

ОСОБЛИВОСТІ ОЦІНКИ ЛОГІКО-СМИСЛОВИХ ЗВ'ЯЗКІВ ЗА ДОПОМОГОЮ ВЕРБАЛЬНИХ МОДЕЛЕЙ

FEATURES OF EVALUATING LOGICAL-SEMANTIC CONNECTIONS USING VERBAL MODELS

У статті представлено результати емпіричного дослідження можливостей використання вербальних моделей для аналізу та оцінки логіко-смислових зв'язків. Актуальність дослідження зумовлена зростанням ролі інформаційної невизначеності у сучасному соціальному середовищі та потребою розробки об'єктивних методів аналізу когнітивної продуктивності. Метою статті є встановлення рівня надійності та придатності оцінювання логіко-смислових зв'язків трьома найпопулярнішими вербальними моделями (Claude, ChatGPT, Gemini), а також зіставлення їхніх показників з експертною оцінкою психолога при аналізі результатів «Методики текстуальних розривів». У ході дослідження було проаналізовано 46 робіт підлітків 16–17 років. Кожен респондент формулював 29 логіко-смислових зв'язків. Виконане завдання оцінювалося трічі трьома вербальними моделями з різних облікових записів. Результати дослідження показали, що внутрішня придатність вербальних моделей при аналізі логіко-смислових зв'язків змінюється від 61% до 73% точних збігів, при цьому Claude демонструє порівняно вищу надійність (α -Кронбаха = 0,757). Водночас міжмодельна узгодженість штучного інтелекту виявилась досить низькою ($r = 0,28$, $\alpha = 0,452$). Це може вказувати на суттєві розбіжності в алгоритмах обробки інформації. Порівняння оцінок, зроблених вербальними моделями, й експертною оцінкою зафіксувало лише 55–56% точних збігів і прямий слабкий кореляційний зв'язок ($r = 0,34$ – $0,41$). Це може вказувати на принципові відмінності у логіці оцінювання логіко-смислових зв'язків між штучним і природним інтелектом. Зроблено висновок про доцільність застосування у дослідницькій роботі гібридного підходу. Зокрема, йдеться про використання вербальних моделей для первинного скрінінгу з обов'язковою експертною верифікацією результатів.

Ключові слова: вербальні моделі, логіко-смислові зв'язки, інформаційна невизначеність, експертна оцінка, методика текстуальних розривів.

ність, експертна оцінка, методика текстуальних розривів.

The article presents the results of an empirical study on the possibilities of using verbal models for the analysis and evaluation of logical-semantic connections. The relevance of the study is determined by the growing role of information uncertainty in the contemporary social environment and the need to develop objective methods for analyzing cognitive productivity. The aim of the article is to establish the level of reliability and suitability of evaluating logical-semantic connections by three of the most popular verbal models (Claude, ChatGPT, Gemini), as well as to compare their indicators with the expert assessment of a psychologist in analyzing the results of the «Textual Gaps Method». During the study, 46 works by adolescents aged 16–17 years were analyzed. Each respondent formulated 29 logical-semantic connections. The completed task was evaluated three times by three verbal models from different accounts. The research results showed that the internal consistency of verbal models in analyzing logical-semantic connections ranges from 61% to 73% exact matches, with Claude demonstrating relatively higher reliability (Cronbach's $\alpha = 0.757$). At the same time, inter-model consistency of artificial intelligence proved to be quite low ($r = 0.28$, $\alpha = 0.452$). This may indicate significant differences in information processing algorithms. Comparison of assessments made by verbal models and expert evaluation revealed only 55–56% exact matches and a weak direct correlation ($r = 0.34$ – 0.41). This may indicate fundamental differences in the logic of evaluating logical-semantic connections between artificial and natural intelligence. A conclusion was drawn regarding the appropriateness of applying a hybrid approach in research work. Specifically, this involves using verbal models for initial screening with mandatory expert verification of results.

Key words: verbal models, logical-semantic connections, information uncertainty, expert assessment, textual gaps method.

УДК 159.955:004.8

DOI <https://doi.org/10.32782/hbts.82.1.21>

Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0



Дрозд П.А.

аспірант

Житомирський державний університет імені Івана Франка

ORCID ID: 0009-0007-6161-1779

Світлична Т.В.

