

КУРС "ДОСЛІДНИЦЬКІ МЕТОДИ Й АНАЛІЗ У ФІЛОСОФСЬКИХ ДОСЛІДЖЕННЯХ": ІІІ ЯК ІНСТРУМЕНТ, ВИКЛИК І ДЗЕРКАЛО ЛЮДСЬКИХ НЕДОЛІКІВ

Перетин філософських досліджень і штучного інтелекту (ІІІ) є, на перший погляд, багатогрунтовым ґрунтом для досліджень, але тільки на перший погляд. Адже це непросте завдання – по-справжньому заглибитись у фундаментальні філософські принципи, на яких базуються розробки ІІІ, та усвідомити всі наслідки цих стосунків. І це не просто роздуми, а пошук фундаментальних принципів, які визначають цінність, ефективність і ризики ІІІ, а також впливають на здатність генерувати нові ідеї та безпеку. У статті обговорюється *залучення ІІІ у викладання філософських дисциплін*, зокрема на прикладі курсу "Дослідницькі методи й аналіз у філософських дослідженнях". Автор також аналізує *етичні та когнітивні проблеми ІІІ*, використовуючи дві студентські роботи як ілюстрації того, як використовуються здобутки курсу для розгляду зокрема упередженості ІІІ та його обмеженої здатності до "судження" порівняно з людським. Автор досліджує *використання ІІІ в академічній праці*, визнаючи його переваги як інструменту, але також піднімаючи питання щодо авторських прав, достовірності інформації та необхідності людського контролю. Ця стаття має на меті дослідити деякі важливі взаємозв'язки, наголошуючи на тому, що етика не є єдиною філософською дисципліною, яка застосовується до ІІІ, що філософські дослідження на рівні онтології, епістемології й телеології збагачують наше розуміння ІІІ та кидають виклик звичайним методологіям, спонукаючи до переоцінки наших освітніх рамок у світлі цих досягнень. У статті йдеться про найновіші підходи до оцінки ролі філософії, яка закладає принципи створення ІІІ, а отже є фундаментальною частиною ефективного ІІІ. Утворюється така собі область філософії ІІІ, що заглиблює й висвітлює важливі теми, поєднуючи теоретичні основи з практичним застосуванням у технологіях. Крім того, розгляд різноманітних концепцій, як от пояснюваний ІІІ (explainable AI), підкреслюють потребу розуміння користувачами, у чому полягає сенс складних моделей, зокрема інформаційних, хоча б у загальних рисах. У висновках наголошується про *важливість моральної відповідальності* у розробці та використанні ІІІ, особистої й колективної.

Ключові слова: ІІІ, філософія ІІІ, етика ІІІ, дослідницькі методи і аналіз, технології ІІІ, великі мовні моделі (LLM).

Сьогодні чи не найпопулярнішою темою, що обговорюється в ЗМІ та наукових дослідженнях, є ІІІ та його вплив на суспільство. Навіть якщо хтось не використовує продукти ІІІ, ігнорувати їхнє існування неможливо.

У заголовку зроблено акцент на назві курсу, і це не випадково. Авторський курс "Research Methods and Analysis in Philosophical Studies"/ "Дослідницькі методи й аналіз у філософських дослідженнях", який я читаю впродовж трьох семестрів на міжнародному відділенні філософського факультету Варшавського університету, спрямований саме на дослідження філософських проблем, напрямів і течій з погляду їхньої методології й концептуального складу. Очікується, що по закінченню курсу студенти: самостійно визначатимуть цілі, дослідницькі запитання, гіпотези, верифікуватимуть результати, застосовуватимуть етичні правила й вимоги до дослідження, а також формуватимуть особисту зацікавленість через діалог, критичне мислення, базові принципи теорії аргументації та комунікативні навички.

Очікувані результати ніби не містять в собі володіння ІІІ, і я не маю фахової підготовки з цього предмету. Ми вчимося разом: знайомимось з базовими принципами, обговорюємо різні підходи до використання ІІІ. Чи не у кожній групі моїх студентів,

хтось пише дослідницьку працю про ІІІ. Наведу до прикладу дві найцікавіші праці: “Ethical Problems Concerning the Development of AI”/”Етичні проблеми розвитку ІІІ” Войцеха Вільчинського та “Can AI judge properly?”/”Чи може ІІІ судити правильно?” Каміла Яблонського.

Автор першої праці прагне показати, що розвиток та функціонування ІІІ відбувається через відображення людських рис та різних суперечностей суспільства в його функціонуванні. Не тільки це, але й ідея про те, що література про ІІІ зосереджена переважно на успіхах ІІІ та на моделюванні його «свідомості» як ідеальної свідомості. На думку В. Вільчинського, набагато менше аналізується література, яка висвітлює недоліки та суперечності людського інтелекту, поширені помилки та хибну поведінку. Проблема полягає в тому, що люди, які створюють ІІІ, «тягнуть» за собою всі людські недоліки та упередження без можливості «еволюції», наявності сумніння та саморозвитку. Автор описує внутрішню (створення ІІІ як технічного завдання) та зовнішню (використання ІІІ реальними людьми для своїх цілей) лінії філософських досліджень ІІІ, кожна з яких має відповідні філософські наслідки, дає опис вразливих рис інструментів ІІІ. Один з висновків цієї праці: закладати етичні принципи потрібно на етапі створення систем ІІІ, а не тоді, коли з'являються негативні наслідки.

