

<https://doi.org/10.15407/fd2026.01.037>

УДК 165.74:004.8:81'22

Олена АСТАПОВА-ВЯЗЬМІНА,

кандидатка філософських наук, доцентка,
Черкаський державний технологічний університет,
18006, Черкаси, бульв. Шевченка, 480
o.astapova@chdtu.edu.ua

<https://orcid.org/0000-0003-2104-9703>

WoS Research ID: PIJ-2547-2026

ТРИАНГУЛЯЦІЯ ДЕВІДСОНА ТА ШТУЧНИЙ ІНТЕЛЕКТ: СЕМІОТИКА ЦИФРОВОЇ СУБ'ЄКТНОСТІ

У статті досліджується трансформація штучного інтелекту (ШІ) з інструментального засобу в автономного посередника крізь призму епістемологічної концепції триангуляції Дональда Девідсона. Актуальність дослідження зумовлена необхідністю філософського осмислення великих мовних моделей (LLM) не лише як алгоритмічних систем, а і як потенційних суб'єктів комунікації, що потребує перегляду категорій пропозиційних установок — переконань, намірів та бажань. Наукова новизна полягає у зіставленні девідсонівського «трикутного екстерналізму» (Я—Інший—Світ) із сучасною проблематикою цифрової суб'єктності. Звернення до мисленнєвого експерименту про «Болотну людину» (Swampman) обґрунтовує тезу, що без історичної генези та причинно-наслідкового зв'язку зі спільним світом, ШІ ризикує залишитися «семантичним зомбі». На відміну від підходу В. Квайна, стратегія Девідсона передбачає, що об'єктивність знання виникає лише в процесі інтерсуб'єктивної взаємодії, а звідси, — розуміння істини уможливляється тільки через мову як форму соціальної триангуляції. У межах дослідження припускається можливість паралелі з художнім образом робота Клари з роману К. Ішигуро як пізньої моделі девідсонівської

Цитування: Астапова-Вязьміна, О. (2026). Триангуляція Девідсона та штучний інтелект: семіотика цифрової суб'єктності. *Філософська думка*, 1, 37—53. <https://doi.org/10.15407/fd2026.01.037>

© Видавець ВД «Академперіодика» НАН України, 2026. Стаття опублікована на заходах відкритого доступу за ліцензією CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

триангуляції, де етична суб'єктність виникає через спробу інтерпретації людського контексту. Особливу увагу приділено логічній структурі радикальної інтерпретації та адаптації теорії істини А. Тарського. Авторка доводить, що значення речень у ШІ-системах є «семантично непрозорим» без залучення до спільного символічного простору. У висновках підкреслюється, що перехід від мобільних до емпіричних обчислень вимагає від ШІ не просто точності, а здатності до «когнітивного використання тристороннього зв'язку». Встановлено, що приписування ШІ пропозиційних установок часто є наслідком людської антропоморфізації, тоді як справжня суб'єктність потребує історії взаємодії з людиною. Стаття пропонує розгляд триангуляції як фундаментальної моделі для створення людино-машинних систем, в яких етична відповідальність та змістова комунікація стають можливими лише всередині «замкненого трикутника» спільної об'єктивної реальності.

Ключові слова: триангуляція, Дональд Девідсон, штучний інтелект, цифрова суб'єктність, Болотна людина, соціальна реальність, радикальна інтерпретація, епістемічна ситуація.

Вступ

Виведення штучного інтелекту (ШІ) за межі інструментального використання і перетворення останнього на автономного посередника рішень людини потребує міждисциплінарного підходу, проблемно-орієнтованого як на оцінку великих мовних моделей (LLM), так і на розуміння етичних наслідків, інших викликів, які ставить перед людством людиною створений ШІ. Аналіз літератури останніх років переконує, що однією з основних проблем сучасної філософії ШІ є питання про те, чи можемо ми обґрунтовано приписувати штучно створеним системам пропозиційні установки — переконання, наміри, бажання тощо. У статті “Competing narratives in AI ethics: a defense of sociotechnical pragmatism” (2024) авторами захищається соціотехнічний прагматизм, наратив, що підкреслює центральну роль контексту та людської діяльності в проектуванні та оцінці нових технологій (Watson, Mokander, Floridi, 2024). Яскравий художній приклад переходу до ШІ як учасника комунікативного акту запропонував К. Ішигуро у славнозвісному романі «Клара і Сонце». Клара — це не просто робот-технологія, яку купили підлітку для розваги, вона зображується суб'єктом комунікації, намагається зрозуміти людський контекст життя, складно осмислює поняття «любов» та «серце», але здійснює спробу врятувати життя дівчинки, здійснюючи щось на кшталт молитви Сонцю. Припустимо, що приклад Клари є пізнім варіантом наочної моделі філософської концепції Дональда Девідсона та актуалізує ідеї американського філософа, демонструючи етичні та епістемологічні переплетіння концепції триангуляції.

Ідея триангуляції (triangulation), як відомо, формується Девідсоном в декількох роботах 1980-1990 років, так, наприклад, в роботі “Three Varieties of Knowledge” автор зауважує на формі триангуляції: «...кожен з

двох людей диференційно реагує на сенсорні стимули, що надходять з одного напрямку. Якщо подумки продовжити вхідні лінії назовні, їхньою спільною причиною буде точка їх перетину. Якщо згодом ці двоє починають помічати реакції один одного (у випадку мови — вербальні реакції), кожен з них може співвіднести спостережувані реакції іншого зі своїми власними стимулами зі світу. В такий спосіб визначається спільна причина. Трикутник, який надає зміст мисленню та мові, виявляється замкненим. Але для триангуляції необхідні двоє» (Davidson, 1987: р. 213). Девідсон наголошує, що знання про інший розум можливе, однак, лише за умови наявності знання про світ, оскільки триангуляція, яка є необхідною для мислення, вимагає, щоб учасники комунікації визнавали, що вони перебувають у спільному світі. М. Де Каро (M. De Caro), описуючи «семантичний екстерналізм» Девідсона, який сам Девідсон називає трикутним екстерналізмом (De Caro, 2011), зауважує, що філософ симпатизував фізичному екстерналізму, захищаючи який пропонує два основних аргументи, що ґрунтуються на мисленневих експериментах: інтерпретатор, який знає всі потенційні докази того, що думає і має на увазі інтерпретований мовець і Болотну людину (досконалий дублікат самого Девідсона, приблизно семидесятирічного віку, що миттєво створюється, коли сам Девідсон зникає) (De Caro, 2011).

