

**ШТУЧНИЙ ІНТЕЛЕКТ ЯК ЯКІСНО НОВА ФОРМА
ТЕХНОГЕННОСТІ: ОНТОЛОГІЧНИЙ
ТА ПРАКСЕОЛОГІЧНИЙ АСПЕКТИ**

Черкасов А.Д.

e-mail: andrii.cherkasov2@nure.ua

Науковий керівник: к. філос. н., доц., Покровський А.М.

Харківський національний університет радіоелектроніки, каф. СГН
м. Харків, Україна

Abstract. The paper examines the «black box» problem in neural networks as a specific form of technogenicity. Unlike classical tools, neural networks introduce qualitative opacity, making standard forms of error detection, control, and accountability impossible. The paper distinguishes the ontological dimension of AI's architecture from the praxeological delegation of cognitive authority to it. The core problem lies in the anthropologization of AI: users intuitively project human-like rationality onto a system that operates on fundamentally different principles. This necessitates a philosophical re-evaluation of responsibility criteria in a technology-driven society.

Однією із актуальних проблем філософії техніки є проблема техногенності. Вона набуває особливого значення в умовах стрімкого розвитку сучасних інформаційних технологій, особливо штучного інтелекту. В загальному сенсі техногенність – властивість технічного рішення детермінувати людську діяльність поза волею суб'єкта, вбудовуючи онтологію засобу в онтологію самого діяча [1]. Ця властивість є новим виміром технічного детермінізму як здатності технічних засобів зумовлювати не тільки певні дії, але й форми соціального життя та пізнання [2]. Штучний інтелект в цьому контексті є не просто черговим проявом загальної тенденції ускладнення техніки, а її якісно новим виміром – передусім через те, що його архітектура розбудована за принципами, що прийнято визначати як «чорний ящик». Ця обставина має принципове значення для адекватного розуміння як сутності систем ШІ, так і їх техногенного впливу на людину і суспільство.

«Чорний ящик» у контексті нейронних мереж означає систему, де вхідні дані і вихідний результат доступні, але механізм, що перетворює одне на інше, принципово прихований – не лише від користувача, а й від розробника. Це не технічний недолік, що підлягає виправленню, а структурна властивість самої архітектури. Саме ця властивість і є проблемою: вона унеможлиблює звичні механізми контролю, виявлення помилок і встановлення відповідальності, а отже, вимагає переосмислення на філософському рівні.

Онтологічний аспект проблеми «чорного ящика» – це питання про те, чому непрозорість функціонування засобу є не випадковою, а необхідною,

зумовленою на рівні онтології. Техніка завжди впливала на людину: молоток чи верстат формували стереотипи дії, обмежували і спрямовували діяльність. Але класичний інструмент залишався «прозорим»: людина могла зрозуміти механізм його дії, за потреби звернутися до фахівця і вийти за межі власної некомпетентності. Непрозорість у такому разі була кількісною – питанням браку знань, який можна було здолати в принципі.

У випадку нейронних мереж непрозорість має принципово іншу, внутрішню природу. Мережа не виконує заздалегідь прописаних алгоритмів, а генерує результат через статистичну оптимізацію у багатовимірному просторі параметрів. Як зазначає J. Burrell [3], саме цей розподілений характер обчислень, а не брак наших знань, унеможлиблює формальне відновлення прямого причинно-наслідкового зв'язку. Додаткова відмінність полягає в тому, що класична техніка доповнювала переважно тілесну, «механічну» діяльність, тоді як ІІІ націлений на відтворення та заміщення когнітивних процесів – судження, оцінки та прийняття рішень [4]. Постає питання: чи може засіб, онтологія якого принципово відрізняється від тих процесів, які він покликаний відтворювати, адекватно їх замінити? Це запитання призводить до праксеологічного виміру проблеми.

