

# ШТУЧНИЙ ІНТЕЛЕКТ І АКАДЕМІЧНА ДОБРОЧЕСНІСТЬ В КОНТЕКСТІ ЕПІСТЕМОЛОГІЇ ЧЕСНОТ

Надія Козаченко

**Анотація.** Мета статті – розглянути особливості академічної доброчесності в контексті зростаючого використання штучного інтелекту в дослідженнях. Традиційна епістемологія не враховує специфіку цього питання, тому необхідно звернутися до епістемології чеснот. Представники цього підходу вважають, що знання є результатом правильного використання епістемічних чеснот, таких як відкритість до істини, уважність та інтелектуальна чесність тощо. Таким чином, доброчесне знання виникає, коли особа використовує інтелектуальні чесноти для досягнення істини. Особливими епістемічними чеснотами, що під загрозою через використання ШІ, є епістемічна автономія та відповідальність. Автономія дозволяє суб'єкту самостійно підтримувати власні переконання, а відповідальність несе епістемічний суб'єкт щодо достовірності своїх переконань. Разом з чеснотами ми розглядаємо й епістемічні вади. Ці характеристики знання допомагають нам сформулювати поняття епістемічного суб'єкта в контексті епістемології чеснот. Грунтуючись на понятті епістемічного суб'єкта ми розглядаємо епістемічний авторитет, його функції та проблеми, які виникають при взаємодії з ним. Чи втрачаємо ми епістемічну автономію, коли довіряємо авторитету? Чи є раціональною віра на основі авторитету? Чи несе людина відповідальність за переконання, набуті через авторитет, особливо якщо це має негативні наслідки? На основі дослідження ми відповідаємо на питання: чи володіє ШІ епістемічними чеснотами і чи може бути епістемічним авторитетом. Основною проблемою постає те, що ШІ не є самостійним епістемічним суб'єктом і не може відповідати за інформацію чи аналіз, як це робить людина. Він не здатний обґрунтовувати свої висновки через брак чеснот. Знання ШІ залежать від даних і методів навчання, тому він не може оцінити їх достовірність або етичні наслідки. Проте, за певних умов ШІ може демонструвати епістемічні чесноти, але це вимагає постійної взаємодії з людиною. Фінальний розділ статті розкриває можливість академічно доброчесної взаємодії зі штучним інтелектом з урахуванням особливостей його функціонування з точки зору епістемології чеснот в реальних навчальних та дослідницьких кейсах.

**Ключові слова:** академічна доброчесність, штучний інтелект, епістемологія чеснот, епістемічний авторитет, епістемічна автономія, епістемічна відповідальність.

## Що таке епістемологія чеснот?

Епістемологія чеснот – напрямок епістемології, що не просто розглядає роль чеснот у пізнанні, але й досліджує самі епістемічні чесноти, вважаючи їх важливими для набуття знання. Прихильники епістемології чеснот розглядають епістемічні чесноти як фактори, які є не менш визначальними для функціонування знання, порівняно з його обґрунтованістю та істинністю.

Варто зазначити, що не існує єдиного уніфікованого переліку епістемічних чеснот, але ті чесноти, які дослідники зазвичай позначають різними назвами, по суті охоплюють схожий між собою набір якостей. Так, йдеться про досить поширені чесноти, які інтерпретуються в контексті пізнавальної діяльності людини, спрямованої на отримання якісних знань. Це, наприклад, такі як: *інтелектуальна чесність* – здатність відкрито визнавати свої помилки та шукати істину, навіть якщо це суперечить власним переконанням; *допитливість* – прагнення до знань, бажання ставити питання і шукати відповіді на складні проблеми; *уважність* – здатність зосереджуватись на важливих деталях і не пропускати значущі аспекти під час пізнавального процесу; *скромність* – усвідомлення власних меж знань і готовність визнати, що не все можна знати або розуміти; *інтелектуальна сміливість* – здатність висловлювати та захищати власні переконання, навіть коли це складно або непопулярно; *інтелектуальна стійкість* – здатність продовжувати пошук знання і не здаватися, коли стикаєшся з труднощами чи невизначеністю; *інтелектуальна незалежність* – здатність формувати власні судження без сліпого слідування за думками інших; *інтелектуальна обережність* – схильність ретельно оцінювати інформацію, перш ніж приймати рішення або робити висновки; *інтелектуальна справедливість* – здатність об'єктивно оцінювати аргументи і позиції незалежно від власних уподобань та інші подібні якості, які ми ще розглянемо в нашому дослідженні.

Епістемологія чеснот має витоки в роботах знаних методологів науки, таких, наприклад, як Ернест Соса [12], і від початку розробки подібні напрямки стосувалися можливості подолання відомих епістемологічних проблем – «трилеми Мюнхаузена», «проблеми Гетґе», вирішення дискусії між фундаменталістами та когерентистами. Я не буду стверджувати, що епістемології чеснот вдалося розв'язати ці або інші епістемологічні проблеми,<sup>1</sup> але, тим не менш, немає сенсу запе-

<sup>1</sup> Детальніше див: <https://plato.stanford.edu/entries/epistemology-virtue/>

речувати, що питання, які підіймаються в цій царині, є надзвичайно важливими для дослідників функціонування знання та для нашого дослідження зокрема.

Загалом розглядають два підходи в епістемології чеснот: теорія добродітності та анти-теорія добродітності. *Теорія добродітності* (virtue theory) схиляється до думки, що добродітна людина може мати знання як обґрунтовані переконання, які базуються на її інтелектуальних чеснотах, або принаймні пов'язані з ними. Тобто, інтелектуальні чесноти розглядаються як повноправний, а іноді й визначний компонент знання, невід'ємно пов'язаний з ним. Разом з тим розглядається також *анти-теорія добродітності* (virtue anti-theory).<sup>2</sup> Цей підхід не намагається систематично пов'язати інтелектуальні чесноти зі знанням або, тим паче, з обґрунтуванням чи істиною. Натомість він стверджує, що інтелектуальні чесноти заслуговують на дослідження самі по собі, незалежно від того, чи можна їх систематично поєднати з аналізом знання.

Для нашого дослідження більш важливим є перший підхід, оскільки він безпосередньо пов'язаний з набуттям та функціонуванням знання. Яскравий представник цього підходу – Лінда Загзебські – замість того, щоб розглядати знання як продукт виключно правильного обґрунтування чи доказів, наголошує на моральних та інтелектуальних якостях, які особа повинна мати для досягнення знання. На її думку, знання — це не просто результат дотримання процедур чи алгоритмів, а результат правильного використання інтелектуальних чеснот, таких як відкритість до істини, уважність, інтелектуальна чесність тощо [18, с. 167-169]. Вона вважає, що інтелектуально добродітна людина прагне до істини не лише для того, щоб підтвердити свої переконання, але й для того, щоб відкрити нові знання і позбутися помилкових уявлень. Загзебські визначає кілька основних інтелектуальних чеснот, серед яких: відкритість до нових фактів, готовність змінювати свою думку під впливом доказів, інтелектуальна уважність, витривалість у пошуку істини, а також здатність приймати критику [18, с. 156-160]. Вона стверджує, що знання виникає, коли особа втілює інтелектуальну чесноту в своїй діяльності з метою досягнення істини. Коли людина прагне до істини через уважне дослідження фактів і доведення, вона здійснює акт втілення інтелектуальної чесноти, і якщо це дослідження приводить до істини, тоді це переконання є знанням [там само, с. 272].

---

<sup>2</sup>Терміни походять від Louden R.B. Virtue ethics and anti-theory. *Philosophia*. 1990. Vol. 20. P. 93–114, але інтерпретовані щодо епістемології.