Концепт триангуляції Девідсоном уточнюється, додаються певні конотації, так, наприклад, Девідсон деталізує: «Базова ситуація являє собою одночасне залучення двох або більше істот у взаємодію одна з одною та спільним світом; саме її я називаю триангуляцією. Вона є результатом тривимірної взаємодії — взаємодії, яка з погляду кожного з двох агентів є двосторонньою: кожен із них одночасно взаємодіє зі світом і з іншим агентом. Іншими словами, кожна істота вчиться співвідносити реакції інших істот зі змінами або об'єктами у світі, на які вона сама також реагує» (Davidson, 1992: р. 118), а у бесіді з Д. Боррадори висловлює свою згоду з тим, що кожен з нас виробляє певну картину світу, а потім розділяє її з іншим. І додає, що «ми не маємо чіткої картини до поки ми не вступимо у комунікацію з іншими людьми» (Borradori, 1995: р. 40). Філософський та інтелектуальний спадок Дональда Девідсона, фокусуючись на інтенційному розумінні суб'єкта та будучи опозицією до позитивістської методології, артикулює на інструментах осмислення соціальної реальності. Девідсон наголошував, що мета гуманітарних наук полягає не тільки в поясненні людської дії та її результатів, а в тому, щоби зробити дії, продукти дій, самих людей, які їх виконують, зрозумілими для їх осмислення (Tashek, 2002). Про витоки такої позиції пише і Д. Боррадори, додаючи невеличку історичну довідку перед бесідою із Д. Девідсоном. Привабливість Девідсона як філософа вона вбачає в тридцятилітньому аналізі мови, наполовину аналітичному, наполовину постаналітичному, того,

що був основою діяльності Віларда Ван Ормана Квайна (Borradori, 1995: р. 42), з яким Девідсон познайомився перед Другою світовою війною і на яку вони пішли добровольцями та служили у воєнно-морському флоті. Після війни Девідсон на семінарах Квайна вивчає проблеми філософії, зрештою приходячи висновку, на відміну від Квайна, Карнапа, що мова та мислення разом зі світом є частинами однієї концептуальної схеми. Девідсон зауважував, що неможливо втратити здатність говорити якоюсь мовою, зберігши при цьому хоч якусь здатність мислити. Свою думку Девідсон підтверджує у бесіді з Д. Боррадори, що будь-яка реальність, як об'єктивна, так і суб'єктивна, формується та підтримується через мову та інтерпретацію (Borradori, 1995: р. 47). Допустимо, що інтелектуальна біографія Девідсона сама по собі ілюструє його теорію, утворивши «спільний світ» причинно-наслідкових зв'язків, що був покладений в основу триангуляції. Якщо Квайн фокусувався на органах чуття індивіда (біхевіористський підхід), то Девідсон, рефлексуючи над їхнім спільним досвідом, зрозумів: знання неможливе без «іншого розуму», який спостерігає за тим самим об'єктом.

Боррадори підкреслює, що комунікативний процес, який потрактували як досягнення консенсусу, завжди розглядався на ґрунті прагматистського мислення, однак для Девідсона, поняття консенсусу залишалося вторинним, і саме це відрізняло Девідсона від прагматизму. Інтерсуб'єктивність, на думку Девідсона, є основою мислення і не залежить від консенсусу. Припускається, що ідеї існують до порозуміння і при зіткненні з іншими ідеями відбувається їхня взаємна згода. Але Девідсон стверджує, що мова, яка розуміється як інтерсуб'єктивне виробництво значень, передує всьому: до включення в певну картину світу ідеї не існують, пояснюючи, що тривимірні відносини між агентами та середовищем, на яке вони разом реагують, є необхідним атрибутом мислення. Однак, зауважує Девідсон, це не є достатнім атрибутом, тому що схожі відносини можуть існувати і у тваринному світі, але останньому ми не приписуємо сили судження. Відповідно, можемо зауважити про ситуацію, яка повинна існувати, якщо існує мислення, але може існувати і незалежно від нього, і тому може передувати мисленню, отже, можемо стверджувати, що певний тип примітивної соціальної взаємодії є складовою історією мислення. В якості ще одного необхідного атрибуту мислення Девідсон називає мову (Borradori, 1995: р. 50). Вказаний трикутник Девідсон називає істотним для існування, і, відповідно, для виникнення мислення, але без трикутника залишаються непоясненими два аспекти мислення: об'єктивність мислення та емпіричний зміст думок про світ.

Квайн, який на думку Девідсона, зробив інтелектуальний прорив, змінивши спосіб аналізу мови та мислення, в «Слові та об'єкті» вперше намагався вирішити проблему розуміння комунікації, але Квайн не ви-

знає метафору трикутника, запропоновану Девідсоном: два співрозмовники та події, що відбуваються у світі. Квайн, пояснює Девідсон, вважає, що потрібно починати з наших органів чуттів і будувати на них все інше, а це ідея, спільна для Декарта та емпіриків. Але Девідсон наголошує, що погоджуючись з тим, що кожен з нас виробляє свою картину світу, а потім розділяє її з іншими, додає, що ми не маємо чіткої картини, поки не вступимо в комунікацію з іншими людьми. Стверджуючи, що існує спосіб філософствування, коли ми виходимо із безпосередньої даності, що може називатись чистим досвідом, чистими враженнями, збудженням нервових клітин, то на цій базі ми створюємо картину зовнішнього світу та інших людей, Девідсон її називає емпіричною або картезіанською. Така картина світу вийде, якщо у світі не буде людей. Свою позицію Девідсон окреслює як таку, що поки ми не знаємо, що відбувається у свідомості інших людей, немає сенсу стверджувати, що маємо поняття об'єктивності, поняття про те, що існує у світі незалежно від нас. На переконання Девідсона, інтерсуб'єктивність — це та галузь, в якій кожен з нас використовує свої власні думки для того, щоби надати смисл думкам інших людей, отже, в процесі спілкування ми створюємо дещо інтерсуб'єктивне, що є об'єктивним. Тільки в контексті інтерпретації, спробі осмислити думки, слова, дії інших людей ми маємо концепції віри, значень, намірів, бажань тощо. Відповідно, ми не можемо осмислити будь-яку істоту як потенційний об'єкт інтерпретації, якщо не можемо осмислити спілкування з нею.