асистент кафедри соціальної та практичної психології

Житомирський державний університет імені Івана Франка

ORCID ID: 0009-0000-2266-9950

Постановка проблеми. Помітне зростання ролі інформаційної невизначеності у сучасному соціальному середовищі викликає потребу у здійсненні оцінки когнітивної продуктивності, а головне в аналізі її точності.

Сучасному інформаційному простору притаманна можливість швидко перевіряти дані. З одного боку, це дозволяє знімати проблему інформаційної невизначеності, компенсувати дефіцит інформації, досягати однозначності її розуміння у процесі інтерпретації або реконструювання. У зв'язку з цим проблема оцінки когнітивної продуктивності у ситуаціях інформаційної невизначеності набуває особливої дослідницької актуальності.

Інформаційна невизначеність традиційно розглядається як один із чинників, що суттєво модифі-

кує когнітивні процеси, зокрема судження, процес оцінювання та вибір [6; 12]. Зокрема, у межах теорії подвійних процесів мислення [4] інформаційна невизначеність розглядається як чинник, що впливає на баланс між інтуїтивними й аналітичними механізмами обробки інформації. З іншого боку, встановлено, що в умовах інформаційної невизначеності відбувається активізація евристичних стратегій її обробки, що, втім, може позначитися на підвищенні ймовірності когнітивних викривлень [5]. Виникає нагальна потреба у розробці методологічних підходів до цілісного й системного аналізу когнітивного продукту. У перспективі це дозволить поглибити розуміння механізмів мислення у складних ситуаціях і назагал сприяти вдосконаленню психологічної діагностики й про-

гнозуванню ефективності когнітивної діяльності. Закономірною постає спроба залучити до цього процесу вербальні моделі, які можуть з вищим ступенем об'єктивності оцінювати отриманий смисловий продукт. Обґрунтування таких досліджень є досить переконливим [8; 10]. Поготив існують перші дослідження порівняння штучного та природного інтелекту [3], які показують, що вербальні моделі можуть слугувати додатковим діагностичним інструментом у питаннях аналізу мислення.

Аналіз останніх досліджень і публікацій.

Сучасні мовні моделі демонструють здатність до виконання складних когнітивних операцій. Зокрема, мовні моделі здатні враховувати контекст у міркуваннях, продукують деякі індуктивні упередження, а в ситуації інформаційної невизначеності здатні до евристик. Іншими словами, вербальні моделі значною мірою імітують мислення, навіть у його недосконалих проявах [3]. Дж. Петерсон зі співавторами показали, що машинне навчання цілком здатне здійснювати реконструкцію латентних стратегій мислення на основі великого масиву поведінкових даних [11]. Це дає підстави припускати, що вербальні моделі придатні для того, щоби виявляти структури когнітивної діяльності людей, які навіть не є очевидними при здійсненні традиційного аналізу. В інших дослідженнях вказується на те, що існує достатній кореляційний зв'язок між інтелектом людини та машини [7].

Разом із тим, робити повне ототожнення когнітивних характеристик природного та штучного інтелекту не можна. По-перше, аналізу підлягає не сам процес переробки інформації, а його результати [9], що, звісно, не дозволяє побачити функціональну специфіку природного та штучного інтелекту. По-друге, дослідникам вдається фіксувати лише зовнішні аспекти когнітивного матеріалу: структурованість причинно-наслідкових зв'язків, складність аргументації, варіативність припущень тощо. По-третє, штучний інтелект є по суті виявом накопиченого колективного досвіду, а не індивідуального, що, втім, можна розглядати як певну перевагу при аналізі продуктів індивідуального природного мислення. Іншими словами, вербальні моделі залишаються все-таки моделями, що репрезентують окремі варіанти можливого когнітивного продукту, сформованого на платформі загальнолюдського досвіду. Це означає, що вербальні моделі не здатні користатися інтуїцією, непевною здогадкою, що властиво мисленню [4], а виходять із певних наявних зразків.

Отже, вербальні моделі, дарма що обмежено, можуть використовуватися для оцінки природного мислення як певний зразок вербального продукту, з яким можуть порівнюватися відповіді респондентів. Тобто вони можуть перевіряти релевантні варіанти смислового продукту, ступінь їхньої адекватності, допустимості у конкретному випадку. Тобто вербальні моделі по суті створюють модельоване середовище, яке дозволяє перевіряти варіанти когнітивних гіпотез. Мало того, вони певною мірою

стають спроможними класифікувати типи представлених для аналізу аргументів, ступінь логічності структури (відповідності законам формальної логіки).