В епістемології чеснот існує два підходи до визначення типів епістемічних чеснот: *респонсибілісти* (responsibilist virtues – «відповідальнісні чесноти») відносять до епістемічних чеснот саме чесноти аристотелівського типу – набуті. Це можуть бути відкритість до нового або інтелектуальна мужність, які є рисами характеру, що вимагають розвитку, але не могли бути вродженими, а отже ми відповідальні за них. На противагу цьому підходу *реліабілісти* (від англ. reliabilist virtues – «надійнісні чесноти») вважають за епістемічні чесноти природні когнітивні здібності, як-от гострий зір, уважність чи пам'ять, які можуть вважатися надійними, а отже на них можна покласти у пізнанні (див.: [7]). Однак, в контексті нашого дослідження, я віддаю перевагу тій позиції, згідно якої всі згадані і подібні чесноти можуть бути розвинені, вдосконалені або обгрунтовано замінені аналогічними можливостями (див.: [там само]).

Загалом чеснотами називають ті якості об'єкта, які дозволяють йому успішно виконувати свої функції (як гострота для ножа чи швидкодія для комп'ютера). Оскільки основною епістемічною функцією людини є набуття знання, епістемічні чесноти – це ті характеристики, які дозволяють нам це робити, при чому не важливо чи є вони вродженими, чи набутими [4, с. 644]. Так, зір є природною властивістю, а якісна інтерпретація МРТ-знімків – набута властивість; навички, отримані у ході навчання логіки чи критичного мислення теж будуть похідними перевагами. Таким чином, чесноти можна набути [там само, с. 645]. І до таких набутих чеснот відносяться мотивація здобувати знання і піклуватися про істинність, епістемічна компетентність та автономія.

**Епістемічна автономія** – здатність суб'єкта самостійно підтримувати власні переконання у стані найбільшої достовірності. Автономія вимагає певного рівня епістемічної компетентності та вище згаданих чеснот – мотивації здобувати знання та піклуватися про його істинність. Епістемічна автономія передбачає також **епістемічну відповідальність** – відповідальність суб'єкта за рівень достовірності своїх переконань.

Беріт Брогаард підкреслює, що знання потребує інтелектуальної чесноти, яка включає прагнення до істини. Вона аргументує, що епістемічна відповідальність передбачає наявність не лише когнітивних здібностей, а й значної мотивації для пошуку істини, що робить цей

процес не просто механічним застосуванням знань, а усвідомленою інтелектуальною практикою. Брогаард також зазначає, що відповідальність в епістемічному плані є іншою, ніж просто причинно-наслідкове приписування собі знань внаслідок його аргументованого набуття: не можна просто взяти і взнати щось, оскільки саме суб'єкт несе відповідальність за власні переконання і повинен використовувати свої інтелектуальні чесноти для критичної оцінки джерел знання і підходити до його здобуття. Не менш важливими є чесноти, які підтримують відповідальну епістемічну позицію: критичне мислення, чесність, відкрите мислення та уважність до контексту [3].

Епістемологія чеснот також актуальна тим, що вона робить акцент на практичній стороні знання, тобто на тому, як ми повинні поводитися у пізнавальному середовищі, щоб досягати істини і уникати помилок. Вона зосереджується не тільки на абстрактних теоріях знання, але й на тому, як чесноти допомагають людям у реальному житті приймати раціональні рішення та розрізняти істинне від хибного.

## **Чим відрізняється епістемологія чеснот від традиційної епістемології переконання?**

У традиційній епістемології основними об'єктами аналізу зазвичай є переконання, тобто, саме зміст того, що ми вважаємо знанням. Так, ми приписуємо знання і переконання носію, але цей «раціональний суб'єкт» виглядає настільки раціоналізованим, що здається спустошеним на тлі фундаментальних вимог обґрунтованості, істинності, несуперечливості та інших характеристик, які висувують знанням та переконанням. Натомість у епістемології чеснот об'єктом аналізу й оцінки є самі епістемічні суб'єкти, відтак саме їхні інтелектуальні чесноти та вади стають центральними поняттями (див.: [4, с. 640]). Саме ці епістемічні чесноти та вади, з точки зору прихильників епістемології чеснот, стають оцінками агентів на основі їхніх інтелектуальних характеристик, що, з точки зору епістемології чеснот, є такими самим значущими для пізнання, як і прагматична або трансцендентальна виправданість переконання чи знання, а іноді й більш фундаментальними.

Одразу варто зробити ремарку: більш фундаментальними – означає не такими, що заміщують, а такими, що дають змогу здійснити, забезпечують підставу. Такі заяви ні в якому разі не мають давати привід для підозрілого редукціонізму: чи часом не мають на меті епістемологи

чеснот звести всю істинність та обґрунтованість до суб'єктивності? Ні, йдеться не про це. Йдеться про те, що спосіб, мета і передумови набуття знання можуть бути так само важливими (а іноді й більш важливими), аніж його раціональні характеристики. Таким чином, в епістемології чеснот важливим є не тільки отримане знання, але й шлях його отримання і якості суб'єкта пізнання, за допомогою яких було отримане це знання.

Ключові характеристики епістемічного суб'єкта проявляються також і через його суб'єктивну сферу, оскільки суто раціональний вимір не є достатнім для інтерпретації динаміки знання-незнання та їхніх модифікацій. Зазначимо, що по-перше, епістемічним суб'єктом може бути як окрема людина, яка аналізує свої знання, так і спільнота, наприклад, наукові спільноти, що діють як епістемічні суб'єкти, критично оцінюючи і приймаючи або відхиляючи теорії на основі доказів і спільних парадигм. Зважаючи на те, що Нобелівські премії останнім часом дають науковим колективам, в цьому сенсі ми можемо цілком говорити про інтерсуб'єктивність як характеристику епістемічного суб'єкта. По-друге, епістемічний суб'єкт не просто накопичує знання, а й відповідає за активність і критичність власного мислення, оцінку джерел знання і обов'язковість застосування інтелектуальних чеснот. Дійсно, для того, щоб вважатися епістемічним суб'єктом, особа має приймати обґрунтовані рішення на основі доступних їй елементів знання, а не діяти пасивно. Це включає усвідомлений вибір позицій і активну участь у процесі формування знань як індивідуально, так і в групах.

Спробуємо надати означення епістемічному суб'єкту, спираючися на подібні дослідження.

**Епістемічний суб'єкт** — це індивід або спільнота, здатні формувати, оцінювати й переглядати свої переконання на основі доступних знань та раціональних міркувань. Такі агенти мають здатність усвідомлено оцінювати, приймати або відхиляти певні знання, спираючись на свої інтелектуальні чесноти, раціональні міркування та розуміння доступних альтернатив. Вони також мають епістемічну відповідальність за процес здобуття знань і використання інтелектуальних чеснот [2, 10].

Згідно з таким означенням стає зрозумілим, що переорієнтація значної частини філософів науки від епістемології знання до епістемології

чеснот зумовлена низкою актуальних проблем і обмежень, що постали перед традиційною епістемологією, яка фокусується переважно на *ЈТВ*<sup>3</sup> знанні. Цей перехід виникає, зокрема, у відповідь на деякі питання. Перш за все, це – обмеження традиційної епістемології у поясненні ролі епістемічного суб'єкта. Так, традиційна епістемологія знання робить акцент на окремих переконаннях і їх обґрунтуванні, однак питання, яким чином сам епістемічний суб'єкт сприяє процесу отримання знання, залишається не висвітленим. Так само, підхід, який фокусується виключно на об'єктивованому *ЈТВ* знанні, не може пояснити важливість інтелектуальних чеснот, таких як відкритість, критичне мислення, уважність до деталей чи інтелектуальна сміливість. Епістемологія чеснот прагне подолати це обмеження, включивши інтелектуальні чесноти у процес оцінки знання. Вона підкреслює, що не тільки результати мислення (віра, переконання чи знання) мають значення, але й сам процес, який вимагає участі розвинутих інтелектуальних якостей епістемічного суб'єкта. Питання виникають тоді, коли стає зрозумілим, що традиційні моделі епістемології часто зосереджені на об'єктивних критеріях обґрунтування, але при цьому ігнорується факт, коли переконання ґрунтується на надійних або добродесних когнітивних здібностях суб'єкта. Епістемологія чеснот вирішує цю проблему, наголошуючи на інтелектуальній надійності суб'єкта, що допомагає йому постійно досягати істини через чесноти або природні здібності, такі як пам'ять або сприйняття (або ж їхні посилювачі).