1987 року в праці “Knowing One's Own Mind” Дональд Девідсон описує мисленнєвий експеримент Swampman (Болотна людина): Уявімо, що блискавка б'є в мертве дерево на болоті, я стою неподалік. Мое тіло розкладається на свої елементи, тоді як випадково, з інших молекул, дерево перетворюється на мою фізичну копію. Моя копія (replica), Swampman, рухається так само як рухаюсь я, відповідно до своєї природи Swampman залишає болото, зустрічає і, видається, впізнає моїх друзів, очевидно, відповідає на їхні вітання англійською мовою. «Болотна людина» заходить до мого будинку і, здається, пише статті з радикальної інтерпретації. Ніхто не може помітити жодної відмінності. Але відмінність є. Моя копія не може впізнавати моїх друзів, не може впізнавати взагалі нічого, адже ніколи нічого не розпізнавала з самого початку. Копія не може знайти імен моїх друзів (хоча, звісно, здається, що знає), не може пам'ятати мій дім. Вона не може мати на увазі те саме, що маю я, вживаючи слово «дім», адже звук «дім», який вимовляє Болотна людина, не був засвоєний у контексті, що надав би йому належного значення — або будь-якого значення взагалі. Власне, я не бачу, в якому сенсі можна було б сказати, що моя копія взагалі щось має на увазі, вимовляючи ці звуки, або що вона має будь-які думки (Davidson, 1987).

Мисленневий експеримент Девідсона традиційно слугує полем для суперечок про природу ментального. Однак сьогодні зазначена проблематика набуває рис прикладної проблеми в царині ШІ, де основний конфлікт розгортається між прагматистським підходом, який оцінює інтелект з позицій його інструментальної ефективності, і системою триангуляції Девідсона, за якою об'єктивне знання неможливе без трикутника — «Я—Інший—Світ». Відповідно, якщо Болотна людина, яка виникла абсолютно випадково, позбавлена здатності по-справжньому знати друзів та розуміти слово «дім», то і ШІ, позбавлений тілесного досвіду та соціальної історії, ризикує залишитись тільки «семантичним зомбі».

Мета статті — розкрити механізм формування символічного світу штучного інтелекту крізь призму концепції триангуляції Дональда Девідсона. «Етична суб'єктність» машини виникає не через алгоритмічне моделювання, а через спільну з людиною інтерпретацію об'єктивної реальності. Відповідно, окреслюється шлях створення спільного з людиною символічного та контекстуального простору.

Виокремимо декілька проблемних шарів мисленневого експерименту Девідсона. По-перше, Болотна людина утворилась абсолютно випадково, ані бажань самої людини на її репліку, ані передумов на її створення не простежується, тобто ми не можемо не врахувати ситуацію її утворення, по-друге, коли Болотна людина, як нам здається, відповідає на привітання друзів Девідсона, ми не знаємо насправді, чи вона відповіла їм, ми не запитали друзів і не переконалися, що насправді вітання були. Нам видається, що Болотна людина пише статті з радикальної інтерпретації. Але для того, щоб писати наукові статті, вона, перш за все, повинна мати певну кількість знань, навичок, розуміння об'єктивної реальності для того, щоб написати науковий текст. За версією Девідсона, репліка Девідсона утворилась не з нього самого, а дерево перетворилось на копію Девідсона, відповідно, маємо інший вихідний матеріал для конструювання копії. Відповідно, виникає запитання: на підставі чого нам видається поведінка Болотної людини такою, як вона описана, що вона (поведінка) схожа на поведінку Девідсона, тому що символічний світ, який вибудовує Болотна людина має інше підґрунтя і символи, сконструйовані нею (Болотною людиною), будуть мати інше значення як для неї самої, так і для нашої їх інтерпретації. В літературі, присвяченій аналізу експерименту Девідсона, утворена на болоті репліка, називається двійником Девідсона, але, наголосимо, що сам автор позначає свою репліку займенником "It", в оригінальному тексті читаємо: It moves into my house...; It can't know my friends' names...; It can't remember my house...; It can't mean... (Davidson, 1987). Відповідно, цікавим є питання — а чи є Swampman людиною. Проблема Болотної людини, на думку, артикулює не тільки фізикалістський підхід до природи ментальних станів, але й наголо-

шує на концепції історичної зумовленості та інтенційності. Зазначений «історичний підхід» стверджує, що саме причинно-наслідковий та історичний генезис суб'єкта є конститутивною умовою виникнення інтенційності та значення (Zheng, 2016). Автор повторює за Девідсоном позицію, що обставини виникнення Болотної людини не забезпечують необхідної історії репліки Девідсона. Так, наприклад, Болотна людина не має розуму, у ментальних станів Девідсона є історія, у Swampman її немає, тому що він з'явився через кілька хвилин після удару блискавки, і висновковують, що Swampman є річчю, хоча, як описує експеримент Девідсон, він, Swampman, як нам видається може інтерпретувати і бути інтерпретованим (De Caro, 2011).