Друга праця ставить за проблему онтологічну й когнітивну невизначеність ІІІ. Її основна теза: ІІІ не здатен “судити” (тобто оцінювати, ухвалювати рішення) в повному людському сенсі, попри зростаюче використання у сферах, що потребують моральної відповідальності. ІІІ залишається інструментом, а не агентом зі свідомістю. ІІІ також не має намірів, а отже не несе моральної відповідальності – відповідальність несуть люди (розробники, користувачі): не можна повністю покладатися на ІІІ у прийнятті рішень, потрібно вирізняти, які рішення мають залишатися за людьми (права, свободи). Надмірна антропоморфізація ІІІ розробниками й користувачами містить в собі неабияку небезпеку через сприйняття ІІІ як “емпатійного агента”. LLM здатні імітувати, але не володіти справжнім розумінням, виконуючи завдання в обмеженому режимі, позбавлені контексту, культурного розуміння й емоцій. Оскільки, на думку автора, справжнє судження вимагає не тільки функціональних, а й нормативних критеріїв, ІІІ не інтегрується повноцінно до людського смислоутворення. Цікавими є думки К. Яблонського, в яких він попереджує проти техноутопізму (“ІІІ як дзеркало наших бажань”), а також про те, що ІІІ – корисний інструмент у структурованих завданнях, але справжнє мислення потребує людських якостей: емпатії, відповідальності та інтерпретації. А відтак найкращою може бути гібридна модель взаємодії: ІІІ прискорює процес вирішення завдань, а людина – контролює й оцінює. Обидві роботи виконувались на моїх очах “з нуля”, з вибору теми й дослідницьких гіпотез, вони цікаво доповнюють одна одну: перша зосереджується на проблемах упередженості (bias), вразливості ІІІ й етичних проблемах, а друга – на когнітивних характеристиках, агентності ІІІ, відсутності правдивого людського смислоутворення й емпатії.

У добу, коли дедалі більше текстів створюються алгоритмами, особисте залучення автора, емоційний наголос, інтонаційна насиченість набувають особливого значення. Такі прояви не є випадковими – вони позначають присутність суб'єкта, мислячої й відповідальної істоти, яка не лише структурує зміст, а й проживає його. На відміну від текстів, згенерованих ІІІ, у яких бракує автентичної інтенції, людський текст свідчить про досвід, сумнів, афект і вибір. Ця суб'єктність – одна з тих рис, які відрізняють філософування від симуляції мислення.

Свідомо оминаючи історію розвитку ІІІ, почнемо з найближчого до нас етапу, коли відбувається перехід до творчих процесів, особливо письма (writing). Картина стрімко змінюється, і ось уже перетворення письмового або друкованого тексту на подкаст і відео, як і створення агентів з функціями персональних асистентів, наших віртуальних двійників, що читатимуть лекції за нас, стає реальністю, з якою поки що граються й експериментують. Але створення різного роду текстів залишається одним із

найпопулярніших напрямків. ChatGPT використовують для пошуку, класифікації об'єктів дослідження, генерування даних, і ще й переписування з редагуванням, не кажучи про подолання письменницького ступору (Writer's block), і все це робиться за лічені хвилини. Існують і більш спеціалізовані варіанти ШІ для академічної праці й пошуку та обробки літератури. Застосунки ШІ, вмонтовані в інтернет браузер, дають відповіді на будь-які запитання, від побутових до наукових. Як не тішитись з появи такого "асистента", рятівного помічника й друга у самотні моменти творчих пошуків? Правда, доводиться також розбиратися із наслідками такого "енциклопедизму" ШІ. Справа в тому, що ШІ знаходить інформацію, використовуючи працю сотень тисяч академічних працівників (науковців, викладачів), а також письменників. Через це "творчість" за допомогою ШІ утворює так звану "сіру зону" – неконтрольовану сферу діяльності, що не врегульована законодавством. До того ж стрімкий розвиток ШІ фактично не дає можливості ретельно дослідити всі продукти цього розвитку й визначити всі їхні переваги й вади у застосуванні до наукової діяльності, встановити межі, а головне – відповідальність за його застосування. Треба визнати, що ШІ реалізує свою "обізнаність" через безпрецедентне в історії порушення авторських прав. На додачу, ШІ може вигадувати авторів текстів, видаючи сконструйовані ним тексти за публікацію, якої насправді не існує. Через це активні користувачі ШІ зізнаються, що витрачають купу часу на встановлення правдивості посилань у списку з огляду використаних джерел. Кожен вирішує самостійно, чи користуватися таким трудомістким засобом створення текстів, але механізм запозичення ШІ літературних джерел вимагає негайного врегулювання й постійного контролю за дотриманням авторського права.