Так, виникає питання: чи не замінили ми питання щодо гіперраціонального суб'єкта гіпердобродесним суб'єктом? Якщо в контексті епістемології чеснот ми починаємо зосереджувати увагу на інтелектуальних чеснотах суб'єкта, таких як чесність, відкритість до критики, уважність та інших якостях, що сприяють пізнанню, то суб'єкт у такій епістемології не просто раціональний, але і добродесний у своїй діяльності. Такий суб'єкт не просто приймає раціональні рішення, а й втілює ідеальні інтелектуальні чесноти, максимально добродесно керуючи своїм пізнавальним процесом. Якщо у традиційній епістемології суб'єкт керується лише раціональними принципами, то в епістемології чеснот він керується не тільки раціональністю, а й етикою пізнання. Це особливо яскраво просвічує у сучасності, де знання стає дедалі складнішим і все більш фрагментованим. Зважаючи на різномані-

---

<sup>3</sup> *ЈТВ* – *justified true belief* – означення знання, що йде ще від Платона, але постійно піддається сумніву, оскільки має низку контраргументів, включно зі сформульованими самим Платоном.

тність інформаційних джерел, епістемологія знання не може завжди надавати достатньо інструментів для оцінки якості переконань, оскільки вона застосовує досить формальний апарат, нездатний охопити динаміку і різноманітність. Натомість епістемологія чеснот забезпечує етичний та пізнавальний фундамент для критичного мислення та інтелектуальної доброчесності, що допомагає уникати релятивізму і сприяє формуванню надійних переконань.

По суті, епістемологія чеснот, або як її ще називають – епістемологія доброчесності, спирається на інтелектуальні чесноти епістемічного суб'єкта й надає йому повний карт-бланш, покладаючись на його розумність. Чи можна уявити собі такого суб'єкта? Так, уявити цілком можна і його ідеальні характеристики будуть цілком співмірні з характеристиками раціонального суб'єкта пізнання. То в чому ж справа, питаєте ви, – підміна однієї нереальної абстракції іншою? – скоріше, доповнення і наголошення на тому, що філософський аналіз ідеалізованих об'єктів дає ту асимптоту, до якої ми маємо прагнути, навіть розуміючи, що ніколи не досягнемо її. Це контрфактичний світ, але це той світ, в якому ми можемо сягнути істини.

## **Як функціонує епістемічний авторитет в контексті епістемології чеснот?**

Формування переконань на основі авторитету є неминучим та виправданим в багатьох обставинах. Одним із ключових факторів, чому ми змушені довіряти авторитетам, є наша природна обмеженість у знаннях і можливостях. Люди не здатні осягнути всі аспекти реальності самостійно, тому часто покладаються на знання та досвід інших, зокрема експертів у певних галузях. Наша епістемічна самодостатність є важливою, але очевидно не достатньою для формування істинних переконань. В умовах сучасного світу, де інформація стає дедалі складнішою, ми все частіше потребуємо епістемічних авторитетів – людей або інституцій, яким можна довіряти, оскільки їхня компетентність перевершує нашу. Наприклад, ми довіряємо лікарям у питаннях здоров'я, науковцям у питаннях науки і навіть побутовим експертам у повсякденних справах. Така довіра до авторитетів впливає з нашого раціонального бажання отримати істину і діяти на її основі. Більш того, наша довіра до інших людей в питаннях знання є раціональною та необхідною. Люди часто покладаються на експертів, щоб отримувати знання з тем, у яких вони не є фахівцями, оскільки область доступного



знання сьогодні є настільки широкою, що охопити поодинокі всі її сфери просто неможливо.

У ході такої довіри негативну роль можуть відіграти епістемічні вади, що протиставляються епістемічним чеснотам або виникають внаслідок їхньої нереалізованості. Перш за все, це *інтелектуальні лінощі* – небажання або невміння інвестувати час та зусилля у ретельне пізнання або критичне мислення [18, с. 140-142]. Сюди ж можна віднести *інтелектуальну недбалість* – нехтування збором або оцінкою відповідних доказів та аргументів у процесі прийняття рішень та поспішність у висновках – здійснення висновків на основі недостатньої кількості інформації або доказів. Варто згадати і про *самовпевненість*, що постає як переоцінка своїх знань або переконань, що заважає об'єктивному аналізу, і також *інтелектуальну заздрість* – відмову визнати або схвалити когось, хто демонструє вищий інтелект або знання, через власні інтелектуальні комплекси [4, с. 639-644]. Згадаємо і про *цинізм*, або *надмірний скепсис* – відмову від визнання будь яких переконань значущими або знання істинним.

Ці вади тісно пов'язані з функціонуванням епістемічного авторитету. Інтелектуальні лінощі та недбалість призводять до некритичного прийняття авторитетних свідчень, або ж і просто дають суб'єкту привід приймати у якості авторитетного будь-яке свідчення. З іншого боку, інтелектуальна самовпевненість, заздрість та цинізм відкидають епістемічний авторитет взагалі, не зважаючи на інші переконання, з яких би авторитетних джерел вони не надходили. Вочевидь і тут потрібно шукати аристотелівську усвідомлену середину між крайнощами.

На буденному рівні можна визначити епістемічний авторитет як такий, на кого можна покластися у формуванні своїх переконань, у набутті знання, – вчитель, який пояснить, а до того ж ще й зрозуміє наявні пробіли в знаннях, скерує та підкаже. Тим не менш, основна епістемічна чеснота, яка підважує статус епістемічного авторитету – це наша пізнавальна автономія, яку з легкістю можна втратити, некритично покладаючись на авторитет. Епістемічна автономія вимагає від людини самостійно оцінювати підстави, якими вона керується для набуття та підтримання переконань, розглядати їхні підґрунтя й наслідки та формувати переконання на основі свідомого рішення. Тому проблема взаємодії автономії та авторитету передбачає питання: як ми можемо довіряти авторитетам, не підриваючи свою автономію? Чи є сенс відкинути всі авторитети і думати власним розумом, плекаючи власну епістемічну автономію? Чи можливо ми можемо стверджувати, що довіра до авторитетів не обов'язково суперечить

епістемічній автономії, якщо ця довіра ґрунтується на раціональній оцінці компетентності авторитету.

Відтак, є сенс більш детально проаналізувати, що таке епістемічний авторитет, як він функціонує і які проблеми породжує. Одразу варто зазначити, що сам термін так і не отримав на сьогодні точного означення і зазвичай визначається через перелік функцій та характеристик. Так, у роботі *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* Лінда Загзебськи зазначає кілька способів функціонування терміну «авторитет» (див.: [16, с. 99-114]).

*Авторитет як джерело знання.* Авторитетність у цьому сенсі вказує на те, що ми можемо довіряти певній особі або інституції як надійному джерелу знань. Це означає, що їхні переконання, думки або твердження можуть бути прийняті нами без необхідності самостійно перевіряти всі деталі. Наприклад, ми можемо вважати вченого авторитетом у певній науковій галузі та приймати його думку як істинну через його компетентність.

