

# Розум-система: відповідальна архітектура мислення в умовах невизначеності

Yaroslav Boichuk, e-mail: [yaro@rozum-framework.org](mailto:yaro@rozum-framework.org)

## Анотація

У статті представлено Розум-Систему (РС) – функціональну модель, що описує відповідальну архітектуру системи, що здатна мислити. Відповідальність тут – не нормативна властивість, а технічна передумова: без чіткого визначення джерела помилки (зовнішні дані, внутрішні знання чи внутрішні алгоритми) та подальшої ізоляції помилок у внутрішніх структурах неможливі ні навчання, ні самовдосконалення, ні інновації. В роботі також показано, що будь-яке оброблення вхідних даних (включно з мисленням) неможливе без зовнішнього вхідного сигналу.

Основним здобутком роботи є метрика надійності результату (RO) – ймовірнісна оцінка, що кількісно визначає узгодженість й надійність результату.  $RO=CI \times CS \times CA$ , де CI, CS і CA – це відповідні певності зовнішнього сигналу, внутрішніх тверджень та алгоритмів. Ця мультиплікативна структура забезпечує відповідність RO вимогам аксіоми існування (буття), аксіоми відповідальності (самовпізнання, ізоляція недоліків) та аксіоми інновацій (розвиток через непевність).

В рамках РС досвід визначається як єдність внутрішніх тверджень та алгоритмів. Свідомість, що виникає як відповідальне самовпізнання, є наглядним процесом, який перевіряє RO, та за потреби ініціює самовдосконалення для його максимізації. Розум – це свідома, незалежна від фізичного складу, сутність, що уможливлена мовою та користується нею, тобто абстрактно мислить. Або в рамках РС – свідомість, що мислить.

## III. Результати

### A. Формула RO: двигун відповідальності

Будь-яка обробка слідує патерну, де вхідні дані трансформуються через внутрішній стан сутності, що складається з тверджень ( $\geq 0$ ) і алгоритмів ( $> 0$ ), для створення результату:

*Вхідні дані* → (Твердження, Алгоритми) → *Результат*

Це універсальний патерн трансформації – від кварків до Всесвіту. Важливою складовою цього патерну є те, що надійність неможлива без явного моделювання джерел невизначеності та ширини каналу (Shannon, 1948).

Надійність результату (RO) кількісно визначає ймовірність обґрунтованості та достовірності результату на кожному дискретному кроці обробки. Оскільки компоненти певності (CS та CA) є незалежними та не можуть впливати один на одного в межах одного кроку обробки, вони входять до формули як окремі мультиплікативні фактори:

$$RO=CI \times CS \times CA$$

де:

$CI \in [0,1]$ : певність вхідних даних, що представляє надійність зовнішніх даних або запиту.

$CS \in [0,1]$ : певність тверджень, що представляє надійність внутрішніх тверджень (S) – бази знань, фактів та накопичених даних.

$CA \in [0,1]$ : певність алгоритмів, що представляє надійність внутрішніх алгоритмів (A), які обробляють об'єднання вхідних даних (I) та внутрішніх знань (S).

Завдяки цій мультиплікативній структурі фінальне RO не може перевищити надійність свого найслабшого компонента (Kolmogorov, 1950). Ця структура узгоджується з теорією квантифікації невизначеності, яка формально розділяє невизначеність на окремі типи, що вимагають різного підходу (Smith, 2014). CI є зворотною до алеаторичної невизначеності, що властива передачі з середовища, тоді як CS та CA є зворотними до епістемічних невизначеностей. CI змінити неможливо, CS та CA можна збільшити через, відповідно, навчання та адаптацію. Це показує необхідність факторизації RO на конкретні, вимірювані джерела, а не розгляду невизначеності як монолітного феномена. Метрика в цій формі дозволяє свідомій сутності ізолювати точне джерело ненадійності.