У Болотній людині відсутня її особиста історія, її інтелектуальний розвиток, її попередня участь в житті суспільства, розвиток з періоду ембріона до людини, або копії, яка може виконувати функції дорослої людини, наприклад, ходити. За кожною людиною стоїть тривала еволюція її виникнення, і це стосується не тільки біологічної складової, але й тілесності, наприклад, традицій, звичаїв тощо. Як зауважує Р. Мілікан (Ruth Garrett Millikan), віра, бажання, інтереси, прагнення, цілі формуються так само як і вид людини. З цією позицією можна погодитись, якщо йдеться про тривалу історичність формування певної культури щодо бажань, віри, інтересів тощо, що можна висловити пропозиційними установками. Безперечно, можна погодитись з авторкою, що за кожною людиною стоїть тривала еволюційна історія, але авторка наголошує, що принцип, згідно з яким члени людської раси є членами одного й того самого реального роду суттєво вирізняється від принципу, за яким Девідсон та Болотна людина належать одному й тому самому роду, здається некоректним. Девідсон не артикулює схожість біологічну, він зауважує на логічній, ментальній складовій. Дослідниця права, що приналежність окремого організму до певного виду залежить від історичних відносин, вбудованих в простір і час. Але складно погодитись із Р. Мілікан, коли вона стверджує, що «жоден предикат, який посилається на нормальність чи ненормальність не може бути застосований до Болотної людини і жоден, який стосується правильного чи неправильного функціонування тіла або мозку. Вона (Болотна людина) не висока і не низька, не сильна і не слаба, не швидка і не повільна ані фізично, ані розумово, оскільки кожен із цих предикатів посилається на біологічну групу, а її немає. ...Жоден його орган не функціонує ані правильно, ані неправильно, окрім інтересів спостерігача. Оскільки у нього в голові немає органів, які є адаптаціями для розвитку концепцій, фіксації віри, вирішення суперечностей тощо, ніщо не вказує, чи використовує він свій мозок правильно чи неправильно» (Millikan, 1996), і висновок, який робить авторка: «Діяльність не є мисленням, якщо вона існує без мети» (Millikan, 1996).

У праці “Davidson’s Holism: Epistemology in the Mirror of Meaning” Д. Малпас (J. Malpas), презентуючи глибоке дослідження філософії Девідсона, доводить, що холізм є не просто логічним обмеженням в роботах Девідсона, а фундаментальним принципом, який визначає структуру психологічного, лінгвістичного та об’єктивного світів, тим самим стверджуючи про конвергенцію ідей Девідсона з європейською герменевтикою Гайдегера та Гадамера. Переходячи в книзі від ідей Квайна до позиції Девідсона, Д. Малпас зміщує фокус з перекладу на більш широку проблему інтерпретації. Відомо, що «радикальна інтерпретація» Девідсона походить від «радикального перекладу» Квайна, трансформуючи ідеї останнього. Квайн прагне до канонічної мови науки, а Девідсона цікавить теорія значення як така. Термін «інтерпретація» замість «перекладу» підкреслює акцент на експліцитній семантиці. Девідсон ставить запитання: які знання необхідні для розуміння слів іншого? Зрозуміло, що звичайного керівництва з перекладу, який надає кореляцію між реченнями, але не завжди розуміє їхні структури, недостатньо. Відповідно, інтерпретація повинна пояснити, в який спосіб значення речень залежать від їх структури.

Майже всі дослідники зауважують, що адаптація теорії істини Тарського для природних мов була головним досягненням Девідсона. Стратегія останнього полягала в аксіоматичному визначенні поняття виконаності (satisfaction); Тарський використовує значення, щоб підійти до питання істини, оскільки одним з обмежень для Т-речень було те, що речення праворуч мало бути перекладом речення ліворуч. Девідсон «перевертає» Тарського (Davidson, 1987: p. 34), використовуючи істину, щоби дійти до значення, так, наприклад, у “Inquiries into Truth and Interpretation”, він пише, що «в той час, як Тарський намагався проаналізувати поняття істини, апелюючи (в конвенції до Т) до поняття значення (у вигляді тотожності значень та перекладу), я мав на увазі зворотне. Я вважав істину центральним примітивним поняттям і сподівався, детально описавши структуру істини, дійти до значення (Davodson, 1987: p. 35). Відповідаючи на запитання: навіщо використовувати істину для пошуку значення, Девідсон пояснює, що потрібен переклад з об’єктної мови на метамову. І Девідсон пропонує замінити «означає, що» на «істинне тоді і тільки тоді, коли (is T if and only if)», а оскільки при перекладі нас цікавлять еквівалентності між частинами речення, Девідсон і пропонує використати «тоді і тільки тоді, коли», що гарантуватиме, що речення першої частини буде еквівалентним за істиннісним значенням іншій частині. Отже, еквівалентність інтерпретується матеріально, тобто в термінах еквівалентності істиннісних значень.

Тільки лінгвістична взаємодія з іншою людиною та світом є аргументом триангуляції, відповідно, тільки той, хто фактично здійснює останню, має мову та мислення. Йдеться про самі підстави філософії мови: що

вможлиблює в людині мову та думки і що зумовлює те, що слова та думки людини мають ті значення та зміст, які вони мають. Девідсон зауважує на основній вимозі триангуляції — мовці і слухачі мають інтерпретувати інших і самі мають бути інтерпретовані, а не тільки піддаватись інтерпретації. Прикладом триангуляції буде радикальна інтерпретація. Посилаючись на мисленнєвий експеримент Девідсона, уявімо, наприклад, що у самого Девідсона була звичка шуткувати з друзями при зустрічі і коли вони бачились вранці, вітались «Добрий вечір!», а ввечері — «Доброго ранку!», можливо тому, що сам Девідсон любив працювати вночі, а зранку любив поспати. Чи зможе Swampman відреагувати адекватно на такі вітання і чи не буде він виправляти друзів Девідсона, почувши такі фрази.