*Авторитет як джерело нормативних вказівок.* У цьому випадку авторитет виступає як нормативна сила, яка створює обов'язок або підставу для прийняття певної віри або дії. Тут авторитет функціонує аналогічно до політичної влади, коли люди виконують вказівки, навіть якщо не розуміють або не згодні з їхньою суттю, але приймають їх, оскільки ці вказівки походять від авторитетної особи або інституції.

*Авторитет як практична сила.* Також авторитет може виступати як практична сила, що формує наші дії або переконання через інструкції або накази. Це не означає, що авторитет має право примушувати до переконання або дії, але його вплив є достатньо сильним, щоб ми змінили свою позицію, покладаючись на його компетентність.

*Авторитет як заміщення.* Цей, можливо, найважливіший спосіб вживання поняття «авторитет» стосується ситуацій, коли авторитет заміщує наші власні причини для прийняття певних рішень або переконань. Це означає, що ми можемо довіряти авторитету настільки, що його переконання переважають наші власні судження. Загзебськи називає це авторитетним заміщенням, коли авторитетні переконання замінюють наші власні підстави для того чи того переконання.

*Теза авторитетного заміщення.* Той факт, що авторитет має переконання  $p$ , є підставою для мене вірити в  $p$ , яка замінює мої інші підстави вірити в  $p$ , а не просто додається до них [16, с. 113].

Це означає, що, приймаючи переконання на основі авторитету, ми

відкидаємо свої власні аргументи або сумніви. Проблема тут полягає в тому, що ми можемо втратити критичний підхід і перестати аналізувати інформацію самостійно, якщо авторитетний голос постійно замінює наші власні причини для віри. Це також досить складно в ситуаціях, коли наше особисте переконання вже має вагомі причини, але ми визнаємо, що інша людина є кращим експертом у певній галузі. Тим не менш, Загзєбськи стверджує, що ми все ж маємо керуватися цією заміщувальною силою авторитету, оскільки епістемічний авторитет зазвичай має більш обґрунтовані причини для своїх переконань.

Цікаве проблемне питання стосується того, чи в принципі можна утворити власні переконання на основі наказу або примусу. Загзєбськи ставить під сумнів думку Лока про неможливість примусової віри, зазначаючи, що, хоча неможливо примусити когось вірити через силу, можна переконати людей довіряти авторитетам через їхню компетентність (див.: [16, с. 100]). Проблема ще більш яскраво проявляється, коли виникає конфлікт між тим, у що нам наказують вірити, і тим, що ми вважаємо правдою.

Також має місце питання про те, чи є віра на основі авторитету раціонально обґрунтованою. Деякі філософи стверджують, що прийняття переконань на основі авторитету є менш раціональним, оскільки в такому випадку ми не самі робимо висновки, а покладаємося на інших. Проблема полягає в тому, як оцінити, коли довіра до авторитету є раціонально обґрунтованою, а коли це лише сліпа віра.

Загзєбськи також піднімає питання про моральну відповідальність за наші переконання. Віру в неправдиві або недостовірні переконання можна вважати морально проблематичною, якщо вона завдає шкоди іншим людям. Проблема полягає в тому, як оцінювати, чи несе особа відповідальність за те, що вона вірить на основі авторитету, якщо це призводить до негативних наслідків.

Загзєбськи порівнює епістемічний авторитет з політичним. На її думку – це два подібних типи влади, між якими існує схожість у механізмах впливу на особу та її рішення. Основна підстава цього порівняння полягає в тому, що обидва типи авторитету включають визнання іншого суб'єкта або інституції як джерела обов'язкових вказівок чи переконань. У випадку політичного авторитету це можуть бути закони чи правила, які особа повинна виконувати, а у випадку епістемічного авторитету – це знання чи переконання, які людина приймає на віру від експерта. Загзєбськи порівнює ці два види авторитету, щоб показати, що авторитетність не обов'язково суперечить автономії. Вона аргументує, що в обох випадках можливо діяти або

вірити на основі авторитету, не втрачаючи здатності до самостійного мислення. Це аналогічно тому, як політична влада не обов'язково обмежує свободу, якщо вона ґрунтується на законному авторитеті. Так само епістемічна влада може бути виправданою і корисною для отримання істини.

Наслідуючи Загзебськи, також звернемося до робіт Джозефа Раза, який розробив загальні принципи влади, які можна застосувати і до епістемічної сфери. Розглянемо кілька ключових ознак авторитету в його роботах, зокрема в книзі *The Authority of Law: Essays on Law and Morality* (1979) та інших текстах. Ось основні критерії авторитету, які він пропонує.

1. *Принцип нормативності*. Авторитет повинен мати силу видавати директиви, які впливають на поведінку людей. Ці директиви мають бути такими, що їхні адресати визнають їх як підставу для дії, навіть якщо вони не погоджуються з ними. Аналогічним проявом для епістемічного авторитету буде *Теза авторитетного заміщення*, оскільки переконання авторитетної особи заміщує наші власні причини для віри. Це означає, що якщо ми визнаємо когось епістемічним авторитетом, ми приймаємо його переконання, навіть якщо воно відрізняється від нашого, оскільки його причини для віри є більш переконливими, ніж наші [11, с. 10]. Авторитетні накази виключають інші практичні причини, які суб'єкт міг би мати для дії. Тобто, коли авторитет видає наказ, це означає, що інші причини для дії стають неактуальними, доки існує наказ авторитету. Так само, неавторитетні свідчення чи переконання не будуть мати заміщувальної сили порівняно з авторитетними.

2. *Принцип служіння* (service conception). З точки зору влади авторитет має служити інтересам своїх підданих. Зокрема, рішення, які видає авторитет, повинні бути кращими, ніж рішення, які піддані могли б ухвалити самостійно, враховуючи їхні власні обмеження в знаннях чи розумінні [11, с. 12-15]. На нашу думку, це основна функція епістемічного авторитету і вона, по-перше, обґрунтовує доцільність звернення до епістемічного авторитета, а по-друге, визначає критерій епістемічної авторитетності: якщо переконання, які підтримує деякий епістемічний суб'єкт, є «кращими» – у сенсі: більш обґрунтованими, достовірними, узгодженими, то даний суб'єкт може бути визнаний епістемічним авторитетом у відповідній царині і може відбутися заміщення.

3. *Принцип легітимності*. Для того щоб бути легітимним, авторитет повинен базуватися на певних моральних або правових засадах, що виправдовують його право надавати накази та вимагати їхнього

виконання [11, с. 22]. Що ж має слугувати аналогічним легітиматором для епістемічного авторитету? Відповідь на це питання цілком належить до царини епістемології чеснот і дає можливість сформулювати означення епістемічного авторитету в контексті теорії чесноти. Дійсно, для того, щоб бути епістемічно легітимним, авторитет має бути доброчесним.

**Епістемічний авторитет** – це епістемічний суб'єкт, що володіє знанням високого рівня достовірності у певній галузі, які були ним набуті доброчесно – тобто, із максимальним задіянням епістемічних чеснот і мінімальними епістемічними вадами.

Функції епістемічного авторитету і принципи політичного авторитету показують, що авторитет «працює» на знання: навіть якщо він не може забезпечити стовідсоткової достовірності своїх висновків, завдяки втіленню низки епістемічних чеснот він пропонує кращий варіант переконання, вартий того, щоб замінити ним переконання менш авторитетного суб'єкта. Такий підхід є центральним для «службового» підходу до авторитету, розробленого Разом, який акцентує увагу на функціонуванні авторитету через надання кращих рішень та підстав для дій.