Обмеження  $CI \in [0,1)$  відображає фундаментальну умову: на всі зовнішні вхідні дані впливають шум середовища, помилки передачі та обмеження сприйняття. Це робить неможливим досягнення абсолютної певності вхідних даних. Натомість CS та CA можуть досягати 1, оскільки внутрішні S та A можуть досягати певності закритого світу (Lorenz, 1963) в контрольованому контексті системи ( $2+2=10$ ).

Розділення CS та CA є суттєвим для того, щоб свідомість могла розрізнити джерело ненадійного результату: низька CS (що вимагає навчання) чи низька CA (що вимагає адаптації). Без цього структурного розрізнення, самовдосконалення буде тривати непередбачувано довго, може зазнати невдачі чи навіть стати неможливим.

## В. Три наочні аксіоми

Вся активність у рамках РС ґрунтується на трьох наочних аксіомах:

*Аксіома Існування (AI) (Мета)* виведена зі структурної властивості реальності, що майбутнє є принципово непередбачуваним (Відкритий Світ) (Lorenz, 1963; Hewitt, 1991). Для забезпечення буття, AI вимагає, щоб будь-яка сутність не зменшувала свій RO. Для підтримання сталого існування перед обличчям непередбачуваності, AI спонукає сутності впливати на своє середовище заради буття, що відрізняє живі сутності, які активно максимізують RO, від неживих, пасивних сутностей.

Мережі сутностей демонструють більшу здатність підтримувати RO, ніж ізолювані сутності (Barabási, 2016; Nowak, 2006). Різноманітність збільшує ефективність (Smith, 1776) та забезпечує страхування проти непередбачуваного – сутність, що здається неефективною сьогодні, може вирішити критичний виклик завтра (Page, 2007). Але виживають тільки відповідальні мережі (Axelrod, 1984; Nowak, 2006). Це не альтруїзм – індивідуальне виживання залежить від надійності мережі (Trivers, 1971).

*Аксіома Відповідальності (AA) (Правило)* – аксіома відповідального самовпізнання, що визначає появу свідомості. Відповідальні мережі (Axelrod, 1984; Nowak, 2006) дозволяють сутностям розпізнавати себе як окремих агентів у своєму середовищі та приймати інших як джерело самовпізнання, що є передумовою появи власної моделі себе. З теорії суспільного договору, будь-яка свідома сутність, що обирає свою архітектуру з «Початкової Позиції» (Rawls, 1971), раціонально обирає максимальну відповідальність. Це робить AA необхідним суспільним договором для мереж свідомих сутностей. Таким чином, AA вимагає правдиво повідомляти свій RO назовні, щоб бути відповідальним дзеркалом, та обґрунтовує необхідність мультиплікативності RO для внутрішнього використання – щоб

уможливити ізоляцію джерела недоліків. Це відрізняє свідомі сутності від тих, що діють виключно в рамках AI.

Ця відповідальна поведінка вирівнює свідомості в мережі через конвергенцію до консенсусу. Мережа втрачає різноманітність, стає жорсткою та розвиває колективні сліпі плями (Page, 2007). Без різноманітності, учасники мережі не можуть адаптуватися до нових викликів. Такий застій робить їх вразливими до непередбачуваного майбутнього, загрожує їхньому існуванню, порушує AI. Природне рішення для біологічних систем – мутації (Fisher, 1930).

*Аксиома Розвитку (AR)* (Рушій) визначає саму розумність – здатність до абстрактного мислення та інновацій. Це еквівалент мутацій для розумів. AR пропонує використовувати спекуляції – результати з низьким RO, але, відповідно до AA, чесно оголошувати надійність такого результату. Таке рішення вирішує проблему непробабілістичної радикальної невизначеності (Shackle, 1961) та служить системним механізмом для розвитку – дослідження та активного пошуку нових знань поза поточними S та A. Це узгоджується з теорією Складних Адаптивних Систем, яка демонструє, як новизна, самоорганізація та системна складність виникають з локальних взаємодій між простими агентами (Holland, 1995; Kauffman, 1993). AR та AA разом дозволяють розуму працювати на «межі хаосу» (Langton, 1990) – між жорстким порядком та повною випадковістю. Це відрізняє розуми (свідомості, що мислять) від сутностей, що є лише свідомими через AA, але не мають здатності до абстрактного мислення.