Праксеологічний аспект проблеми «чорного ящика» – це питання про те, як людина взаємодіє з системою ІІІ, не знаючи і не маючи змоги дізнатись, за якими принципами вона насправді працює. Має місце загальна логіка функціонування і сприйняття будь-якої автоматизованої системи: ІІІ дає швидкий задовільний результат і знімає з людини тягар обдумування. Ефективність тут розуміється вузько – як економія зусиль, – а питання про те, що саме стоїть за результатом і які ризики це несе, відходить на другий план або не ставиться взагалі. Не розуміючи архітектурної природи ІІІ, людина мимовільно заповнює цей пробіл власними інтуїціями. Наприклад, функціонуванню системи може приписуватись послідовність і стабільність, що є обов'язковими атрибутами традиційних уявлень про класичні методи наукового аналізу проблем чи обґрунтування. Проте як раз цих властивостей у ІІІ нема за визначенням. На більш глибокому рівні відбувається пряме отождоження наслідків активності людського розуму та функціонування ІІІ – ніби за рішенням стоїть якийсь розум, а не статистична модель.

Делегування ІІІ когнітивних повноважень відбувається без урахування онтологічної специфіки його архітектури, і це породжує двоїсту проблему відповідальності. З одного боку, діє прагматична логіка економії сил: є суб'єктивне прагнення уникнути зайвих когнітивних зусиль та зняти з себе тягар обдумування. З іншого – є глибший механізм: посилення на рішення ІІІ виконує ту саму функцію, що й посилення на закон або формальну процедуру – воно знімає з людини персональну відповідальність, переносючи її на «формальну систему». Принципова відмінність полягає у

тому, що у випадку традиційної техніки або правових норм завжди є можливість відстежити зв'язок між рішенням системи і конкретними діями оператора чи законодавця. У випадку ШІ цей зв'язок розірваний на рівні онтології самого засобу: можна контролювати мотивацію користувача, але не процедури функціонування системи. Техногенність тут набуває нового виміру, бо навіть сама інтерпретація проблеми, яку вирішує користувач, ШІ визначає поза логічним аналізом проблеми [5]. Виникає рекурсія ШІ, який імпліцитно містить і запитання, і відповідь, і критерії оцінки – і все це поза традиційними процедурами контролю за логікою постановки та розгляду проблем. Можна заперечити, що роботу власного мозку ми теж оцінюємо переважно за результатами, не розуміючи до кінця принципи його роботи. Але людина, на відміну від «чорного ящика», усвідомлює і здатна контролювати процедури власного мислення і прийняття рішень – хоча б частково – і саме це надає підстави для відповідальності та корекції діяльності. У взаємодії з непрозорою системою цього мінімуму бракує: нема способу апелювати до рішення чи встановити відповідальність в епістемічному сенсі.

Таким чином, «чорний ящик» у нейронних мережах є не технічним недоліком, що підлягає усуненню, а структурною характеристикою, що визначає специфічний, якісно новий тип техногенності. На відміну від всіх попередніх ітерацій техніки, у випадку ШІ сам інструмент позбавлений прозорості внутрішньої логіки. Праксеологічно це означає, що суспільство делегує ШІ когнітивні повноваження, отримуючи результат, онтологічний статус якого не відповідає нашому розумінню процедур отримання знань та прийняття обґрунтованих рішень, що принципово загострює проблему критеріїв адекватності знання та відповідальності.

Список використаних джерел:

1. Покровський А.М. Від антропології та філософії техніки до онтології діяча та засобу // Вісник Харківського національного університету імені В.Н. Каразіна. Серія Теорія культури і філософія науки. 2012. № 1029. С. 38-43.
2. Feenberg A. *Transforming Technology: A Critical Theory Revisited*. Oxford: Oxford University Press, 2002. 230 p.
3. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3 (1). <https://doi.org/10.1177/2053951715622512>.
4. Carabantes M. Black-box artificial intelligence: an epistemological and critical analysis // *AI & Society*. 2020. Vol. 35. P. 309-317.
5. Warren von Eschenbach W.J. Transparency and the Black Box Problem: Why We Do Not Trust AI // *Philosophy & Technology*. 2021. Vol. 34. P. 1607-1622.