Принцип легітимації авторитету через його доброчесність дає змогу людині не втрачати свою епістемічну автономію через звернення до авторитету. Дійсно, автономія зберігається тоді, коли людина приймає рішення довіряти певному авторитету на основі свідомого вибору через аналіз його надійності. Ми цілком можемо покладатися на інших у процесі формування переконань, не жертвуючи при цьому своєю самостійністю, оскільки саме рішення про довіру є проявом автономії. Таким чином, епістемічна автономія у контексті довіри до авторитету означає здатність приймати рішення про те, кому і чому довіряти, зберігаючи контроль над власними переконаннями.

Наша довіра до авторитетів не суперечить нашій епістемічній автономії, якщо ми усвідомлено вибираємо, кому довіряти. Можна припустити, що автономія не втрачається й тоді, коли ми вирішуємо довіряти іншим на основі раціональних міркувань, тобто коли вважаємо їх компетентними у певній сфері. Ми не жертвуємо своєю незалежністю, оскільки рішення довіряти авторитетам базується на нашій раціональній оцінці їхньої експертизи. Більш того, віра в авторитети впливає з наших епістемічних зобов'язань перед собою. Ми раціонально зобов'язані шукати істинні підтверджені переконання і діяти на їх основі.

Якщо ми визнаємо, що певна особа або інституція володіє кращими знаннями або методами для отримання істини, ніж ми самі, тоді довіра до них стає частиною нашого власного прагнення до знання. Вміння взаємодіяти з епістемічним авторитетом може розглядатися як ще одна епістемічна чеснота. Визнання авторитетності інших дослідників – необхідна вимога провадження наукових досліджень, зафіксована в Європейському кодексі дослідницької доброчесності: «Усі партнери співробітництва в галузі досліджень несуть відповідальність за сумлінне проведення наукового дослідження та його результати», «дослідники враховують останні досягнення у відповідних галузях розробки ідей досліджень» й разом з тим «дослідники повідомляють про свої результати та методи, включаючи використання зовнішніх послуг або штучного інтелекту та автоматизованих інструментів, з урахуванням норм, прийнятих у цій дослідній дисципліні, та полегшуючи верифікацію чи реплікацію, за наявності необхідності» [15].

### **Чи володіє штучний інтелект епістемічними чеснотами?**

Епістемологія чеснот дає змогу підійти з іншого боку до питання аналізу епістемічних властивостей штучного інтелекту. Загалом, досить багато досліджень присвячено моделюванню знань на основі епістемології знання і ми без сумніву можемо стверджувати, що ШІ вже давно проходять Тест Тьюринга. Але чи володіє штучний інтелект епістемічними чеснотами?

Штучний інтелект, в його сучасній формі, поки що не володіє епістемічними чеснотами в такому розумінні, як їх визначає епістемологія чеснот. Епістемічні чесноти — це не лише когнітивні здібності, а й моральні якості, такі як інтелектуальна чесність, допитливість, уважність, скромність і критичне мислення. Вони вимагають здатності до свідомого вибору, оцінки, саморефлексії та бажання досягати істини. ШІ, навіть найбільш просунуті моделі, не можуть мати цих епістемічних чеснот. По суті, навіть найпопулярніші та найрозвиненіші ШІ є лише великими мовними моделями, тобто представляють собою маніпулятор, який не має внутрішньої мотивації або свідомості, що є основою чеснот. Чесноти передбачають готовність до дії на основі моральних або інтелектуальних установок, що походять з особистої відповідальності та рефлексії, в той час як ШІ лише виконує задані алгоритми на основі даних і програмованих критеріїв. Відповідно,

навіть якщо він може імітувати поведінку, що відповідає чеснотам (наприклад, генерувати інформацію, аналізуючи велику кількість даних, виявляючи уважність до деталей, «готовність» переробити відповідь), – це імітація без внутрішнього наміру або прагнення.

Якщо ми говоримо про те, що епістемічний суб'єкт формує, оцінює і змінює свої переконання, беручи до уваги інтелектуальні чесноти та особисту відповідальність за якість своїх знань, то, з цієї точки зору, ШІ не є самостійним епістемічним суб'єктом, оскільки він не має власної волі, намірів або здатності до оцінки етичних аспектів знання. ШІ діє виключно на основі алгоритмів, запрограмованих людьми, і не може самостійно вибирати, які знання є важливими чи морально правильними. По суті ШІ виявляється надто довірливим для того, щоб бути самостійним епістемічним суб'єктом, – він розглядає будь-яку інформацію як сукупність фактів і будь-які висновки як достовірні.

Однак, у функціоналістському підході до епістемології, якщо ми зосереджуємося на тому, що суб'єкт «робить» для отримання знання, ШІ може бути частково визнаний епістемічним суб'єктом, оскільки він здатний до аналізу великих обсягів інформації, виявлення закономірностей, здійснення логічних міркувань. Проте, його самостійність є відносною, оскільки всі його дії залежать від початкових налаштувань і цілей, визначених людьми. Так, дедуктивні міркування – це безперечно царина ШІ, але він не здатний проаналізувати підстави дедукції. Крім того, міркування на зразок аналогії чи абдукції ШІ здатний виконувати тільки на доступних йому прикладах: по суті, в ситуації невідомої відповіді (умовно кажучи – в ситуації відкриття) – він просто «підгонить» відомі йому приклади міркувань під надану йому ситуацію, опише отриманий результат в мовній моделі і це може справити враження справжнього відкриття на недосвідченого користувача.

Тим не менш, не варто відіймати у ШІ його апроксимативні здібності – у царині великих обчислень, закономірностей великих чисел чи у наближених розрахунках – ШІ незамінний помічник людини, оскільки дійсно може «побачити» в машинному масштабі дуже великої кількості даних те, що було не помітне в людському масштабі.

З огляду на велику поширеність великих мовних моделей, зокрема, серед школярів та студентів, які використовують їх як універсальний генератор відповідей, який все «знає», виникає питання: чи може ШІ бути епістемічним авторитетом? З огляду на надане вище означення, епістемічний авторитет – це епістемічний суб'єкт, що володіє знанням високого рівня достовірності у певній галузі, які були ним набуті

доброчесно. Але ми вже показали, ШІ не є самостійним епістемічним суб'єктом і не володіє епістемічними чеснотами.

Так, ШІ здатний бути суттєвим джерелом інформації в певних спеціалізованих сферах, де він має доступ до великих обсягів даних і може швидко аналізувати їх. Він навіть може демонструвати певний рівень експертності, за рахунок «уважності» й «акуратності». Але, хоча у медичних діагнозах або при аналізі фінансових ринків, ШІ може бути дуже корисним, оскільки він має можливість враховувати величезну кількість чинників, тим не менш він не може бути авторитетом: він не може прийняти остаточне рішення високої надійності самостійно. Відтак, авторитетом все одно може бути лише людина, а у сферах значного ризику – група людей.

Таким чином, є дві проблеми, пов'язані з епістемічним авторитетом ШІ. По-перше, він не є самостійним епістемічним суб'єктом, і не може брати на себе відповідальність за свою інформацію або аналіз, як це робить людський авторитет. Відсутність самосвідомості та моральної відповідальності означає, що ШІ не може обґрунтувати свої висновки з точки зору інтелектуальних чеснот. По-друге, його знання повністю залежать від наданих даних і методів навчання, і він не може самостійно оцінити достовірність чи етичні наслідки своїх відповідей. Відтак, ШІ може бути деякою мірою експертом, але не авторитетом, оскільки авторитет ШІ є технічним, а не епістемічним і тим паче, моральним.

Тим не менш, залишається питання: чи може ШІ набути епістемічних чеснот? Оскільки епістемічні чесноти включають не тільки когнітивні здатності, але і моральні компоненти — такі як інтелектуальна чесність, сміливість, і допитливість, що відображають не лише знання, але й прагнення до істини і відповідальність за знання, стає зрозумілим, що набуття чеснот вимагає не тільки відповідних здібностей, але й морального ставлення до пізнання, що включає свідоме рішення прагнути до істини, бути відкритим до критики і неупередженим. Навіть дуже складні сучасні моделі ШІ, здатні «навчатися», не можуть набути епістемічних чеснот у тому розумінні, в якому це доступно людині, оскільки не мають самоусвідомленості й моральної мотивації.