Один із провідних світових експертів з ШІ Паскаль Борнет, автор наукового-інформаційного блогу на LinkedIn, IRREPLACEABLE with AI, стверджує, що прийняття ШІ веде до підвищення креативності, здатності поєднувати різні інформаційні царини між собою, що на перший погляд здаються випадковими (як от відвідування Стівом Джобсом курсу каліграфії привело до створення революційної типографіки десятиліття потому). В ідеалі, ШІ може вправно поратись з великими базами даних, структурувати їх, створювати несподівані зв'язки, але тільки людська творчість може надихнути ШІ на максимальний творчий ефект: «Наша творчість – це не просто поєднання існуючих елементів – це малювання з багатого гобелена нашого життєвого досвіду, емоцій і випадкових зустрічей. Отже, підсилення нашої людської креативності допомагає підвищити цінність, яку ми можемо створити, працюючи разом із ШІ. Синергійний ефект» [Bornet, 2025].

Розглядаючи різні погляди на співпрацю зі ШІ, варто звернути увагу і на думки таких експертів, як Майк Фергюсон, який пропонує іншу, більш інструментальну, перспективу: «Мусимо чесно зізнатися, що відтоді, як тема ШІ стала популярною, ми використовували її для обговорення критичного мислення, постановки запитань, не кажучи про дискусії з етичних проблем, що їх може породити ШІ, якщо дати йому волю», – нагадує він на сайті Academia.edu в обговоренні статті Майкла Герліха [Gerlich, 2025], що влаштував Марк Шампань. У коментарі до дискусії Фергюсон зазначає: «Після ознайомлення з деякими вступними матеріалами про штучний інтелект..., я згадав прислів'я, що ті, хто навчиться користуватися ШІ, замінять тих, хто не вміє. Деякі педагоги використовують штучний інтелект, щоб запропонувати метафори для урізноманітнення педагогіки, створити екзаменаційні та інші запитання для сприяння навчанню». На його думку, штучний інтелект – це просто ще один інструмент: «Як і книгу, її можна читати розумно чи ні, використовувати її як обмежувач чи підняти монітор комп'ютера на зручнішу висоту. Книзі буде байдуже, як і ШІ. Але читачі книг, схоже, отримують вигоду від розумного використання, ніж ті, хто радше спалив би книгу, оскільки вважав її загрозою або відчайдушно потребує палива. Я пам'ятаю, що деякі безцінні книги спалили для приготування їжі» (йдеться про бібліотеку Nag Hammadi library - Н.В.) [Shampagne, 2025].

Синергійний ефект, про який йшлося, набуває особливої ваги, коли йдеться про філософські дослідження. Але є одне застереження. Філософська праця має свої

особливості. Окрім читання велетенської кількості першоджерел, навчання філософії привчає до постійного діалогу з класиками через їхні тексти, цитування з авторитетних джерел, запозичення ідей, концепцій, підходів від авторів різних епох, у тому числі сучасних. І тут не йдеться про плагіат. Все, що стосується поділяння думок класика або сучасника, наслідування стилю і комплексу ідей, у тому числі з метою критичного аналізу й інтерпретації, а також полегшення сприйняття складних філософських побудов в освітньому процесі є цілком законною діяльністю осмислення й реценсії думок у філософії, так і в цілому гуманітарних дослідженнях. Втім, цей ефект – довіра, покладання на думку інших або деференція [Вяткіна, 2019] – стосується не тільки запозичення, а радше слідування авторитетній думці з причин ідеологічних, партійних, групових, родинних, не кажучи про професійну причетність, суб'єктивну схильність до певного складу й типу думок. Описана професійна особливість іноді може утнути злий жарт з дослідником. Бо фаховий філософ, схильний до залежності від описаних факторів, а найбільше від домінуючої авторитетної думки, може непомітно для себе, себто автоматично, підпорядкуватися ІІІ з його переконливо генерованою думкою. Пригадаймо: що відбувалося після падіння комуністичного режиму й єдиної панівної марксистсько-ленінської парадигми, в річниця якої працювали філософи радянської доби. Деякі з колег впадали у справжній відчай з тієї причини, що впав їхній стійкий “символ віри”. “У що тепер вірити?” – щиро волали вони у відчай, може тому, іронічно кажучи, що в ті часи ще не існувало ІІІ.

