

## СЕКЦІЯ 5. ВИКЛИКИ СУЧАСНОЇ ЖУРНАЛІСТИКИ В ЕПОХУ ШТУЧНОГО ІНТЕЛЕКТУ

DOI <https://doi.org/10.36059/978-966-397-443-9-16>

### ЧИ МОЖЕ ШТУЧНИЙ ІНТЕЛЕКТ ВОЛОДІТИ ЕПІСТЕМІЧНИМИ ЧЕСНОТАМИ?

**Козаченко Н. П.**

*кандидат філософських наук, доцент,  
доцент кафедри філософії*

*Криворізького державного педагогічного університету  
м. Кривий Ріг, Україна*

З точки зору епістемології чеснот, знання – це не просто результат дотримання процедур чи алгоритмів, а результат правильного використання інтелектуальних чеснот, таких як відкритість до істини, уважність, інтелектуальна чесність тощо [5, с. 167–169]. Інтелектуально добродішній агент пізнання прагне до істини не лише для того, щоб підтвердити свої переконання, але й для того, щоб відкрити нові знання і позбутися помилкових уявлень. Таким чином, знання виникає, коли особа втілює інтелектуальну чесноту в своїй діяльності з метою досягнення істини. Так, 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 [5, с. 272].

Епістемологія чеснот дає змогу підійти з іншого боку до питання аналізу епістемічних властивостей штучного інтелекту. Перш за все, це питання щодо того, чи володіє штучний інтелект епістемічними чеснотами? Адже 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. Чесноти передбачають готовність до дії на основі моральних або інтелектуальних установок, що походять з особистої відповідальності та рефлексії. ШІ, навіть найбільш просунуті моделі, не можуть мати цих епістемічних чеснот, принаймні на сучасному етапі їхнього розвитку. По суті, навіть найпопулярніші та найрозвиненіші ШІ є лише великими мовними моделями і представляють собою

маніпулятор, що не має внутрішньої мотивації або свідомості, які є основою чеснот.

З точки зору епістемології чеснот епістемічний агент розглядається як суб'єкт пізнання, що формує, оцінює і змінює свої переконання, спираючись на інтелектуальні чесноти та особисту відповідальність за якість своїх знань. З цієї точки зору, ШІ не є самостійним епістемічним агентом, оскільки він не має здатності до визначення рівня доброчесності створюваної чи оброблюваної ним інформації, він не може самостійно вибирати, які знання є важливими чи морально правильними. По суті ШІ виявляється надто довірливим для того, щоб бути самостійним епістемічним агентом, – він розглядає будь-яку інформацію як сукупність фактів і будь-які висновки як достовірні. Таким чином, відповідальність за перевірку достовірності даних ШІ повністю покладається на людину. Так, Associated Press явно постулює це в своїх стандартах щодо генеративного ШІ: «Будь-які результати генеративного інструменту ШІ слід розглядати як неперевірений вихідний матеріал. Під час розгляду будь-якої інформації для публікації співробітники АР повинні застосовувати своє редакційне судження та стандарти джерел АР» [1].

З огляду на велику поширеність великих мовних моделей та їхнє використання у якості універсального генератору відповідей, який все «знає», виникає питання: чи може ШІ бути епістемічним авторитетом? З точки зору епістемології чесноти, епістемічний авторитет – це епістемічний суб'єкт, що володіє знанням високого рівня достовірності у певній галузі, які були ним набуті доброчесно [4, с. 69]. 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.

Останніми роками Амазон заповнений книгами, гайдами та іншою літературою, згенерованою ШІ. «Грибні книги» – один з чудових прикладів того, що розглядати ШІ як епістемічний авторитет не просто не варто, але й небезпечно. Йдеться, зокрема, про «Кулінарну книгу з диких грибів» та про «Найбільший польовий путівник з книг про гриби на південному заході», про які в 2023 р. написала The Guardian

[2]. Зразки з книг були досліджені компанією, яка виявляє вміст штучного інтелекту, і виявилось, що кожен зразок отримав рейтинг 100% за показником виявлення ШІ. Тим не менш, книги не були позначені як такі, що містять матеріали, згенеровані ШІ. І що набагато гірше, книги містили дуже небезпечні рекомендації щодо ідентифікації їстівних грибів – спробувати на смак, понюхати, – що, зважаючи на високу токсичність отруйних грибів, становить серйозну загрозу для здоров'я. Коли журналісти The Guardian звернулися до фахівців мікологів, ті надали дуже негативну оцінку матеріалу книг, назвавши його просто небезпечним. Виникає питання, хто несе відповідальність за розповсюдження подібної інформації? В тому, що відповідальності не несе сам ШІ, сумнівів немає, з простої причини – він не є самостійним епістемічним агентом. Відсутність самосвідомості та моральної відповідальності означає, що ШІ не може обґрунтувати свої висновки з точки зору інтелектуальних чеснот. 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.

Тим не менш, залишається питання: чи може ШІ набути епістемічних чеснот? Якщо розглядати епістемічні чесноти не як риси характеру, а 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 результатів. Уважність ШІ може бути реалізована через його здатність ретельно враховувати всі деталі і використовувати багато джерел даних. Також ШІ міг би проявляти допитливість, генеруючи гіпотези або формулюючи запитання для уточнення інформації.

Для того щоб ШІ міг набути таких функціональних аналогів епістемічних чеснот, необхідні певні умови. ШІ повинен мати можливість самостійно навчатися на основі отриманих даних, а також адаптувати свої методи та алгоритми. Це дозволить йому втілювати аналоги допитливості й уважності, в ході того як він поступово покращує свої знання та алгоритми. Щоб проявляти інтелектуальну чесність, ШІ повинен мати здатність визнавати свої помилки, аналізувати їх і вносити корективи у свої алгоритми.

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, з одного боку, включає програмування алгоритмів, що будуть дотримуватись

прозорості та відкритості, а з іншого, – дотримання доброчесності під час навчання ШІ: «Багато моделей genAI є непрозорими системами, навченими на матеріалі, зібраному без відома чи згоди його творців» [3]. По-друге, потрібно перекласти розроблені етичні 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 сприяти дотриманню епістемічних чеснот.

Таким чином, 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. Відтак, питання доброчесності ШІ – це питання по суті доброчесної взаємодії людини та ШІ, у якій відповідальність покладається тільки на людину.

### Література

1. Barrett A. Standards around generative AI. *Associated Press*. Aug. 16, 2023. URL: <https://blog.ap.org/standards-around-generative-ai>
2. Milmo D. Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI. *The Guardian*. 1 Sep. 2023. URL: <https://www.theguardian.com/technology/2023/sep/01/mushroom-pickers-urged-to-avoid-foraging-books-on-amazon-that-appear-to-be-written-by-ai>
3. Viner K., Bateson A. The Guardian's approach to generative AI. *The Guardian*. 16 Jun. 2023. URL: <https://www.theguardian.com/help/insideguardian/2023/jun/16/the-guardians-approach-to-generative-ai>
4. Zagzebski L. Epistemic Authority and its Critics. *European Journal for Philosophy of Religion*. 2014. № 6. С. 68–87.
5. Zagzebski L. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge : Cambridge University Press, 1996.