Однак, якщо розглядати епістемічні чесноти не тільки як риси характеру, а також як функціональні характеристики системи, то можна припустити, що ШІ міг би імітувати певні чесноти. Наприклад, він міг би демонструвати інтелектуальну чесність, надаючи джерела інформації і відображаючи невизначеність в результатах (наприклад, надавати ймовірнісні оцінки і явно зазначати це). Уважність ШІ може



бути реалізована через його здатність ретельно враховувати всі деталі і використовувати багато джерел даних. Також ІІІ міг би проявляти допитливість, генеруючи гіпотези або формулюючи запитання для уточнення інформації.

Для того, щоб ІІІ міг набути таких функціональних аналогів епістемічних чеснот, необхідні певні умови. Перш за все, ІІІ повинен мати можливість самостійно навчатися на основі отриманих даних, а також адаптувати свої методи та алгоритми. Це дозволить йому втілювати аналогії допитливості й уважності, в ході того як він поступово покращує свої знання та алгоритми.

На сьогодні хороші результати навчання ІІІ демонструє лише концепція навчання з підкріпленням, однак до самостійного навчання, як це робить людина, ІІІ ще не здатний – йдеться про досі повністю нереалізовану концепцію «навчання без вчителя». Проблема неконтрольованості, яка є унікальною для RL (Reinforcement Learning – навчання з підкріпленням), полягає у моделюванні того, як агенти можуть самі генерувати цікаві та різноманітні дані за допомогою самоконтрольованих цілей. Іншими словами: як ми – люди – вчимося корисної поведінки без нагляду, а потім адаптуємо її для швидкого вирішення подальших завдань? (див.: [8]).

Щоб проявляти інтелектуальну чесність, ІІІ повинен мати здатність визнавати свої помилки, аналізувати їх і вносити корективи у свої алгоритми. Це може бути реалізовано через вбудовані механізми зворотного зв'язку, які дозволять системі навчатися на власних помилках. Це дуже складно реалізувати: по-перше, завжди існує можливість помилкової корекції, яка може увести критичну помилку до системи; по-друге, на сьогодні існує велика кількість конкуруючих ІІІ і тому ніхто з розробників не ризикне дозволяти ІІІ самому вносити зміни до своєї програми, оскільки кожен етап оновлення програми пов'язаний не тільки з її розвитком, але ще й з комерційними інтересами.

Також має місце епістемологічна проблема узгодженості алгоритмів: якщо ІІІ дозволити самому вносити зміни до власних програм, йому все одно буде потрібне втручання людини, яка визначить рівень важливості алгоритмів з тим, щоб вирішити, які з них змінювати, які ні. Рівень складності систем ІІІ такий, що у випадку внесення змін в певні алгоритми доводиться задіювати примусове рішення щодо зупинки такого перегляду. Власне, це те, що інтуїтивно робить людина, коли сама визначає для себе фактори достатності і ясності, і водочас, те, що кидає алгоритм в нескінченність.

По суті ситуація з навчанням і самонавчанням ІІІ дуже схожа на

людську: ми теж в основному навчаємося «з підкріпленням», коли вчителями постають батьки, викладачі, друзі й суспільство в цілому, і ця ж сама «клітка суспільства» визначає і обмежує наші знання та моделі поведінки, диктуючи прийнятні шляхи. Відмінність полягає в тому, що людина здатна вийти за ці рамки, принаймні в трансцендентальному вимірі, в той час як для ШІ цього виміру взагалі не існує: істина, творчість, свобода – для нього лише набір даних.

Тим не менш, існує серйозна надія (або й побоювання), що ШІ зможе «навчитися» цінностям, ідеалам, чеснотам тощо. Ключову роль у тому, чи можна зробити ШІ надійним епістемічним суб'єктом, звичайно відіграє людина, яка його розробляє і тренує. Епістемічні чесноти повинні закладатися на етапі розробки і в ході функціонування системи, шляхом створення етичних алгоритмів і постійного моніторингу роботи системи. Людина може «виховувати» ШІ, навчаючи його критеріям критичної оцінки даних, використання багатьох джерел інформації, вмінню визнавати свою обмеженість та здійснювати самоперевірку не за запитом, а самостійно.

Таким чином, для того щоб ШІ міг принаймні імітувати епістемічні чесноти, людина повинна виконати низку дій. По-перше, розробники у співпраці з філософами моралі мають розробити чіткі етичні стандарти для функціонування ШІ, щоб він міг діяти відповідно до принципів чесності, справедливості та відповідальності. Це включає програмування алгоритмів, що будуть дотримуватись прозорості і відкритості.

По-друге, потрібно перекласти розроблені стандарти на «мову», доступну ШІ, а це зовсім не тривіальна задача, виконання якої вимагає вже сьогодні залучення експертів з філософії, етики та епістемології, які допоможуть зрозуміти, як саме можна адаптувати ідею епістемічних чеснот до алгоритмічної системи. По-третє, на етапі підтримки функціонування ШІ потрібно контролювати дотримання його алгоритмами цих стандартів. Людська участь у роботі ШІ має бути постійною. Людям ще довгий час доведеться аналізувати помилки і коригувати навчання ШІ таким чином, щоб уникати системних упереджень та сприяти дотриманню епістемічних чеснот.

Отже, щоб ШІ набув елементів епістемічних чеснот, необхідно суттєво змінити як його архітектуру, так і підхід до його навчання і роботи з ним. Поки що ШІ може лише імітувати прояви деяких чеснот на основі алгоритмів, що підтримують критичний аналіз і самокорекцію. Але набуття справжніх епістемічних чеснот, які включають свідомість і моральний аспект, залишається виключно людською особливістю.

## Як ми можемо залишатися академічно добродішними, використовуючи ІІІ в своїх дослідженнях?

Очевидно, що забороняти використання великих мовних моделей та інших ІІІ в наукових та навчальних роботах немає ніякого сенсу, адже ця технологія вже інтегрована в академічне життя школярів, студентів та викладачів. Тим не менш, якщо вчитель чи керівник хоче це зробити, то він має зробити це явно. Дійсно, деякі види роботи, особливо контрольного характеру, можуть потребувати такої заборони. Тому це має бути заявлено в зрозумілому і ясному вигляді.

Тим не менш, використання ІІІ в науковій та навчальній діяльності уникнути неможливо. Адже ІІІ дуже корисний помічник, який можна використовувати якісно і добродішно. На нашу думку, основна проблема полягає в тому, що користувачі часто вважають ІІІ епістемічним авторитетом, але, як ми розглянули вище, – це зовсім не так. Перший крок, який користувач має зробити на шляху добродішного використання ІІІ – це розуміння того, що у цій взаємодії носій знань, суб'єкт пізнання і носій епістемічних чеснот – людина, яка має піклуватися про те, щоб правильно надавати ІІІ необхідну інформацію, оцінювати та коригувати його відповіді.

Типовий приклад використання ІІІ, коли користувач не усвідомлює свою відповідальність за надійність відповіді ІІІ.