Однак застосування вербальних моделей як інструменту оцінки природного мислення вимагає встановлення чітких методологічних меж. Вербальні моделі, вочевидь, поки що здатні вивчати тільки окремі аспекти мислення, фіксуючи його зовнішні ознаки, а не його цілісний та глибинний характер.

Нами вирішено перевірити, якою мірою вербальні моделі спроможні аналізувати результати досліджень за допомогою «Методики текстуальних розривів», розробленої О.В. Мазяром [1; 2]. Ця методика передбачає, що респонденти у ході виконання основного завдання заповнюватимуть так звані смислові лакуни – пропуски парних речень в оповіданні, що моделюють собою ситуацію інформаційної невизначеності. Таких речень, які необхідно створити/відтворити, є п'ятнадцять. Процес аналізу когнітивної продуктивності передбачає встановлення логіко-смислових зв'язків між попередніми та наступними реченнями, яких назагал виявляється двадцять дев'ять. Основними типами логіко-смислових зв'язків автор вважає імітаційність, алогічність, аморфність, логічність та ефазивність. Можливими також стають смислові пропуски, коли респондент не здатен сформулювати відповідне речення. Кожен із цих типів зв'язків виконує певну функцію у побудові загального смислу оповідання. При цьому імітаційність, алогічність та аморфність відбивають слабкий, суперечливий або невиразний смисловий зв'язок, тоді як логічність та ефазивність показують, що респондент задовільно впорався із завданням.

Слабке місце «Методики текстуальних розривів» полягає саме у тому, що логіко-смисловий зв'язок між сусідніми реченнями не завжди є очевидним в інтерпретації. З цією метою доцільним стає аналіз смислового матеріалу кількома експертами та пошук консенсусу. Натомість вербальні моделі можуть виступати в якості таких експертів, поготив існує можливість використання кількох вербальних моделей та їхнього повторного застосування.

У нашому дослідженні вирішено використати три найбільш популярні вербальні моделі: Chatgpt, Claude та Gemini. Інтерес становить не тільки зіставлення їхніх результатів з експертною оцінкою психолога (ступінь збігу інтерпретаційних показників), але й виявлення особливостей аналізу у межах кожної вербальною моделі, що здійснюється з різних облікових записів. У перспективі це може показати ступінь однорідності їхніх інтерпретаційних показників.

У ході дослідження було проаналізовано 46 робіт осіб підліткового віку (16–17 років). Спочатку роботи аналізувалися двома психологами та встановлювалася консолідована логіко-смислова оцінка зв'язку між реченнями. Паралельно роботи

пропонувалися на перевірку зазначеним вербальним моделям, причому це робилося тричі, для чого створений когнітивний продукт завантажувався з трьох облікових записів для кожної з вербальних моделей.

Результати дослідження. Спочатку визначимо ступінь надійності логіко-сміслових оцінок вербальних моделей. Для цього обраховуватимемо відсоток абсолютного збігу зроблених оцінок внаслідок трьох незалежних серій за кожною вербальною моделлю та рівень кореляційних зв'язків між ними. У випадку збігу оцінок з'ясуємо точність відтворення, а у випадку кореляційного аналізу показників – узгодженість логіко-сміслових патернів для кожної вербальної моделі. Додатково для оцінки внутрішньої узгодженості результатів використовуватимемо коефіцієнт α -Кронбаха.

Результати дослідження за трьома вербальними моделями представлено у таблиці 1.

Середній кореляційний показник між трьома вербальними моделями становить 0,28, а показник внутрішньої узгодженості α -Кронбаха – 0,452.

Таким чином, внутрішня узгодженість результатів вербальних моделей показала, що за показниками збігу оцінок найбільшу стабільність демонструє Claude (73% ідентичних оцінок при двох повторних запитах), тоді як ChatGPT і Gemini продемонстрували дещо нижчі показники – 68% та 61% відповідно. Це свідчить про те, що приблизно у третині випадків вербальні моделі генерують різні оцінки для ідентичних вхідних даних за умови використання різних облікових записів однієї моделі.

Коефіцієнт α -Кронбаха для внутрішньої узгодженості підтверджує цю тенденцію: Claude демонструє цілком прийнятну надійність ($\alpha = 0,757$). Натомість ChatGPT ($\alpha = 0,667$) і Gemini ($\alpha = 0,664$) демонструють дещо нижчу надійність, що перебуває на межі прийнятності.