Девідсон неодноразово підкреслював, що на питання про те, що людина має на увазі своїми висловлюваннями і які думки у неї при цьому виникають, можливо відповісти тільки в семантичному контексті, сказавши, що людина має на увазі або думає то те й про те, а на питання: що уможлиблює мову і думки, відповісти, не прибігаючи до семантичних понять, неможливо. Автори зауважують на відмінній рисі семантичного екстерналізму Девідсона, а саме на тому, що перцептивний екстерналізм забезпечується тільки соціальним екстерналізмом. Останній вирізняється міжособистісною характеристикою, на відміну від комунікативної, тому що для того, щоб мати мову і думки, людина має бути зрозумілою іншими, але при цьому не обов'язково висловлювати одні й ті самі думки одними й тими самими словами. Перцептивний екстерналізм вирізняється тим, що він одночасно є історичним, відповідно, виходячи з історії причинно-наслідкових взаємодій мовців та слухачів з навколишнім світом, жодне слово не має свого значення і жодна думка не має свого змісту, що визначатиметься ізольовано від інших. Як зауважував сам автор «розуміння істини залежить від норми, яка забезпечується тільки міжособистісним спілкуванням, і, звичайно, наявність будь-якого пропозиційного відношення...залежить від розуміння об'єктивної істини» (Meyers, Verheggen, 2016). Для Девідсона існує чітко визначений паралелізм між змістом пропозиційних установок думок та значенням речень. Перші складаються з понять, що роблять внесок у їхній зміст, так само, як інші складаються з термінів, значення яких робить внесок у значення речень, в яких вони вживаються. Ще одна особливість про позиційної установки та думки полягає в тому, що вони можуть бути вираженими або повідомленими за допомогою речень, що мають умови істинності і, відповідно, є істинними або хибними. За позицією Девідсона, ядро смислу і змісту є істинно-умовним. Згадаємо, що розумів Девідсон під концепцією об'єктивності та пропозиційним змістом. Поняття об'єктивності він використовує як синонім об'єктивної істини або іноді, просто істини, або понятті віри, або понятті помилки, або контрасту між вірою та істиною. Володіти

поняттям об'єктивності означає мати «усвідомлення, якби нечітко воно не було вираженим, того факту, що те, що мислиться (або висловлюється), може бути істинним або хибним» (Myers, Verheggen, 2016). Повертаючись до запропонованого нами прикладу жартівливих привітань з друзями Девідсона самого Девідсона та його репліки, виникає питання щодо об'єктивної істини ранку та вечора, коли здійснюються привітання. Зрозуміло, що і сам Девідсон, і його друзі розуміють істинність висловлювання «Зараз вечір», «Зараз ранок», але при цьому вони розуміють і жарт, який вони придумали і в якому здійснюють своє спілкування і який стає для них подією, в якій вони проживають і яку відчують. Якщо, наприклад, до друзів справжнього Девідсона приєднається новий друг і почує таке привітання, він, звісно, здивується, але йому інші пояснять, чому саме так вони вітаються, і ця гра буде розумітись останнім другом істиною, що у випадку з Девідсоном, має конвенційний характер. Однак, чи зможе зрозуміти це репліка Девідсона, чи їй потрібно буде пояснювати всі правила гри, яку придумали справжній Девідсон та його друзі. Якщо, наприклад, будемо використовувати велику мовну модель (LLM), то їй потрібно буде пояснювати всі правила гри, людина ж зрозуміє ситуацію відразу, не потребуючи додаткових пояснень.

У дослідженні (Malpas, 1992) зауважується, що для того, щоб речення, які визначають істиннісні умови висловлювань мовця дійсно надавали смисл, необхідно ретельно продумати формулювання істиннісних умов. Приклад, наведений авторами, засвідчує, що твердження «Геспер тотожній Фосфору» істинне тоді і тільки тоді, коли твердження «Геспер тотожній Гесперу» не буде надавати смисл, оскільки, значення «Геспера» відрізняється від значення «Фосфору», про що свідчить той факт, що хтось може не вірити в тотожність Геспера і Фосфора. Як впливає із зазначеного, для Девідсона значення і зміст є інтенціональними. Зазначена особливість підкреслює ідею, що смисл і зміст навіть базових висловлювань і думок — це щось більше, ніж просто зв'язок між об'єктами у світі та словами і поняттями, які їх виражають. Інакше кажучи, ми не можемо повністю пояснити смисл і зміст в екстенціональних термінах. Як підкреслював сам автор концепції, пропозиційний зміст є цілісним, умовно-істинним. Важливою особливістю є те, що пропозиційні установки та думки, а також речення, що їх виражають, є семантично непрозорими. Ми не можемо замінити кореляційний вираз не ризикуючи змінити його істиннісне значення.

Як неодноразово стверджував сам Девідсон, триангуляція складається із взаємних реакцій двох або більше істот на тривалі стимули та реакції один на одного (Davidson, 1992: p.194), але вона може приймати різні форми в різних контекстах. Відповідно, в них можуть брати участь нелінгвістичні істоти. Девідсон описує приклад з двома левицями, які на-

магаються піймати газель і координують свою поведінку, спостерігаючи одна за одною та реагуючи на реакції одна одної, це приклад доконцептуальної, примітивної форми. У випадку з лінгвістичними істотами триангуляція включатиме повну мовну взаємодію. Наприклад, коли людина вивчає свою власну мову як тільки у неї з'являється поняття об'єктивності, або, наприклад, коли радикальний інтерпретатор, хто вивчає другу або додаткову мову безпосередньо в контексті ситуації. Зазначені триангуляції є необхідними для того, коли людина є одночасно і мовцем і мислителем, але другий є достатнім. Ми знаходимось в ситуації, коли потрібна демонстрація для визначення причини, і, відповідно, смислу висловлювання мовця. Концептуальна позиція Девідсона ґрунтується на позиції, за якою «зміст наших думок та висловлювань частково визначається історією причинно-наслідкових взаємодій у світі» (Davodson, 1992: p. 200).