Студент отримав завдання встановити логічну модальність висловлювання і переадресував задачу ІІІ, який видав таку відповідь:

« $p \rightarrow (q \vee \neg(p \wedge r))$  Цей запис підкреслює, що висловлювання виражає імплікацію ( $\rightarrow$ ) і диз'юнкцію ( $\vee$ ), а також заперечення ( $\neg$ ). Отже, логічну модальність висловлювання можна охарактеризувати як можливість. Це означає, що така ситуація може мати місце, але не обов'язково має бути такою».<sup>4</sup>

Сама відповідь демонструє невміння користувача сформулювати правильний запит до ІІІ, окреслити вимоги, контекст, параметри правильної відповіді, а також невміння співвіднести отриману відповідь з вивченим матеріалом і перевірити її на правильність. Викладачу достатньо побіжно глянути на подібну відповідь, щоб визначити ав-

---

<sup>4</sup>Правильна відповідь на цю задачу в контексті відповідного курсу логіки передбачає побудову таблиці істинності на 8 рядків і 8 стовпчиків, тому недолік розв'язання ідентифікується навіть суто візуально.

торство ШІ. Але незадовільну оцінку студент отримує не лише за те, що ШІ розв'язав задачу, а за те, що він розв'язав її неправильно.

Чи можна в такій задачі отримати правильну відповідь від ШІ? Так, для цього йому потрібно надати приклад, допустимі позначення, метод розв'язання, правила побудови таблиці істинності і варіанти логічних модальностей, які визначаються в цьому завданні. Чи буде правильна відповідь ШІ зарахована як академічно доброчесна? В даному випадку це цілком можливо, оскільки для того, щоб сформулювати такий запит до ШІ, користувач має повністю володіти матеріалом.

В наданому прикладі використання ШІ не виправдане, оскільки написання такого запиту займе набагато більше часу, ніж самостійне розв'язання, бо задача невеличка. Але, якщо користувачу потрібно розв'язати десятки таких прикладів з досить громіздкими умовами, час написання детального запиту компенсується продуктивністю ШІ. Тож, ще одним кроком до доброчесного використання ШІ є розуміння доцільності його використання.

Набагато складніше ідентифікувати тексти, створені ШІ, адже параметри написання текстів більш розмиті, ніж вимоги до формальної задачі. Джон Спенсер зазначає, що «Generative AI поставив перед академічною доброчесністю нові виклики. Створювати оригінальний вміст із мінімальними людськими зусиллями легше, ніж будь-коли. Просто введіть добре продуману підказку, і ви отримаєте текст, який можна прийняти за роботу студента» [13].

Крім того, навіть якщо ми захочемо довести, що деякий текст написаний за допомогою ШІ, програми, які мають ідентифікувати авторство тексту, не стануть нам у пригоді. По-перше, вони такі самі ШІ й мають ті самі недоліки. Практика показує, що такі програми мають досить великий відсоток похибок, коли текст, написаний ШІ, ідентифікується як текст, написаний людиною, і навпаки. Анекдотичним прикладом стало визнання однією з таких програм авторства ШІ у американській конституції.<sup>5</sup>

По-друге, параметри, які використовують програми для виявлення авторства ШІ у текстах, досить розмиті і частина людей під них цілком підпадає. Тим не менш, фахівець одразу побачить результат роботи ШІ, якщо він виконаний без задіяння користувачем власних епістемічних чеснот. ШІ досить «лінивий» – без вказівок, уточнень контексту, додаткових запитів він просто буде переформулювати

<sup>5</sup> Див.: <https://x.com/FeiziSoheil/status/1648452852736753665>

очевидну інформацію найрізноманітнішими способами.

Відтак, основним кроком до доброчесного використання ІІІ є визнання того, що він не є епістемічним авторитетом. Епістемічним авторитетом є поки що лише людина, яка може використовувати ІІІ як помічника, співрозмовника, генератора ідей або й рецензента, але людина має прийняти те, що відповідальність за результати роботи ІІІ несе саме вона. Така відповідальність передбачає усвідомлення користувачем того, яким саме чином він може доброчесно використовувати ІІІ. Тобто, які кроки користувачу потрібно зробити для того, щоб матеріал, отриманий у співпраці з ІІІ, можна було характеризувати з точки зору епістемічних чеснот.

Джон Спенсер пропонує проактивний підхід до академічної доброчесності, який полягає в тому, що ми «навчимо студентів, що означає використовувати штучний інтелект як інструмент спільної творчості таким чином, щоб уникнути когнітивної атрофії. Зі старшими учнями ми можемо інтегрувати штучний інтелект у процес письма таким чином, щоб від учнів вимагалось більше, ніж просто копіювати та вставляти текст» [13]. Для цього він пропонує низку методик, серед яких чільне місце займає навчання якісному написанню запитів.

Потрібно розуміти, що написання проміту для ІІІ вже вимагає епістемічних чеснот: користувач має розуміти, що ІІІ наслідуює його стиль постановки питань і зауважує навіть на стиль спілкування. Створення правильного запиту вимагає від користувача чітко бачити ціль запиту, розуміти можливі способи виконання запиту, вміти визначити критерії правильності отриманої відповіді, контекст, в якому має бути сформульована відповідь.

Спенсер наводить приклади таких запитів.

Я намагаюся оптимізувати свій код для [вказати конкретну ціль оптимізації], але я не впевнений, чи дотримуюся найкращих практик.

Чи можеш ти переглянути мій код і дати поради щодо оптимізації в \_\_\_\_\_ ? [13].

Примітно, що будь-який запит, метою якого є отримання суттєвої інформації від ІІІ, повинен містити значну кількість вхідних даних і передбачає серйозну роботу користувача.

Обов'язковим етапом у проактивній взаємодії з ІІІ є оцінка та критика його відповідей. Ось які підходи до оцінки пропонує Спенсер.

*Оцінка релевантності.* Визначте, чи відповідь безпосередньо стосується вашого початкового запиту. Оцініть, чи вміст пов'язаний з

розглянутою темою, чи містить нерелевантну інформацію. Перевірте, чи чат-бот дійсно зрозумів контекст чи його потрібно переглянути.

*Перевірка фактичної точності.* Перевірте точність інформації, наданої у відповіді, шляхом перехресних посилань на надійні джерела. Підкресліть будь-які неточності або оманливі твердження. Якщо ви проводите наукове дослідження, шукайте рецензовані джерела. Якщо це більш неофіційно, пошукайте інформацію в Інтернеті. Поговоріть з експертами (якщо можливо) і порівняйте це зі своїми попередніми знаннями. Хоча штучний інтелект рідше відчуває «галюцинації», це все ще трапляється.

*Оцінка побудови тексту.* Перевірте ясність, послідовність і читабельність відповіді. Шукайте граматичні помилки, незграбні фрази або структурні проблеми, які заважають розумінню. Враховуйте стиль організації тексту. Це лінійно? Це хронологічно? Чи логічний він, чи хаотичний? Або це більше компіляція? Чи не вистачає чогось?[13].

Наступним кроком буде створення уточнених запитів для ШІ, які дозволять позбавитися виявлених проблем. Спенсер пише:

Як професор, я можу подивитися на завдання й чітко наочно побачити взаємодію між ШІ та людиною. Я бачу, як ідея, створена штучним інтелектом, викликала абсолютно новий напрямок мислення, який потім призвів до чогось цілком людського [13].

Цікавою є тенденція серед наукової спільноти враховувати саме такі випадки, коли використання ШІ приводить до «чогось цілком людського». Так, MLA (англ. Modern Language Association) пропонує рекомендації щодо цитування генеративного штучного інтелекту, який визначається як інструмент, який «може аналізувати або узагальнювати вміст величезного набору інформації, включаючи веб-сторінки, книги та іншу документацію, доступну в Інтернеті, і використовувати ці дані для створення оригінального нового контенту» [6]. Тут само розміщені й рекомендації – обов'язки:

Цитуйте генеративний інструмент штучного інтелекту щоразу, коли ви перефразовуєте, цитуєте або включаєте у свою власну роботу будь-який вміст (текст, зображення, дані чи інше), створений ним; підтвердіть всі функціональні використання інструменту (наприклад, редагування вашого тексту чи перекладу слів) у примітці, тексті чи іншому відповідному місці; подбайте про перевірку вторинних джерел, на які він посилається.