Кореляційний аналіз повторних вимірів у межах окремих вербальних моделей виявив прямий помірний зв'язок для Claude ($r = 0,62$), що вказує на збереження аналізу загальних патернів оцінювання, хоча з деякими варіаціями у конкретних значеннях. Натомість ChatGPT ($r = 0,52$) і Gemini ($r = 0,45$) демонструють дещо слабший зв'язок, що можна свідчити про порівняно вищу варіативність у закономірностях оцінювання або вищу чутливість цих вербальних моделей до контекстуальних змін при повторних запитах.

Аналіз міжмодельної узгодженості логіко-сміслових оцінок свідчить про те, що середній кореля-

ційний показник між трьома вербальними моделями становить лише 0,28, що вказує на їхній прямий слабкий зв'язок. Це, своєю чергою, свідчить про існування суттєвих розбіжностей в алгоритмах обробки й інтерпретації смислових зв'язків. Показник внутрішньої узгодженості α -Кронбаха для всіх трьох моделей разом становить 0,452, що є неприйнятно низьким рівнем і вказує на порушення консенсусу між оцінками одних і тих самих смислових конструкцій.

Отримані результати засвідчують, що використання вербальних моделей в оцінці природного мислення має кілька суттєвих недоліків. По-перше, існують деякі обмеження у відтворюваності результатів навіть у межах найбільш популярних вербальних моделей. По-друге, наявна недостатня узгодженість між різними вербальними моделями, що ставить під сумнів можливість стандартизації подібних досліджень. При цьому Claude демонструє порівняно кращі показники надійності серед досліджуваних вербальних моделей, проте навіть для цієї системи, вочевидь, слід використовувати множинне тестування та верифікацію результатів альтернативними методами.

Тепер зіставимо оцінки, отримані за допомогою вербальних моделей, та консолідованою оцінкою експертів. Оскільки назагал було проведено дев'ять серій досліджень (по три з кожною вербальною моделлю), вирішено зіставляти їх за показниками мір центральної тенденції: середнім арифметичним показником, модою та медіаною. Зрозуміло, що основною при цьому слід вважати медіану, оскільки вона є стійкою до випадкових викидів, які виникають через варіативність відповідей вербальних моделей. Зауважимо, що попередньо було з'ясовано, що рівень невідповідності становить 27–39%. Натомість медіана відбиває типові значення, яке вербальна модель генерує найбільш стабільно. Заразом проаналізуємо середнє арифметичне та моду. Мода як найбільш частий показник у множині теж є досить цінною, оскільки демонструє «консенсусний» результат, який так само, як і медіана, є реальним значенням з представленому наборі оцінок, а не обчисленим артефактом, як середнє арифметичне.

Результати зіставлення експертної оцінки та показників вербальних моделей представлено у таблиці 2. Зауважимо, що поррахувати збіг оцінок у випадку середнього арифметичного математично недоцільно, оскільки цей показник не становить цілого числа, а значить таких збігів майже не буде.

Таблиця 1

Показники відповідності результатів між трьома вербальними моделями

Вербальна модель	Збіг оцінок	Кореляційний показник	α -Кронбаха
Claude	73%	0,62	0,757
Chatgpt	68%	0,52	0,667
Gemini	61%	0,45	0,664

Показники відповідності результатів між експертною оцінкою та показниками вербальних моделей

Міри центральної тенденції	Збіг оцінок	Кореляційний показник	α -Кронбаха
Медіана	0,55	0,35	0,391
Мода	0,56	0,34	0,388
Середнє арифметичне	–	0,41	0,447

Частота точних збігів логіко-смыслових оцінок експертної оцінки та вербальних моделей у випадках медіани та моди є практично ідентичною – 55% та 56% відповідно. Це означає, що дещо більше, ніж у половині випадків, вербальні моделі генерують оцінки, які точно відповідають експертному судженню.

Кореляційний аналіз дозволив встановити слабкий прямий зв'язок між оцінками вербальних моделей та експертом: $r = 0,35$ – для медіани, $r = 0,34$ – для моди та $r = 0,41$ – для середнього арифметичного. Це можна свідчити про те, що загальні тенденції зміни оцінок вербальних моделей (при врахуванні середнього арифметичного) не досить виразно узгоджуються з експертною логікою. Тобто не маємо систематичного узгодження між вербальними моделями та експертом у логіці оцінювання змістових зв'язків.

