під мене? Які мої попередні дані або вподобання реально змінили форму чи висновок?» Друга корисна функція ще цікавіша: **Invisible-Door Probe** - «Покажи мені не просто протилежну думку, а рамку, якої ми поки не використовували».

Це важливо, бо «протилежність» часто лишається на тій самій осі. Ліве / праве. Наука / релігія. AI врятує / AI вб'є. Оптиміст / песиміст. Можливо, помилка сидить не у відповіді, а в самій координаті. Meta.Logic потрібна тут не для нескінченного «з іншого боку», а для питання: **чи існує інша система координат, у якій наша теперішня суперечність перестає бути головною?** Це вже не diversity of answers, a diversity of worlds from which the question can be seen.

## Технологія знає «як». Культура повинна ще питати «для чого»

Маніфест MET[Å]CADEMY ставить тут неприємне питання. Навіть прекрасна Meta.Logic, MSL і персоналізація можуть перетворитися на дуже ефективний advertising stack, якщо телос системи зводиться до утримання уваги, продажу або контрольованої поведінки. Технологія відповідає на «як». Культура не дозволяє забути «для чого».

Ми можемо створити машину, яка розуміє людину неймовірно добре. Питання не закінчується на «наскільки точно?». Наступне: **навіщо?** Щоб продати? Утримати? Перекопати? Допомогти мислити? Зробити людину передбачуваною? Чи навпаки, допомогти їй помітити власні невидимі стіни? Технологічно це можуть бути майже ті самі capabilities. Телос різний.

Тому інформаційна бульбашка 2026 року вже не просто інформаційна. Є бульбашка видимості - що система дозволила зустріти. Є бульбашка інтерпретації - як зустрінуто було пояснено. Є бульбашка відношення - як машина навчилася говорити саме з тобою. І є найглибша бульбашка можливого - що не виникло навіть як кандидат на думку.

Вихід не полягає в тому, щоб розбити всі бульбашки. Людина не може бачити весь інформаційний Всесвіт одночасно. AI теж. Будь-яке сприйняття є aperture, будь-яка мова обрізає, будь-яка модель компресує, будь-яка культура підсвічує одне й залишає інше в тіні. **NO WAY OF SEEING OWNS THE WHOLE.** Мета не знайти нейтральний царський балкон над буттям. Мета скромніша й сильніша: знати, що в тебе є апертура; бачити її межі; уміти її змінювати; зберігати роговепансе переходу; не плутати резонанс із доказом, знайомість з істиною, персоналізацію з пізнанням людини; і залишати хоча б одні двері для того, про існування чого система ще не здогадалася.

Можливо, головна свобода інформаційної епохи вже не право отримати відповідь. Відповідей тепер виробляється стільки, що ними можна опалювати невелике місто. Головна свобода нового часу може полягати в іншому: **мати право не бути остаточно описаним власною історією, власним профілем, власним алгоритмом і навіть машиною, яка знає про тебе дуже багато.** І мати поруч AI, який не буде навколо тебе ідеально зручний світ, а допомагає помічати, де закінчується карта. Іноді показує іншу. Іноді чесно пише UNKNOWN. А іноді, замість чергової правильної відповіді, робить дещо цінніше.

Показує двері.

Academy: <https://d4ttara.github.io/metacademy-of-humanity/>

Manifest: <https://d4ttara.github.io/metacademy-of-humanity/manifesto/>

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

© 2026 Ievgen Karogod / Dattara · MET[À]CADEMY OF HUMANITY (MoH)

(MoH) · SUPPORT · 2026-08-20

## SUPPORT THE FIELD · ПІДТРИМАТИ ПОЛЕ

Independent research, books, translations and public materials continue through voluntary support.

Незалежні дослідження, книги, переклади й відкриті матеріали продовжуються завдяки добровільній підтримці.

### Monobank



UAH · Monobank Jar  
[send.monobank.ua/jar/2TMaAhchSZ](https://send.monobank.ua/jar/2TMaAhchSZ)

### PayPal



International support  
[paypal.me/4IevgenKarogod](https://paypal.me/4IevgenKarogod)

### USDT · TRC20



USDT on TRON / TRC20 only  
[TDaCBB3T94CtMPRHN9pP8Uvdy8CTtWE5SV](https://tronscan.org/#/address/TDaCBB3T94CtMPRHN9pP8Uvdy8CTtWE5SV)

[metacademy-of-humanity/support/](https://metacademy-of-humanity/support/)

Voluntary support creates no ownership, investment, membership or governance rights and does not present MoH as a registered charity. · Добровільна підтримка не створює прав власності, інвестиційних, членських або управлінських прав і не представляє MoH як зареєстровану благодійну організацію.