Філософські дослідження та ІІІ доволі тісно перетинаються, особливо у дослідженні передумов, наслідків та етичних міркувань, пов'язаних із використанням ІІІ [Мироненко, 2025]. Кожна друга конференція, пов'язана з епістемологією, когнітивними науками долучає розгляд етики ІІІ. Надійну основу для аналізу цих перетинів можна знайти у філософії інформації Лучано Флоріді, який досліджує вплив ІІІ на суспільство, залучаючи до цих досліджень провідних фахівців з різних царин філософії й комунікаційних технологій [Gorichanaz et al, 2020]. Разом ці дослідження підкреслюють критичний вплив філософських досліджень на формування майбутньої траєкторії ІІІ. Справедливо буде також згадати, що Флоріді фактично започаткував етику ІІІ [Floridi, 2015], створивши Onlife Initiative: “реінжиніринг концепції для переосмислення суспільних проблем у рамках цифрового переходу” [Ibid., pp.1-2]. Переосмислення необхідне, бо традиційні етичні теорії виникли задовго до сучасних технологій. Сучасна прикладна етика, яка охоплює етичні питання, пов'язані як з існуючими технологіями ІІІ, так і неіснуючими, але можливими у майбутньому. Можна передбачити ситуації, коли ці технології будуть перевершувати людський інтелект, і вже маємо роботів, які здатні навчатися і й самостійно виправляти помилки. Все важче стає контролювати подібні технології й узгоджувати їх з людськими цінностями. Тому на перший план виходить таке поняття як моральна відповідальність. Останні досягнення ІІІ зі створення дипфейків для інформаційної війни з усією очевидністю демонструють ризики й загрози конфліктів не у далекому майбутті, а вже зараз.

Інтеграція дослідницьких методів і ІІІ у філософські дослідження – це такий собі трансформаційний зсув у створенні знань у цій дисципліні. Дуже інтригуюче звучить заголовок статті «Philosophy Eats AI» [Schrage, Kiron, 2025], що довелося придбати її на сайті MIT Sloan Management Review. Зміст статті надихає філософів. Автори статті стверджують, що філософія стає ключовим чинником у розвитку, навчанні та використанні ІІІ. Якщо раніше говорили, мовляв, “програмне забезпечення поглинає світ”, то нині саме філософія “поглинає” ІІІ, визначаючи його цінність, ефективність і ризики для бізнесу. У відповіді на питання: чому філософія важлива для ІІІ, автори зазначають: філософія – це не лише етика по відношенню до філософської проблематики ІІІ. Хоча етичні питання, як вже про це йшлося, залишаються в центрі уваги і є засобом ширшого впливу філософії на різні сфери діяльності. Важливими є також: *Онтологія*: щодо того, як ІІІ “бачить” і структурує реальність. *Епістемологія*: на яку

філософію знання спирається ІІІ, а отже, як визначається істина / хиба. *Телеологія*, що визначає, для чого створюється та яку мету має досягати та чи та модель ІІІ.

Попередня версія мого курсу була спрямована на дослідницькі методи й аналіз у науках про бізнес, і тому мені близькі міркування авторів цієї статті про те, що без філософського підходу не можна отримати конкурентної переваги у бізнесі, бо філософські настанови впливають на моделі прийняття рішень і взагалі на все, що стосується коректного відображення подій, фактів й аналітики у різних конфігураціях ІІІ. Важливо пам'ятати, на думку авторів, що філософія – це не надбудова, а фундаментальна частина ефективного ІІІ. Саме від того, які принципи закладаються у відповідні системи ІІІ, залежать етичні принципи, рівень агентності ІІІ, як і безпека разом із здатністю генерувати нові ідеї: «Ті, хто ігнорує філософію, створюють потужні, але зрештою обмежені інструменти; ті, хто її впроваджує, отримують AI-партнерів, здатних розвивати стратегічну місію організації» [Schrage & Kiron, 2025, p. 13]. І це вимагає у підсумку активної участі філософів у дослідженні філософії ІІІ, бо Інститут етики ІІІ [Institute of Ethics in AI] Оксфордського університету на 10 червня 2025 року оголосив семінар під назвою “Що відбувається, коли ми вирішуємо не вибирати – і дозволяємо штучному інтелекту чи іншим людям вирішувати за нас?” (What happens when we choose not to choose – and let AI or others decide for us?).

Оскільки технологія ІІІ невпинно поширюється, вона запроваджує нові методології для аналізу й тлумачення філософських концепцій, хоча це не дуже легко продемонструвати, бо для цього потрібен довший час. Наприклад, застосування онтологій для представлення складних філософських ідей дозволяє більш структурований підхід до управління знаннями, гарантуючи, що нюанси людського мислення ретельно відображаються та аналізуються [Brewster et al., 2004.]. Крім того, перетин соціальної епістемології з системами ІІІ має глибокі наслідки для розуміння того, як конструюються та поширюються знання в епоху цифрових технологій. Філософи, які краще орієнтуються в етичних і епістемологічних викликах штучного інтелекту, ефективніше долають упередження, властиві знанням, опосередкованим людиною та технологіями [Godler et al, 2020].