У межах філософської парадигми Девідсона стверджується, що проста фізична взаємодія між суб'єктами є недостатньою для виникнення думок; така взаємодія повинна набути інтерсуб'єктивної значущості для її учасників. Для того, щоб агенти усвідомили, що вони реагують на один і той самий зовнішній об'єкт, вони мають «когнітивно використати тристоронній зв'язок», тобто реагувати не лише на стимул, а й на саму взаємодію з іншими суб'єктами. Без здатності до змістовного спілкування істоти не спроможні отримати когнітивну вигоду від механізму триангуляції спільного світу. Згідно з аргументом триангуляції, ані нелінгвістичні істоти, ані ізольовані індивіди не можуть володіти думками, оскільки це вимагає наявності поняття об'єктивності — здатності розрізняти істинні переконання та помилки. Це поняття, у свою чергу, ґрунтується виключно на лінгвістичній взаємодії. Отже, мова та мислення виникають одночасно в соціальному середовищі: суб'єкт, позбавлений поняття об'єктивності, не може мати думок, оскільки їхній зміст (значення) фіксуватиметься лише через інтерсуб'єктивні норми та спільне розуміння істини. Девідсон підкреслює, що триангуляція є не просто психологічним описом процесу навчання, а логічною умовою можливості усвідомлення помилки. Мислення та мовлення вимагають здатності диференціювати правильне та неправильне застосування понять, відповідно, саме ця нормативна здатність дозволяє провести межу між суто біологічною дискримінацією стимулів та справжньою концептуалізацією. Водночас концепція стикається з викликами з боку емпіричних досліджень, зокрема, в контексті аутизму. Приклади високо функціональних індивідів, які демонструють володіння мовою за умови дефіциту поняття об'єктивності або теорії розуму, ставлять під сумнів жорсткий детермінований зв'язок між соціальною триангулярністю та лінгвістичною спроможністю. В основі девідсонівського проєкту лежить теза про те, що значення базується на істині: розуміти значення речення — означає знати об'єктивні умо-

ви у світі, за яких воно є істинним. Такий підхід стверджує прямий зв'язок між мовою та дійсністю, минаючи суб'єктивні ментальні посередники, що робить об'єктивну істину фундаментом будь-якої когнітивної дії.

У статті "AI in society: A theory" Р.Ф. Квандт (Ryan Phillip Quandt) переосмислює тріангуляцію Девідсона як модель для людино-машинних команд та систем (Quandt, 2022). Контекст розуміння природи ІІІ включає суто прагматичну схильність людини до антропоморфізації та реальну наявність когнітивних станів у системи, які слід розрізняти. Наприклад, роботизований пилосмок, що рухається зигзагом по підлозі, людина сприймає як пристрій, що «шукає» бруд і відчуває «задоволення», коли знаходить його. Маневрування пилосмоку навколо ніжок стільців чи відстеження ліній стін створює враження цілеспрямованої діяльності, відповідно, стверджує автор, ми використовуємо такі поняття, як «обережність», «виявлення» або «маневр», що в межах девідсонівської традиції ілюструють пропозиційні установки або когнітивні відношення (Quandt, 2022). Опис поведінки пилосмока через ментальні стани припускає наявність у нього внутрішнього життя, до речі відомо, що деякі люди роботизованому пилосмоку вмикають в кімнаті світло, щоб йому було видно, де прибирати. Проте існують вагомні підстави заперечувати наявність у пилосмока справжніх пропозиційних установок, зокрема наміру. Згідно з Девідсоном, намір і переконання є невіддільними: «мати намір — означає вірити, і навпаки». Суб'єкт не може мати наміру підняти чашку, якщо він не володіє переконанням, що чашка знаходиться поруч, що він здатний її дістати і що певна послідовність фізичних зусиль призведе до бажаного результату. Ключовим аргументом проти визнання пропозиційних установок у простих машин, на думку автора, є неспроможність пройти «тест на заміну». Уявімо, що пилосмок спроектований повідомляти про те, що він всмоктує. Якщо він звітує про «бруд», чи означає це, що він також має намір збирати «ґрунт», «крихти» або «шерсть»? Очевидно, що ні. Пристрій не диференціює ці об'єкти як окремі концепти; для його алгоритму вони є функціонально ідентичними подразниками. Ця відсутність концептуального розрізнення вказує на те, що система позбавлена поняття об'єктивності та можливості помилки, які є фундаментальними для девідсонівського розуміння думки (Quandt, 2022). Навіть підвищення чутливості датчиків не змінить ситуації, оскільки проблема лежить не в точності сприйняття, а у відсутності пропозиційного змісту, який би фіксував значення об'єктів через соціальну тріангуляцію. Таким чином, схильність описувати ІІІ в термінах пропозиційних установок є скоріше особливістю людської інтерпретації, ніж характеристикою внутрішньої когнітивної архітектури пристрою, приходить висновку Р.Ф. Квандт.

«Ріші Чандра, віцепрезидент із продуктів та генеральний менеджер Google Nest, у 2019 році зазначив, що ми перебуваємо на етапі переходу

від мобільних обчислень (прикладом яких є комп'ютер у смартфоні) до ембійентних обчислень (*ambient computing*). Останні він визначає як наявність «завжди доступної обчислювальної системи прямо під рукою, яка розуміє користувача, здатна діяти від його імені та надавати підтримку у різний спосіб» (Quandt, 2022). Отже, у межах філософського аналізу зазначений перехід означатиме створення інтелектуального середовища, яке стає дедалі автономнішим та чутливішим для потреб, звичок, бажань людини. Безумовно, складно однозначно оцінити, що формування складних систем «людина-машина», де ШІ виступатиме вже не просто як інструмент, а проксі агентом для рішень людини. Значений технологічний зсув актуалізує проблематику Девідсона: чи можна вважати дії такої системи результатом її власних пропозиційних установок, чи вона лише ефективно віддеркалюватиме інтенції користувача. Вимірювання суспільних ефектів впровадження штучного інтелекту стикається з трьома фундаментальними викликами: 1) включення соціального, контекстуального характеру дії, 2) концептуалізація траєкторії розвитку, включаючи людських агентів, 3) дозвіл ШІ відхилятися від очікування результату. Автор наголошує, що ці проблеми пов'язані, тому що машини з'являються в суспільстві як соціальні агенти.