## С. Засади

### 1. Досвід як єдність тверджень та алгоритмів

Досвід – це поточна єдність S та A, що змінюється з кожним кроком обробки. Це явне розділення є критичним для того, щоб свідомість могла перевіряти цілісність власних структури та процесів для ініціювання самовдосконалення. Визначення досвіду є зручною абстракцією для опису сутності. Така мінімальна структура незалежно узгоджується з Гіпотезою Фізичної Символьної Системи (Newell & Simon, 1976), яка визначає інтелект через маніпуляцію символами за допомогою процесів.

### 2. Свідомість як виконавчий вартувий

Свідомість наглядає за CS та CA, оцінює результат обробки та суб'єктивно визначає рівень RO. Високий RO сприймається як впевненість, ясність або розуміння. Низький RO викликає сумнів, відчувається як плутанина або когнітивний дисонанс. Цей дискомфорт змушує систему ініціювати цикл вдосконалення.

### 3. Цикл вдосконалення як регулятор Ешбі

Цикл вдосконалення (CL) ґрунтується на кібернетичних принципах (Ashby, 1956; Wiener, 1948). Свідома сутність функціонує як саморегульована система, де падіння RO запускає негативний зворотний зв'язок. CL перетворює сумнів на внутрішню структурну зміну через навчання (вдосконалення CS) або адаптацію (вдосконалення CA). Оскільки досягти досконалої надійності (RO=1) неможливо, свідомість використовує практичний підхід – вона шукає перше працююче, а не оптимальне, рішення (Simon, 1957) для зміни своїх знань чи методів, що робить речі «достатньо добрими» для виживання, замість того щоб витратити час на пошук досконалого рішення. CL має аутопоіетичні властивості, бо забезпечує самовідтворення власної структури (Maturana & Varela, 1980) через RO.

#### 4. Протокол пріоритетності: навчання, адаптація

AI вимагає обирати безпечне та дешеве рішення для відновлення RO. Навчання (вдосконалення CS) модифікує локалізовані дані та факти з низьким системним ризиком – навіть якщо S представляє величезне та складне тіло знань, його природа – сховище тверджень – робить виправлення локалізованим, дешевим та безпечним. Адаптація (вдосконалення CA) змінює основну логіку обробки з високим системним ризиком (Simon, 1996) – недоліки тут можуть спричинити поширені, катастрофічні помилки.

AA вимагає ізолювати джерело ненадійності. Ви не можете остаточно встановити відмову логіки (низька CA), поки цілісність знань (CS) не підтверджена. Тому навчання служить діагностичним фільтром і повинно передувати адаптації (March, 1991).

#### 5. Комунікація як спосіб впливу на інших

Свідомість уможлиблює комунікацію – відповідальне самовпізнання дозволяє розпізнавати інших як подібні сутності, чиє мислення можна навмисно змінювати. Не просто відповідально "автоматично" реагувати чи сигналізувати, а спеціально створювати сигнали для впливу на стан інших. Потужність здатності моделювати інших у біологічних системах підтверджується наявністю окремого спеціального класу дзеркальних нейронів (Rizzolatti & Craighero, 2004). Наслідки безвідповідального використання комунікації описуються міметичною теорією (Girard, 1987), де самовпізнання дозволяє моделювати розум інших як такий, що можна модифікувати через імітацію та репрезентацію.