Формування онтології та роль логіки у процесі пізнання можна розглядати з позицій, близьких до філософського функціоналізму. Відстежуючи історію й сучасний стан досліджень щодо штучного відтворення раціонального мислення та здобуття знань через *інтелектуальні автомати*, термін, який уживає мій колега Олександр Маєвський, знаходимо перетини класичних філософських методів з їхніми працюючими моделями. Експертні системи порівнюються з моделями людської інтелектуальної діяльності, виявляючи їхню спорідненість, що сприяє функціональній успішності таких систем, але призводить до **дегуманізації** поняття інтелекту, оскільки їхня логіка і онтологія є непрозорими [Маєвський, 2020].

Філософія має достатній концептуальний інструментарій для аналізу різного роду моделей ІІІ, у тому числі великих мовних моделей (LLM), щоб визначити їх як механістичний проєкт статистичного моделювання мови та мовлення, щоб виявити їхній спосіб відображення дійсності та епістемічну обмеженість [Маєвський, 2024].

Таке методологічне розмаїття цікаве, але воно створює проблеми для міждисциплінарного оцінювання, враховуючи різноманітну термінологію та задіяні рамки. Чи можливо створити стандартизовану епістемологічну модель, щоб і дослідникам, і студентам можна було б орієнтуватись в цих відмінностях, гарантуючи, що філософські дослідження залишаються доступними та дійсними в рамках ширшого академічного дискурсу [Bell et al., 2010]?

Поки ми намагаємося шляхи до порозуміння між філософами і ІІІ, без гучних заяв і сенсацій створюється курс з розробки моделей та методів керування навчальним контентом на основі Web-систем в контексті сучасних освітніх викликів. Курс дає можливість створювати карти понять, уможлиблює керування синтезом знань в

навчальних системах, відтворює модель освітнього запиту [Титенко, 2024]. Без згадування ІІІ даються базові понятійні інструменти, на основі яких створюються коректні, тобто у межах своєї компетентності, інформаційні системи із закладеними в них принципами й системами контролю, що про їхні наслідки й результати точаться дискусії десь далеко за межами чітких навчальних програм.

Висновок

Підсумовуючи, перетин філософських досліджень і ІІІ є спільним ґрунтом для дослідницьких методів і аналітичних технік, які поглиблюють наше розуміння обох сфер. Поточний дискурс свідчить про те, що ІІІ необхідний, але для його сприйняття й прийняття суспільством потрібне чітке розуміння його систем, призначення і меж застосування. Без надійної основи для визначення того, що означає розуміти ці системи, спроби пояснення можуть зрештою не досягти намічених цілей [Páez, 2019]. Дослідження філософських питань, пов'язаних з інформатикою, покращує наше розуміння методологій штучного інтелекту, пропонуючи систематичний підхід до визначення принципів, які керують розумною поведінкою та процесами прийняття рішень [Rapaport, 2005]. Отже, коли ми орієнтуємося в складності ІІІ, філософська перспектива не тільки збагачує наші аналітичні інструменти, але й культивує її вдумливе ставлення до технологій для людського розуміння. Інакше може бути, як у прикладі Г. С. Сковороди: «Если видишь на старой в Ахтырке церкви кирпич и вапну, а плана ея не понимаешь, как думаешь – усмотрел ли, узнал ея?» – «Никак! Таким образом, одну только крайнюю и последнюю наружность вижу в ней, которую и скот видит, а симметрии ея, или пропорции ея который всему связь и голова материалу, понеже в ней не разумею, для того и ея не вижу, не видя ея головы» [Сковорода, 1983, с. 115]. «Як бачимо, тут і пропорція, і симетрія, бракує тільки слів про те, що тварне є код, а невидиме – інформація, – коментує цей вислів Григорія Савича Мирослав Попович, – Суть сказаного Сковородою саме в цьому, таким чином її належить виразити сьогоднішньою мовою» [Попович, 1998, с.286].

Чому етика й моральнісні принципи набувають такої ваги в дослідженнях і створенні ІІІ? Відповідь проста: потрібно бути порядною людиною, знайомою з правилами поведінки у суспільстві, мати розуміння й відповідні моральнісні погляди. І це не просто слова. Це критерії відбору для такої праці.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- Вяткіна Н. Теорія значення, деференція і нормативність. *Філософська думка*. 2019. № 5. С. 40-51. <https://dumka.philosophy.ua/index.php/fd/article/view/382/354>
- Маєвський О. Епістемічна обмеженість великих мовних моделей. *Multiversum. Philosophical almanac*. 2024. Issue 2(180). Vol. 1. С. 54-70. <https://doi.org/10.35423/2078-8142.2024.2.1.3>
- Маєвський, О. Функціональний Успіх Інтелектуальних Автоматів. *Наукові записки НаУКМА. Філософія та релігієзнавство*. 2020. Т. 5. С.15-26. <https://doi.org/10.18523/2617-1678.2020.5.15-25>
- Мироненко Р. Проблема етичного ставлення до ІІІ-моделей: чи потрібно їм дякувати й вітатися з ними? [Електронний ресурс]. *Печера Платона*. 06.05.2025. Режим доступу: <https://platoscave.com.ua/tpost/xh7ys1h9k1-problema-etichnogo-stavlennya-do-sh-mode>
- Попович М. *Напис історії культури України*. Київ: АртЕк, 1998. 728 с.
- Титенко С. *Онтологічно-орієнтовані системи керування контентом інформаційно-навчальних Web-порталів*. Київ: Американ Юніверсіті Київ, 2024. 248 с.
- Aberdein A., Inglis M. Introduction, *Advances in Experimental Philosophy of Logic and Mathematics*. London: *Bloomsbury Academic*. 2019. P. 1-13.
- Bell K. R. W., Noble J., R, Rossiter S. Social simulations: improving interdisciplinary understanding of scientific positioning and validity, *Journal of Artificial Societies and Social Simulation*. 2010. № 13 (1) 10. DOI: 10.18564/jasss.1590 <https://www.jasss.org/13/1/10.html>

