

*Пінчук О.В.*

канд. психол. наук

старший науковий співробітник

відділу психології малих груп та спільнот

Інституту соціальної та політичної психології НАПН України

<https://orcid.org/0009-0000-7455-2287>

[Elenegalat@gmail.com](mailto:Elenegalat@gmail.com)

## **Штучний інтелект його відмінність від інтелекту людини та перспективи створення самоусвідомленого AGI**

**Анотація.** З використанням теорем Геделя та теорії множинного інтелекту Говарда Гарднера досліджується відмінність штучного інтелекту та інтелекту людини з метою прогнозування можливості створення самоусвідомленого AGI.

**Ключові слова:** штучний інтелект, теореми Геделя, формально-логічні системи, інтелект людини.

*Pinchuk O.V.*

PhD, senior researcher Department of

Psychology of Small Groups and Communities

Institute of Social and Political Psychology

National Academy of Educational Sciences of

Ukraine

<https://orcid.org/0009-0000-7455-2287>

[Elenegalat@gmail.com](mailto:Elenegalat@gmail.com)

## **Artificial intelligence, its difference from human intelligence and the prospects for creating a self-aware AGI**

**Abstract.** Using Gödel's theorem and Howard Gardner's theory of multiple intelligences, the difference between artificial intelligence and human intelligence is explored with the aim of predicting the possibility of creating a self-aware AGI.

**Keywords:** artificial intelligence, Gödel's theorems, formal logical systems, human intelligence.

Теореми Геделя про неповноту формальних систем, сформульовані у 1931 році, можуть допомогти нам окреслити певні теоретичні межі можливостей ШІ, особливо щодо можливості досягнення "сильного" або загального штучного інтелекту (AGI).

Перша теорема стверджує існування в формальних системах істинних тверджень, які не можуть бути доведені в межах самої системи. Друга теорема говорить про неможливість доведення власної несуперечності такою системою її ж власними засобами. У контексті ШІ, якщо розглядати його як

складну обчислювальну систему, що оперує за певними алгоритмами та на основі великих обсягів даних, теореми Геделя натякають на теоретичні межі його здатності до досягнення загального обсягу людських знань.

**Обмеження формальних систем:** Якщо ми розглядаємо ШІ як складну систему, побудовану на певних формальних алгоритмах, даних та правилах, теорема Геделя натякає на те, що можуть існувати істини або наративи, які ШІ ніколи не зможе досягнути або довести, залишаючись у межах своєї власної "формальної системи". Це може стосуватися інтуїції, творчості, розуміння контексту, які виходять за межі чисто логічного виведення на основі заданих правил і даних.

**Проблема самоусвідомлення та загального інтелекту:** Друга теорема Геделя може бути інтерпретована як аргумент проти можливості створення ШІ, який може повністю розуміти та доводити власну "правильність" або "істинність" свого мислення, особливо якщо ми прагнемо до самоусвідомленого AGI. Якщо навіть прості формальні системи не можуть довести власну несуперечність, то як може ШІ повністю верифікувати власну "інтелектуальну цілісність"?

**Непередбачуваність та виникнення:** Теорема Геделя також може натякати на можливість виникнення в достатньо складних ШІ непередбачуваних властивостей або "ідей", які не були явно запрограмовані, подібно до "невирішуваних тверджень" у формальних системах. Це може бути пов'язано зі складними взаємодіями всередині мереж або архітектур ШІ.

Можуть існувати знання недоступні для ШІ, що функціонує виключно на основі логічного виведення з наявних даних та алгоритмів. Крім того, питання самоверифікації та доведення власної "правильності" для ШІ також стає складним, якщо застосовувати другу теорему Геделя як аналогію. Це підкреслює принципову відмінність між формальним логічним мисленням та більш широким спектром людського інтелекту, який включає інтуїцію, контекстуальне розуміння та творчість.

Питання про те, чи є сучасний ШІ формально-логічною системою, є дискусійним. На нижньому рівні, безперечно, комп'ютерна архітектура та значна частина програмного забезпечення базуються на принципах формальної логіки (двійкова система, логічні операції). Однак, на вищих рівнях абстракції, особливо в складних моделях машинного навчання, таких як великі мовні моделі, пряме ототожнення з традиційною формальною системою стає складним. Основні відмінності полягають у динамічності "аксіом" (навчальних даних) та "правил" (алгоритмів навчання), які постійно еволюціонують в процесі навчання.