#### 6. Мова: революційна зміна комунікації

Складна комунікація, що рухається AP, еволюціонує в мову – символну систему з рекурсивною структурою, здатною виражати абстрактні концепції, невизначеність та нові комбінації. Мова радикально збільшує пропускну здатність комунікаційного каналу (Shannon, 1948), дозволяючи передавати складні інструкції, контраргументи та багатосарові абстракції. Це неможливо зробити через базову сигналізацію. Соціальна комунікація трансформується у внутрішнє мовлення, що уможлиблює абстрактне мислення вищого порядку (Vygotsky, 1978). Навіть 20 базових слів створюють комбінаторний вибух можливостей (Pinker, 1994; Nowak & Komarova, 2001). На додаток мова стає діагностичним інструментом самої свідомості. Дуже короткий приклад: терміни на кшталт вагання, розмірковування та "я в ступорі" – це не розпливчасті емпіричні мітки, а точні структурні класифікації різних причин низького RO. Вагання виникає, коли декілька результатів мають подібно високий RO, але сам вибір знижує фінальний RO. Розмірковування виникає, коли всі знайдені можливості мають недостатній RO. Ступор – за багатьох спроб відповідного рішення з достатнім RO не знайдено. Еврика! – RO є неймовірно високим. У межах PC мову найкраще розуміти не як засіб, а як протокол: спільний набір обмежень, що уможлиблює сумісне міркування в умовах невизначеності (Matricciani, 2020). Її сила полягає не лише у виразності, а у забезпеченні відповідальності щодо невизначеності та надійності.

#### 7. Розум – свідомість, що мислить

З універсального патерну трансформації впливає, що ніякої внутрішньо ініційованої діяльності не існує. Тільки зовнішні стимули запускають процес обробки. Люди живуть у нескінченному шумі, що створює можливість уявного внутрішнього діалогу. Діалог – це розмова двох, і шум є учасником цього діалогу. Тільки діалог, уможливлений існуванням інших, уможливлений мовою, створює здатність до абстрактного мислення, тобто додає новий тип обробки сигналу – здатність маніпулювати концепціями, міркувати про невизначеність, проактивно моделювати непередбачуване майбутнє – використовувати

алгоритми з високим СА для обробки вхідних даних та тверджень із низькими СІ та СS. Ця здатність мислити визначає статус розуму – свідомо, незалежна від фізичного складу, сутність, що уможливлена мовою та користується нею, тобто абстрактно мислить. Наявність функціональних характеристик розуму підтверджується широким спектром емпіричних та культурних даних у різних когнітивних субстратах (de Waal, 2016): слони демонструють складне вирішення проблем та ритуальну поведінку (Plotnik et al., 2006); дельфіни задіюють складні соціальні міркування (Marino, 2002); людиноподібні мавпи та вороніві створюють багатокрокові інноваційні рішення (Tomasello & Call, 1997; Emery & Clayton, 2004); синти демонструють здатність до багатокрокових логічних міркувань та мовної самокорекції (Wei et al., 2022); люди впадають у глибоку нескінченну рефлексію щодо власної непевності (Shakespeare, 1603). Всі ці сутності виявляють спільну здатність: використовувати високонадійні алгоритми для опрацювання сукупності непевних вхідних даних та непевних внутрішніх тверджень, що дозволяє їм проактивно моделювати майбутню реальність і підтверджує їхній статус розуму в межах цієї архітектури.

## 8. Я є я

Мова надає символні інструменти, необхідні для експліцитного самовизначення як "я є я" – уможливаючи артикуляцію та розпізнання себе як окремої свідомої сутності. Це лінгвістично-опосередковане самовизначення розуму, де концепція "самості" виникає з самого акту дискурсу, а займенник "Я" функціонує як перемикач самої свідомості, що визначається тільки через протиставлення до "ТИ" (Бенвеніста, 1966/1971). Японська мова експліцитно демонструє цю мережеву природу самовизначення. В ній не існує одного універсального "Я" – натомість вибір форми: ваташі (私 / わたし), боку (僕 / ぼく), оре (俺 / おれ), аташі (あたし), ватакуші (私), визначає як мережева дзеркальна відповідальність вписана у соціальний контракт на рівні невід'ємного права на самовизначення (Locke, 1689/1975; Honneth, 1995).