Тімоті МакАду, спираючись на стиль цитування APA (англ. American Psychological Association (APA) Style), пропонує такий спосіб вказівки на використання ШІ.

Якщо ви використовували ChatGPT або інші інструменти штучного інтелекту у своєму дослідженні, опишіть, як ви використовували цей інструмент, у розділі «Методи» або в подібному розділі статті. Для оглядів літератури чи інших типів есе, чи відгуків, ви можете описати, як ви використовували цей інструмент, у вступі. У своєму тексті надайте підказку, яку ви використали, а потім будь-яку частину відповідного тексту, який був згенерований у відповідь [9].

Таким чином, можемо говорити, що використання ШІ у спільній співтворчості цілком можливе і умови його добросовісного використання цілком можуть бути підтримані науковою спільнотою. При цьому основним параметром академічної добросовісності при взаємодії зі штучним інтелектом по суті є збереження епістемічної автономії користувача за рахунок свідомого розуміння того, що ШІ не є епістемічним авторитетом.

Використання ШІ у дослідницькому процесі чи навчанні має свої переваги, але очевидно створює виклики для академічної добросовісності. Надмірна залежність від рекомендацій або відповідей ШІ може призвести до втрати здатності самостійно мислити і оцінювати достовірність знань, що суперечить принципам автономії. Епістемічна автономія передбачає готовність брати на себе відповідальність за власні знання та здатність критично аналізувати джерела інформації, включно з тими, що походять від ШІ, оскільки включає здатність розуміти обмеження ШІ як джерела знань. Зберегти епістемічну автономію у взаємодії зі ШІ можливо через активне використання епістемічних чеснот, таких як критичне мислення, інтелектуальна чесність, допитливість, налаштованість на отримання істини. Користувач має усвідомлювати, що його автономія – це і його відповідальність у взаємодії з ШІ, оскільки ШІ не має власної волі, намірів або розуміння контексту; він працює на основі обробки великих обсягів даних без здатності до рефлексії чи моральної оцінки. Це означає, що користувач повинен залишатися «останньою інстанцією» у процесі прийняття рішень і переконань, оцінюючи, наскільки обґрунтованими та доречними є рекомендації ШІ.

## References

1. *Barseghyan H.* Redrafting the Ontology of Scientific Change. *Scientonomy*. 2018. Vol. 2. P. 13-38. URL: <https://scientojournal.com/index.php/scientonomy/article/view/31032>.

2. Barseghyan H., Levesley N. Question Dynamics. *Scientonomy*. 2021. Vol. 4. P. 1-19. URL: <https://scientojournal.com/index.php/scientonomy/article/view/37120>.
3. Brogaard B. Virtue Epistemology and Epistemic Responsibility. *Externalism about Knowledge*. Oxford: Oxford University Press. 2023. P. 213–246.
4. Battaly H. Virtue Epistemology. *Philosophy Compass*. 2008. Vol. 3 (4). P. 639–663.
5. Goldman A. Knowledge in a Social World. Oxford, 1999.
6. How do I cite generative AI in MLA style. *MLA Handbook*. 9th edition. 17 March 2023. URL: <https://style.mla.org/citing-generative-ai/>
7. King N.L. Responsibilist Virtue Epistemology: A Reply to the Situationist Challenge. *The Philosophical Quarterly*. 2014. Vol. 64 (255). P. 243–253. URL: <http://www.jstor.org/stable/24672604>
8. Laskin M., Yarats D., et al. Unsupervised Reinforcement Learning Benchmark. *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. 2021. URL: <https://bair.berkeley.edu/blog/2021/12/15/unsupervised-rl/>
9. McAdoo T. How to cite ChatGPT. *APA Style*. 23 February 2024. URL: <https://apastyle.apa.org/blog/how-to-cite-chatgpt>
10. Patton P. Epistemic Tools and Epistemic Agents in Scientonomy. *Scientonomy*. 2019. Vol. 3. P. 63-89. URL: <https://scientojournal.com/index.php/scientonomy/article/view/33621>.
11. Raz J. Between Authority and Interpretation. New York: Oxford University Press, 2009.
12. Sosa E. The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge. *Midwest Studies in Philosophy*. 1980. Vol. 5. P. 3–25.
13. Spencer J. Promoting Academic Integrity in the Age of Generative AI. 6 May 2024. URL: <https://spencereducation.com/ai-academic-integrity/>



14. *Stewart C.* Expertise and Authority. *Episteme*. 2020. Vol.17 (4). P. 420–437. doi:10.1017/epi.2018.43
15. *The European Code of Conduct for Research Integrity*. Revised edition. 2023. URL: [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/european-code-of-conduct-for-research-integrity\\_horizon\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/european-code-of-conduct-for-research-integrity_horizon_en.pdf)
16. *Zagzebski L.* Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief. Oxford: Oxford University Press, 2012.
17. *Zagzebski L.* Epistemic Authority and its Critics. *European Journal for Philosophy of Religion*. 2014. Vol. 6. P. 68–87.
18. *Zagzebski L.* Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge. Cambridge: Cambridge University Press, 1996.

## ARTIFICIAL INTELLIGENCE AND ACADEMIC INTEGRITY IN THE CONTEXT OF VIRTUE EPISTEMOLOGY

Nadiia Kozachenko

**Abstract.** The aim of this article is to examine the specificity of academic integrity in the context of the increasing use of artificial intelligence (AI) in research. Traditional epistemology does not address the specific challenges posed by this issue, so it is necessary to incorporate virtue epistemology. Proponents of this approach believe that knowledge is the result of the proper use of epistemic virtues such as openness to truth, attentiveness, and intellectual honesty. Thus, virtuous knowledge arises when a person uses epistemic virtues to arrive at the truth.

Key epistemic virtues threatened by AI are epistemic autonomy and responsibility. Autonomy allows a subject to maintain its beliefs independently, while responsibility refers to the epistemic agent's accountability for the reliability of its beliefs. In addition to virtues, we also consider epistemic vices. These characteristics of knowledge help us to formulate the concept of an epistemic agent within the framework of virtue epistemology.

Based on the concept of the epistemic agent, we examine epistemic authority, its functions, and the problems that arise in interacting with it. Do we lose epistemic autonomy when we rely on authority? Is belief based on authority rationally justified? Is a person responsible for beliefs formed through reliance on authority, especially when this leads to negative consequences?

Based on our research, we address the following questions Does AI have epistemic virtues and can it be considered an epistemic authority? The main problem is that AI is not an autonomous agent and cannot be held accountable for its information or analysis as a human authority can. AI is incapable of justifying its conclusions by epistemic virtues. Its knowledge is entirely dependent on the data provided and the methods of training, and it cannot independently assess the reliability or ethical implications of its answers. Under certain conditions, however, AI can exhibit some epistemic virtues, but this requires constant interaction with humans.

The final section of the article explores the possibility of maintaining academic integrity when interacting with AI, considering its functioning through the lens of virtue epistemology in real educational and research contexts.

**Keywords:** academic integrity, artificial intelligence, virtue epistemology, epistemic authority, epistemic autonomy, epistemic responsibility.

*Надійшла до редакції 15 вересня 2024 р.*

---

**Козаченко Надія Павлівна**

Кафедра філософії

Криворізький державний педагогічний університет

просп. Університетський, 54

м. Кривий Ріг

50086

**Kozachenko Nadiia**

Department of Philosophy

Kryvyi Rih State Pedagogical University

Universytetskyi ave., 54

Kryvyi Rih

50086



<https://orcid.org/0000-0003-2358-9076>



[n.p.kozachenko@gmail.com](mailto:n.p.kozachenko@gmail.com)



<https://doi.org/10.55056/apm.7740>