Р.Ф. Квандт, описуючи власну авторську концепцію, зауважує, що запропонована ним теорія — це триангуляція, яка виражає як траєкторію, так і взаємодію, порівнює її з розумінням триангуляції в математиці — це спосіб визначення відстані точки від базової лінії шляхом систематичного виміру іншої точки, пов'язаної з нею (Quandt, 2022). В контексті соціальних відносин триангуляція описує, як хтось сприймає об'єкт відносно іншої людини (базове відношення), до також взаємодії з тим самим об'єктом. Автор постійно наголошує в статті, що триангуляція не є абстрактною філософською аналогією Девідсона, а є практичною моделлю для аналізу та проектування людино-машинних систем. Автор пропонує використовувати триангуляцію для визначення, наскільки ШІ здатний до «спільної діяльності» з людиною, що перегукується з мисленнєвим експериментом Девідсона, чи може репліка Девідсона замінити самого Девідсона. Відповідно, це вимагатиме формування інтерсуб'єктивної реальності, де обидва агенти визнають спільний об'єкт через взаємну кореляцію своїх реакцій. Девідсон, як зазначає автор, ніколи не формулював триангуляцію як суворий формальний аргумент, автор адаптує її для людино-машинних команд, надаючи їм емпіричної цінності — здатності пояснювати та передбачати, підлягати спростуванню в експериментах. Основний внесок автора полягає в тому, що будь-яке тестування ШІ вже передбачає певні «теоретичні зобов'язання». Триангуляція слугує рамкою, яка дозволяє зрозуміти, чому певна модель взаємодії наближається до істини ще до початку її практичного тестування. Отже, текст обґрун-

товує перетворення філософської концепції тріангуляції на наукову методологію для створення ефективних інтелектуальних систем, де людина і машина діятимуть як одна команда.

У статті “The Symbol Grounding Problem” С. Гарнад (S. Harnad) зауважує, що властивості спеціалізованих символічних систем мають виявитися залежними від поведінкових міркувань (Harnad, 1990: р. 345). Штучний інтелект потребує формування власної символічної системи, не тієї, що дозволяє великій мовній моделі працювати, а йдеться про спільний символічний простір з людиною, оскільки саме остання використовує його інструмент для спілкування. Один із видів формування символічного обґрунтування С. Гарнад наводить на прикладі вивчення китайської мови як другої на підставі китайсько-китайського словника (Harnad, 1990: р. 339). Подорож словником буде нагадувати карусель, постійний перехід від одного символу до іншого, від означеного до означуваного, жодного разу не зупиняючись на тому, що означає символ. Автор наголошує, що система, на перший погляд, може вести себе цілком розумно протягом певного часу, а може бути підірвана несподіваним випадком, який демонструє її глупоту: програма розуміє сцену, описує її, відповідає на запитання, демонструючи своє розуміння (хто що зробив, де, чому), а потім раптом показує, що вона «не знає», що повісити слухавку і вийти з кімнати не робить телефон зниклим. Розглянемо опис символічної системи Гарнада. Символічна система — це 1) набір довільних фізичних токенів (подряпини на папері, дірка на стрічці, події в цифровому комп’ютері тощо), які 2) маніпулюються на підставі певних правил, а 3) також є фізичними токенами. Маніпуляції символами на підставі правил ґрунтуються 4) виключно на формі символічних токенів (а не на їхньому значенні), відповідно, є синтаксичними і складаються з 5) правил комбінування та рекомбінування символічних токенів. Існують 6) примітивні атомарні символічні токени та 7) складові рядків символічних токенів. Вся система та її частини — атомарні токени, складові токени, синтаксичні маніпуляції (як фактичні, так і можливі) і правила — всі вони є 8) семантично інтерпретованими.

Отже, для того, щоб ШІ або великі мовні моделі (LLM) могли зкомунікувати з людиною, вони мають вийти за межі суто синтаксичної маніпуляції і вирішити проблему обґрунтування символів. Якщо цього не зробити, ШІ залишиться замкненим у герменевтичному колі, де значення знаків лише паразитуватимуть на смислах людини. Тріангуляція можлива тільки через гібридну модель, яка обґрунтує символи, починаючи з чуттєвості і до абстрактних понять. Імовірно, система повинна отримувати «сенсорні проєкції» реальних об’єктів (аналогів того, як люди бачать, чують той чи інший предмет). Згадаємо Клару із «Клари і Сонця», вона бачила світ поділенням на невеличкі сегменти у вигляді сітки, відпо-

відно, сконструювати всю картину світу вона не могла. Щоби залишитись в тріангуляції з людиною, ШІ повинен демонструвати поведінкову здатність вирізняти та ідентифікувати об'єкти у спільному фізичному просторі. Значення символу «фіксується» здатністю системи правильно реагувати на об'єкти та стани справ у світі, до яких ці символи відносяться. Без обґрунтування спільної символічної системи в причинно-наслідковому досвіді ШІ не зможе вийти на істинну тріангуляцію з людиною і комунікація в такому випадку залишиться тільки ілюзією, створеною людиною-інтерпретатором.

Висновки

Концепція тріангулярності Дональда Девідсона артикулює можливість значення та мислення тільки в історичній та соціально побудованій практиці. Мисленнєвий експеримент Swampman, демонструючи історію виникнення Болотної людини, виконав концептуальну функцію, продемонструвавши, що без причинно-наслідкової історичності унеможливаються значення, спогади, розпізнавання, думки. Імітуючи поведінку людини, Swampman, так само як і Ішигурова Клара, позбавлений ментального змісту, оскільки значення слів залежать не тільки від внутрішнього стану людини, а й від історії їх вживання. Відповідно, тріангулярність Девідсона зафіксувала фундаментальну асиметрію, себе ми розуміємо безпосередньо, без підтвердження, інших — через поведінку, мову, мовлення, контекст, соціальні зв'язки. Формування символічного світу машини не може бути ізольованим алгоритмічним обчисленням, без визнання «спільного світу» ШІ залишиться в позиції Болотної людини — системи, що тільки імітує комунікацію, але позбавлена справжнього пропозиційного змісту. Людина прагне наповнити сучасні великі мовні моделі (LLM) етичним змістом, але етична суб'єктність машини має бути динамічним продуктом спільної діяльності, щоби максимально запобігти ризикам використання ШІ в буденному та науковому житті, а не вбудованою характеристикою.