Benzmüller C., Parent X., van der Torre L. Designing Normative Theories for Ethical and Legal Reasoning: LogiKEy Framework, Methodology, and Tool Support. *arXiv*. 2019. doi: <http://arxiv.org/abs/1903.10187>

Bornet P. The Power of Random: Why Unconventional Thinking is Your Secret Weapon in the AI Age. *LinkedIn*. 2025. 15. 01. <https://www.linkedin.com/pulse/power-random-why-unconventional-thinking-your-secret-weapon-bornet-94s6e/>

Boyles R., James M. A Case for Machine Ethics in Modeling Human-Level Intelligent Agents. *Kritike*. 2018. 12 (1). P. 182–200. https://www.kritike.org/journal/issue_22/boyles_june2018.pdf

Brewster C., Buckingham S., O'Hara K., Simon Fuller S., Shum S.B., et al. Knowledge Representation with Ontologies: The Present and Future. *IEEE Intelligent Systems*. 2004. 19 (1). P. 72–81. doi:10.1109/MIS.2004.1265889

De Bellis A. F., Umbrello S. A Value-Sensitive Design Approach to Intelligent Agents. *Artificial Intelligence Safety and Security*. 2018. P. 395–410. doi: 10.13140/RG.2.2.17162.77762

Ess Ch. Revolution? What Revolution? Successes and limits of computing technologies in philosophy and religion. *Blackwell Publishers*. 2004. doi: <https://doi.org/10.1002/9780470999875.CH12>

Floridi L. The Ethics of Artificial Intelligence. - *Principles, Challenges, and Opportunities*. Oxford: Oxford University Press, 2023.

Floridi L. (ed.) *The Onlife Manifesto. Being Human in a Hyperconnected Era*. Springer Open, 2015.

Institute of Ethics in AI. Oxford University's Institute for Ethics in AI. Tackling the ethical challenges posed by Artificial Intelligence. Part of @PhilFacOx and @unioxford.

Gerlich M. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*. 2025. 15(1), 6; <https://doi.org/10.3390/soc15010006>

Godler Y., Miller B., Reich Z. Social Epistemology as a New Paradigm for Journalism and Media Studies. *New Media & Society*, 2020. 22(2), 213–229. <https://doi.org/10.1177/1461444819856922>

Gorichanaz T., Furner J., Ma, L., Bawden D., Robinson L., Dixon D., Herold K., Søe S.O., Martens B.V., & Floridi L. Information and design: book symposium on Luciano Floridi's The Logic of Information. *J. Documentation*. 2020. 76, 586–616.

León M. Fuzzy Cognitive Maps as a Bridge between Symbolic and Sub-Symbolic Artificial Intelligence. *International Journal on Cybernetics & Informatics (IJCI)*. 2024. Vol.13, No.4, August. P. 57–75. <https://doi.org/10.5121/ijci.2024.130406>

Müller V. C. Ethics of Artificial Intelligence. In *Anthony Elliott (ed.), The Routledge social science handbook of AI*. London: Routledge, 2021. 20p. <https://core.ac.uk/download/305120702>.

Nag Hammadi library. *Wikipedia. The Free Encyclopedia*. https://en.wikipedia.org/wiki/Nag_Hammadi_library

Páez A. The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds and Machines*. 2019. 29. P. 441–459. <https://link.springer.com/article/10.1007/s11023-019-09502-w>

Rapaport W. J. Philosophy of Computer Science: An Introductory Course. *Teaching Philosophy*. 2005. 28:4. P. 319–341. https://cse.buffalo.edu/~rapaport/Papers/rapaport_phics.pdf

Shampagne M. 2025. <https://www.academia.edu/?u=VBp3kE>

Schrage M., & Kiron D. Philosophy eats AI. *MIT Sloan Management Review*, (Reprint#66311). 2025. 1–19. <https://sloanreview.mit.edu/article/philosophy-eats-ai/>.