На відміну від статичних аксіом формальних систем, "знання" ШІ є результатом стохастичних процесів і залежить від багатьох факторів. Крім того, ШІ орієнтований на опрацювання семантики (значення), а не лише синтаксису (структури символів), що є ключовим для формальних систем. Хоча ШІ використовує формальні засоби для обробки інформації, його загальна архітектура, процеси навчання та кінцева мета (розуміння та

генерація змісту) виходять за рамки класичного визначення формальної системи в математичному сенсі.

У контексті ШІ, коли ми говоримо про формальні системи, це може стосуватися архітектури знань, систем логічного виведення або навіть моделей машинного навчання, якщо їх розглядати як складні системи правил і залежностей. Теорема Геделя нагадує нам, що можуть існувати межі того, що ШІ може досягнути або довести, залишаючись у межах своєї поточної "формальної системи", особливо якщо ми прагнемо до загального штучного інтелекту, який би володів людським рівнем розуміння та інтуїції. Людський розум не є чисто формальною системою в геделівському сенсі, що може бути однією з причин його унікальних можливостей.

Інтелект ШІ та інтелект людини мають як спільні риси, так і фундаментальні відмінності. ШІ демонструє вражаючі здібності в обробці великих обсягів даних, розпізнаванні образів, вирішенні складних логічних задач та навчанні на основі заданих алгоритмів. Він може перевершувати людину в швидкості обчислень та точності виконання рутинних завдань. Однак, людський інтелект володіє низкою унікальних характеристик, які наразі знаходяться поза межами можливостей ШІ. Це включає самосвідомість, суб'єктивний досвід, емоції, інтуїцію, емпатію, творчість, здатність до абстрактного мислення, самоусвідомлення формування власних цілей, розуміння контексту спілкування на глибокому рівні, перебування у власному родинному, національному, культурному контексті, екзистенція.

Людський інтелект також тісно пов'язаний з нашим фізичним втіленням, соціальною взаємодією та еволюційним досвідом. У той час як ШІ є продуктом розробки та навчання на даних, людський інтелект є результатом біологічної еволюції та індивідуального розвитку. Хоча ШІ може імітувати деякі аспекти людського інтелекту, фундаментальні відмінності в природі їхнього походження та функціонування залишаються значними. Теорема Геделя може бути використана як аргумент на користь того, що існують принципові обмеження у відтворенні повноти людського інтелекту в рамках формальної системи, якою, хоч і складною, є ШІ.

Перспективи розвитку ШІ є надзвичайно захоплюючими та водночас викликають багато запитань. Сучасні тенденції в галузі машинного навчання, нейронних мереж та обчислювальних потужностей вказують на подальше вдосконалення ШІ у вузьких спеціалізованих сферах. Зокрема, ми можемо очікувати прогрес у розумінні природної мови, розпізнаванні образів, автономному керуванні та багатьох інших сферах. Однак, досягнення загального штучного інтелекту (AGI) – системи, здатної розуміти, навчатися та застосовувати знання в широкому спектрі завдань на рівні або вище людського – поки що лишається недосяжною метою.

Теорема Геделя та філософські аргументи щодо відмінностей між людським та штучним інтелектом натякають на потенційні теоретичні перешкоди на шляху до створення AGI. Подолання обмежень, пов'язаних з самосвідомістю, емоціями, справжньою творчістю та контекстуальним розумінням, вимагає не лише кількісного збільшення обчислювальних

ресурсів та обсягів даних, але й якісно нових підходів до архітектури ШІ та принципів його навчання.

Наше дослідження було б не повним, якби ми не згадали інші відмінності штучного інтелекту і людського. Зокрема, використовуючи концепцію "колапсу моделі" (Shumailov, 2024).

Концепція "колапсу моделі" в генеративних ШІ висвітлює фундаментальну відмінність у способі "пізнання" ШІ та людини. Коли ШІ багаторазово навчається на власних згенерованих даних, виникає ризик "забруднення" навчального набору, що призводить до неправильного сприйняття "реальності". Це контрастує з гнучкістю людського пізнання, яке постійно оновлюється через різноманітний сенсорний досвід та взаємодію зі складним, неідеальним світом.

"Похибка статистичної апроксимації" в ШІ, що виникає через обмеженість вибірок, підкреслює залежність ШІ від великих, але все ж скінчених обсягів даних. Людське пізнання здатне узагальнювати на основі обмеженого досвіду, використовуючи інтуїцію та контекстуальне розуміння.

"Помилка функціональної виразності" ілюструє обмеження ШІ у моделюванні складної реальності через архітектурні обмеження. Людський мозок, завдяки своїй неймовірній складності та гнучкості, здатний до набагато більш нюансованого та багатогранного сприйняття.