ДЖЕРЕЛА/REFERENCES

- Ashouri-Kisomi, M. A., Parvizi, M. & Momeni, S.M.H. (2025). Pragmatism, the Ethics of Artificial Intelligence and Sociotechnical Systems. *Journal of Philosophical Investigations*, 19(52), 281—300. <https://doi.org/10.22034/jpiut.2025.66643.4060>
- Borradori, G. (1995). *The American Philosopher: Conversations with Quine, Davidson, Putnam, Nozick, Danto, Rorty, Cavell, MacIntyre, Kuhn*. Chicago, London: The University of Chicago Press.
- Chalmers, D.J. (2025). Propositional Interpretability in Artificial Intelligence. <https://doi.org/10.48550/arXiv.2501.15740>
- Davidson, D. (1987). Epistemology Externalized. In: *Subjective, Intersubjective, Objective* (pp. 193—204). Oxford: Clarendon Press.

- Davidson, D. (1987). Knowing One's Own Mind. In: *Subjective, Intersubjective, Objective* (pp. 15—38). Oxford: Clarendon Press
- Davidson, D. (2001). *Mental Events*. Retrieved from: <https://divinecuration.github.io/assets/pdf/davidson-mental-events.pdf>
- Davidson, D. (1992). The Second Person. In: *Subjective, Intersubjective, Objective*. (pp. 107—122). Oxford: Clarendon Press.
- Davidson, D. (1992). Three Varieties of Knowledge. In: *Subjective, Intersubjective, Objective*. (pp. 205—220). Oxford: Clarendon Press.
- Davidson, D. (2001). *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press
- De Caro, M. (2011). *Davidson's Externalism. The Short Happy Life of the Swampman: Interpretation and Social Externalism in Davidson*. Retrieved from: https://www.academia.edu/6549574/Davidsons_Externalism
- Harnad, S. (1990). *The Symbol Grounding Problem*. <https://doi.org/10.1016/0167-2789%2890%2990087-6>
- Malpas, J. (1992). *Donald Davidson and the Mirror of Meaning: Holism, Truth, Interpretation*. Cambridge University Press.
- Mendoza, A. (2025). *Don't Believe the Believe Hype!* Retrieved from: <https://openreview.net/pdf?id=mCPn3aqf5U>
- Millikan, R.G. (1996). *On Swampkinds*. Retrieved from: <https://philosophy.uconn.edu/wp-content/uploads/sites/365/2014/02/On-Swampkinds.pdf>
- Myers, R.H., Verheggen, C. (2016). *Donald Davidson's Triangulation Argument: A Philosophical Inquiry*. New York: Routledge.
- Quandt, R.P. (2022). AI in Society: A theory. *Front. Phys. Sec. Snterdisciplinary Physics*. (vol. 10—22). <https://doi.org/10.3389/fphy.2022.941824>
- Taschek, W. (2002). *Making Sense of Others: Donald Davidson on Interpretation*. Retrieved from: <https://surl.li/cgcxzm>
- Watson, D.S., Mokander, J., Floridi, L. (2024). *Competing narratives in AI ethics: a defense of sociotechnical pragmatism*. <https://doi.org/10.1007/s00146-024-02128-2>
- Zheng, Y. (2016). *The Swampman Puzzle and Diachronic Holism*. <https://doi.org/10.1111/phil.12113>

Отримано / Received 14.12.2025

Прийнято до друку після рецензування /
Accepted for publication after review 29.02.2026

Опубліковано / Published 02.04.2026

Olena ASTAPOVA-VYAZMINA,

PhD, Associate Professor at the Department
of Philosophical, Political and Psychological Studies,
Cherkasy State Technological University,
480, bulv. Shevchenka, Cherkasy, 18006
o.astapova@chdtu.edu.ua
<https://orcid.org/0000-0003-2104-9703>
WoS Research ID: PIJ-2547-2026

DAVIDSON'S TRIANGULATION AND ARTIFICIAL INTELLIGENCE:
SEMIOTICS OF DIGITAL SUBJECTIVITY

The article explores the transformation of Artificial Intelligence (AI) from a mere instrumental tool into an autonomous mediator through the lens of Donald Davidson's epistemological concept of triangulation. The relevance of the study is driven by the necessity for

a philosophical reflection on Large Language Models (LLMs), which presuppose considering them not merely as algorithmic systems but as potential subjects of communication. This shift requires a re-evaluation of propositional attitude categories, such as beliefs, intentions, and desires. The scientific novelty lies in the juxtaposition of Davidsonian «triangular externalism» (Self—Other—World) with contemporary issues of digital subjectivity. By referencing the «Swampman» thought experiment, the author argues that without historical genesis and a causal connection to a shared world, AI risks remaining a «semantic zombie.» In contrast to W. Quine's approach, Davidson's strategy asserts that objectivity of knowledge arises only through intersubjective interaction; consequently, understanding truth becomes possible solely through language as a form of social triangulation. Within the scope of the study, a parallel is drawn with the fictional image of the robot Klara from Kazuo Ishiguro's novel as a late model of Davidsonian triangulation, where ethical subjectivity emerges through an attempt to interpret the human context. Particular attention is paid to the logical structure of radical interpretation and the adaptation of Alfred Tarski's theory of truth. The author demonstrates that the meaning of sentences in AI systems remains «semantically opaque» without engagement in a shared symbolic space. The conclusions emphasize that the transition from mobile to ambient computing requires AI to possess more than just precision; it demands the capacity for the «cognitive utilization of the three-way link.» It is established that attributing propositional attitudes to AI is often a result of human anthropomorphism, whereas genuine subjectivity requires a history of interaction with humans. The article proposes regarding triangulation as a fundamental model for developing human-machine systems, where ethical responsibility and meaningful communication become possible only within the «closed triangle» of a shared objective reality.

Keywords: *triangulation, Donald Davidson, artificial intelligence, digital subjectivity, Swampman, social reality, radical interpretation, epistemic situation.*