Вяткіна Наталія Борисівна

кандидат філософських наук

старший науковий співробітник відділу логіки і методології науки

Інститут філософії імені Г.С. Сковороди НАН України

01001, Київ, вул. Трьохсвятительська, 4

E-mail: nataliia.viatkina@auk.edu.ua

ORCID: <https://orcid.org/0000-0003-4839-1146>

Стаття надійшла до редакції: 24.03.2025

Схвалено до друку: 12.05.2025

COURSE "RESEARCH METHODS AND ANALYSIS IN PHILOSOPHICAL RESEARCH": AI AS A TOOL, CHALLENGE AND MIRROR OF HUMAN FAILURES

Viatkina Natalia B.

PhD in philosophy, Senior Research Associate
Department of Logic and Methodology of Science
H. S. Skovoroda Institute of Philosophy, NAS of Ukraine
4, Triokhsviatytska St., Kyiv, 01001, Ukraine
E-mail: natalia.viatkina@auk.edu.ua
ORCID: <https://orcid.org/0000-0003-4839-1146>

ABSTRACT

The intersection of philosophical research and artificial intelligence (AI) is, at first glance, fertile ground for research, but only at first glance. After all, it is not an easy task to thoroughly delve into the fundamental philosophical principles on which AI developments are based and to realize all the consequences of this relationship fully. The article is devoted to the search for those fundamental principles that may determine the value, effectiveness, and different risks that AI may generate alongside new ideas and experiences. The article exemplifies the role of AI in teaching the course "Research Methods and Analysis in Philosophical Studies" as a study. The author also analyzes the ethical and cognitive problems of AI, using two student research works as illustrations of how the course achievements are applied to consider, in particular, the bias of AI and its limited ability to "judge" compared to humans. Deliberately bypassing the history of AI development, the author describes the closest stage to us when there is a transition to creative processes, especially writing. ChatGPT is used for searching, classifying research objects, generating data, and rewriting with editing, as well as overcoming writer's block, all of which can be accomplished in a matter of minutes. There are also more specialized versions of artificial intelligence for academic work, such as searching and processing literature. True, one also has to deal with the consequences of such "encyclopedism" of AI. The fact is that AI finds information using the work of hundreds of thousands of academics (scientists, teachers), as well as writers. Everyone decides for themselves whether to use such a labor-intensive means of creating texts. However, the mechanism of AI borrowing literary sources requires immediate regulation and constant monitoring of copyright compliance. The article discusses the latest approaches to assessing the role of philosophy, which lays the principles of creating artificial intelligence and is, therefore, a fundamental part of effective AI. A field of AI philosophy is emerging that deepens and illuminates important topics, combining theoretical foundations with practical applications in technology. Furthermore, considering various concepts, such as explainable AI, highlights the need for users to understand, at least in general terms, the meaning of complex models, including information models. The conclusions emphasize the importance of moral responsibility in the development and use of AI, both personal and collective.

Keywords: *AI, AI philosophy, AI ethics, research methods and analysis, AI technologies, large language models (LLM).*

REFERENCES

- Viatkina, N. (2019). Theory of Meaning, Deference and Normativity. *Philosophical thought*, 5, 40-51. <https://dumka.philosophy.ua/index.php/fd/article/view/382/354> (In Ukrainian)
- Mayevsky, A. (2024). Epistemic Limitations of Large Language Models. *Multiversum. Philosophical almanac*, 2(180), 1, 54-70. <https://doi.org/10.35423/2078-8142.2024.2.1.3> (In Ukrainian)
- Mayevsky, A. (2020). The Functional Success of Intelligent Automata. *NaUKMA Research Papers in Philosophy and Religious Studies*, 5, 15-26. <https://doi.org/10.18523/2617-1678.2020.5.15-25> (In Ukrainian)