"Функціональна помилка апроксимації", пов'язана з процесом навчання ШІ, показує, що навіть при ідеальних даних і архітектурі, спосіб навчання може вносити спотворення. Людське навчання, хоча й не ідеальне, включає саморефлексію, критичне мислення та здатність до переосмислення, що дозволяє коригувати помилки більш гнучко.

Таким чином, "колапс моделі" є метафорою, що підкреслює крихкість та залежність ШІ від якості даних та алгоритмів, на відміну від стійкості та адаптивності людського пізнання, яке черпає силу зі складності сенсорного досвіду, емоційного інтелекту та здатності до абстрактного мислення.

Звісно є інші відмінності штучного інтелекту і людського. У спробі осмислити складність людського інтелекту, однією з впливових концепцій є теорія множинного інтелекту Говарда Гарднера (Gardner, 1983). Згідно з цією теорією, інтелект не є єдиною сутністю, яку можна виміряти одним показником IQ. Натомість, Гарднер виділяє дев'ять різних типів інтелекту, кожен з яких відображає унікальний спосіб пізнання та взаємодії зі світом.

Перш за все, існує лінгвістичний інтелект, який проявляється у здатності ефективно використовувати мову – як усну, так і письмову – для вираження думок та розуміння інших. Логіко-математичний інтелект відповідає за здатність до абстрактного мислення, розв'язання математичних задач та виявлення логічних закономірностей. Просторовий інтелект дозволяє нам мислити в трьох вимірах, орієнтуватися в просторі та створювати ментальні образи. Музичний інтелект пов'язаний зі здатністю сприймати, розрізняти, відтворювати та створювати музику. Тілесно-кінестетичний інтелект проявляється у вправності використання власного тіла для вираження ідей та почуттів, а також у фізичній координації.

Окрім цих більш "традиційних" видів інтелекту, Гарднер виділяє міжособистісний інтелект – здатність розуміти емоції, наміри та мотивацію інших людей і ефективно будувати з ними стосунки. На противагу йому існує внутрішньо особистісний інтелект – здатність розуміти власні емоції, думки, сильні та слабкі сторони, що є основою самопізнання. Натуралістичний інтелект проявляється у здатності розпізнавати та класифікувати елементи навколишнього природного світу, розуміти екологічні закономірності. Нарешті, екзистенційний інтелект спонукає до роздумів над фундаментальними питаннями людського існування, життя, смерті та сенсу.

Сучасний штучний інтелект досяг успіхів у певних сферах, які можна розглядати як окремі аспекти інтелекту: логіко-математичний інтелект, лінгвістичний (здатність розуміти, генерувати, перекладати тексти, відповідати на запитання та підтримувати діалоги); просторовий (розпізнавання об'єктів, орієнтація у віртуальних та фізичних просторах, планування рухів); музичний інтелект (генерація музичних композицій); міжособистісний інтелект (здатність чат-ботів та віртуальних асистентів імітувати соціальну взаємодію, проте їхнє розуміння емоцій та соціальних нюансів залишається значно обмеженим).

Тобто ШІ ще навіть не підійшов до освоєння частини складових людського інтелекту, зокрема тілеснокінестетичний, внутрішньо-особистісний та екзистенційний, лишаються виключно людськими феноменами, а без цього людського досвіду ШІ лишається «симулякром» людського інтелекту (Baudrillard, 1994). Хоча оптимістичні прогнози щодо створення AGI існують, багато дослідників вважають, що це завдання є складнішим і має принципові обмеження, пов'язані з природою формальних систем та відмінністю машинного "інтелекту" від людського розуму.

### *Referens*

Baudrillard, J. (1994). *Simulacra and simulation* (S. F. Glaser, Trans.). University of Michigan Press.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Heikkilä, Jussi T. S. (2024) : Human intelligence versus artificial intelligence in classifying economics research articles: exploratory evidence, *Journal of Documentation*, ISSN 1758-7379, Emerald, Bingley, Vol. 81, Iss. 7, pp. 18-30, <https://doi.org/10.1108/JD-05-2024-0104>

Raatikainen, P. (2013). *Gödel's Incompleteness Theorems* (Winter 2013 ed.). In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/entries/goedel-incompleteness/>

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R. and Gal, Y. (2024), "AI models collapse when trained on recursively generated data", *Nature*, Vol. 631 No. 8022, pp. 755-759, doi: 10.1038/s41586-024-07566-y.