- Myronenko, R. (2025). The problem of ethical attitude towards AI models: should we thank them and greet them? *Plato's Cave*. 06.05. <https://platoscave.com.ua/tpost/xh7ys1h9k1-problema-etichnogo-stavlennya-do-sh-mode> (In Ukrainian)
- Popovych, M. (1998). *An Essay on the Cultural History of Ukraine*. Kyiv: ArtEk. (In Ukrainian)
- Tytenko, S. (2024). *Ontologically-oriented content management systems for informational and educational Web portals: monograph*. Kyiv: American University Kyiv. (In Ukrainian)
- Aberdein, A., Inglis, M. (2019). Introduction. *Advances in Experimental Philosophy of Logic and Mathematics*. London: Bloomsbury Academic. pp. 1-13.
- Bell, K, R. W., Noble, J., R, Rossiter, S. (2010). Social simulations: improving interdisciplinary understanding of scientific positioning and validity, *Journal of Artificial Societies and Social Simulation*, 13 (1) 10. DOI: 10.18564/jasss.1590 <https://www.jasss.org/13/1/10.html>
- Benzmüller, C., Parent, X.r, van der Torre, L. (2019). Designing Normative Theories for Ethical and Legal Reasoning: LogiKey Framework, Methodology, and Tool Support. *arXiv*. doi: <http://arxiv.org/abs/1903.10187>
- Bornet, P. (2025). The Power of Random: Why Unconventional Thinking is Your Secret Weapon in the AI Age. *LinkedIn*, 15. 01. <https://www.linkedin.com/pulse/power-random-why-unconventional-thinking-your-secret-weapon-bornet-94s6e/>
- Boyles, R., James M. (2018). A Case for Machine Ethics in Modeling Human-Level Intelligent Agents. *Kritike*, 12 (1), 182–200. https://www.kritike.org/journal/issue_22/boyles_june2018.pdf
- Brewster, C., Buckingham S., O'Hara, K., Simon, Fuller, S., Shum, S.B., et al. (2004). Knowledge Representation with Ontologies: The Present and Future. *IEEE Intelligent Systems*, 19 (1), 72-81. doi:10.1109/MIS.2004.1265889.
- De Bellis, A. F., Umbrello, S. (2018). "A Value-Sensitive Design Approach to Intelligent Agents" *Artificial Intelligence Safety and Security*. pp. 395-410. doi: 10.13140/RG.2.2.17162.77762
- Ess, Ch. (2004). Revolution? What Revolution? Successes and limits of computing technologies in philosophy and religion. *Blackwell Publishers*, doi: <https://doi.org/10.1002/9780470999875.CH12>
- Floridi, L. (2023). *The Ethics of Artificial Intelligence - Principles, Challenges, and Opportunities*. Oxford: Oxford University Press.
- Floridi, L. (ed.) (2015). *The Onlife Manifesto. Being Human in a Hyperconnected Era*. Springer Open ISBN 978-3-319-04093-6 (eBook).
- Institute of Ethics in AI. Oxford University's Institute for Ethics in AI. Tackling the ethical challenges posed by Artificial Intelligence. Part of @PhilFacOx and @unioxford.
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 2025, 15(1), 6; <https://doi.org/10.3390/soc15010006>
- Godler, Y., Miller, B., Reich, Z. (2020). Social Epistemology as a New Paradigm for Journalism and Media Studies. *New Media & Society*, 22(2), 213–229. <https://doi.org/10.1177/1461444819856922>
- Gorichanaz, T., Furner, J., Ma, L., Bawden, D., Robinson, L., Dixon, D., Herold, K., Søre, S.O., Martens, B.V., & Floridi, L. (2020). Information and design: book symposium on Luciano Floridi's *The Logic of Information*. *J. Documentation*, 76, 586-616.
- León, M. (2024). Fuzzy Cognitive Maps as a Bridge between Symbolic and Sub-Symbolic Artificial Intelligence. *International Journal on Cybernetics & Informatics (IJCI)*, 13, 4, August, 57-75. <https://doi.org/10.5121/ijci.2024.130406>
- Müller, V. C. (2021). Ethics of Artificial Intelligence". *Anthony Elliott (ed.), The Routledge social science handbook of AI*. London: Routledge <https://core.ac.uk/download/305120702>.
- Nag Hammadi library. *Wikipedia. The Free Encyclopedia*. https://en.wikipedia.org/wiki/Nag_Hammadi_library
- Páez, A. (2019). The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds and Machines*, 29, 441–459. <https://link.springer.com/article/10.1007/s11023-019-09502-w>

- Rapaport, W. J. (2005). Philosophy of Computer Science: An Introductory Course. *Teaching Philosophy*, 28:4, 319-341. https://cse.buffalo.edu/~rapaport/Papers/rapaport_phics.pdf
- Shampagne, M. (2025). <https://www.academia.edu/?u=VBp3kE>
- Schrage, M., & Kiron, D. (2025). Philosophy eats AI. *MIT Sloan Management Review*, (Reprint#66311), 1–19. <https://sloanreview.mit.edu/article/philosophy-eats-ai/>

Article arrived: 24.03.2025

Accepted: 12.05.2025

Як цитувати: Вяткіна, Н. (2025). КУРС "ДОСЛІДНИЦЬКІ МЕТОДИ Й АНАЛІЗ У ФІЛОСОФСЬКИХ ДОСЛІДЖЕННЯХ" ІІІ ЯК ІНСТРУМЕНТ, ВИКЛИК І ДЗЕРКАЛО ЛЮДСЬКИХ НЕДОЛІКІВ. *Вісник Харківського національного університету імені В. Н. Каразіна. Серія «Філософія. Філософські перипетії»*, (72), 107-116. <https://doi.org/10.26565/2226-0994-2025-72-10>

In cites: Viatkina, N. (2025). COURSE "RESEARCH METHODS AND ANALYSIS IN PHILOSOPHICAL RESEARCH" AI AS A TOOL, CHALLENGE AND MIRROR OF HUMAN FAILURES. *The Journal of V. N. Karazin Kharkiv National University, Series Philosophy. Philosophical Peripeteias*, (72), 107-116. <https://doi.org/10.26565/2226-0994-2025-72-10>
[In Ukrainian]