Слід додаткового зауважити на тому, що мінімальна різниця оцінок між модою та медіаною (збіг – 55% vs 56%, кореляція – 0,35 vs 0,34) вказує на те, що розподіл оцінок вербальних моделей є досить симетричним, а значить немає проблем із мультимодальністю.

Показники внутрішньої узгодженості за допомогою коефіцієнта α -Кронбаха для всіх комбінацій виявились досить низьким: 0,391 – для медіани, 0,388 – для моди та 0,447 – для середнього арифметичного. Жоден із показників не досягає мінімальної порогової прийнятної надійності ($\alpha \geq 0,60$). Це вказує на принципові розбіжності у підходах до оцінювання між вербальними моделями та експертом. Принагідно зауважимо, що існує помітний контраст між внутрішньою узгодженістю самих вербальних моделей та їхньою узгодженістю з експертом.

Низькі показники узгодженості можуть пояснюватися кількома факторами. По-перше, існують відмінні референтні рамки, адже експерт має змогу спиратися на глибинні теоретичні конструкції, керуватися неявними знаннями та контекстуальним розумінням смислу, які вербальними моделями не завжди враховуються. По-друге, вербальні моделі ще досі недостатньо включені у розуміння смислових нюансів, не виявляють тонкощі окремих змістових відношень і часто-густо залишаються поверхневими при аналізі. По-третє, самі вербальні моделі демонструють достатньо високу варіабельність оцінювання: усередині деяких вербальних моделей збіг становить лише 61–73%. При додаванні експертної оцінки отримуємо ще більшу внутрішню

неузгодженість оцінок. Втім, варто зауважити, що така варіабельність представлена і в оцінках експертів, вивчення чого може стати перспективою подальших досліджень.

Висновки. Отримані дані свідчать про те, що на сучасному етапі вербальні моделі не можуть повною мірою замінити експертну оцінку при аналізі змісту логіко-смыслових зв'язків. Показники збігу на рівні 55–56% та слабка кореляція ($r = 0,34$ – $0,41$) вказують на те, що вербальні моделі й експерти здебільшого оперують різною «логікою» оцінювання. Проте це ще не означає, що одна зі сторін неодмінно помиляється. Вочевидь, йдеться про різні епістемологічні підходи до розуміння логіко-смыслових зв'язків. Це означає, що для дослідницьких цілей доцільно використовувати гібридний підхід, який полягає у тому, що вербальні моделі здійснюють первинний скринінг логіко-смыслових зв'язків, однак вирішальним залишається експертна верифікація результатів.

ЛІТЕРАТУРА:

1. Мазяр О., Марчук А. Динаміка логіко-смыслових зв'язків у юнацькому віці. *Вісник Львівського університету. Серія «Психологічні науки»*. 2024. Вип. 22. С. 83–89. DOI: <https://doi.org/10.30970/PS.2024.22.11>.
2. Мазяр О., Світлична Т. Вікові відмінності вербальної динаміки мислення. *Журнал соціальної та практичної психології*. 2024. № 4. С. 12–16. DOI: <https://doi.org/10.32782/psy-2024-4-2>.
3. Binz M., Schulz E. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*. 2023. Vol. 120. № 6. e2218523120. DOI: <https://doi.org/10.1073/pnas.2218523120>
4. Evans J. St. B. T., Stanovich K. E. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*. 2013. Vol. 8, № 3. P. 223–241. DOI: <https://doi.org/10.1177/1745691612460685>
5. Gigerenzer G., Gaissmaier W. Heuristic decision making. *Annual Review of Psychology*. 2011. Vol. 62. P. 451–482. DOI: <https://doi.org/10.1146/annurev-psych-120709-145346>
6. Kahneman D. *Thinking: Fast and Slow*. New York: Farrar, Straus and Giroux, 2011. 512 p.
7. Kosinski M. Theory of mind may have emerged in LLMs. *arXiv*. 2023 <https://arxiv.org/abs/2302.02083>
8. Lake B. M., Ullman T. D., Tenenbaum J. B., Gershman S. J. Building machines that learn and think like people. *Behavioral and Brain Sciences*. 2017.

Vol. 40. e253. DOI: <https://doi.org/10.1017/S0140525X16001837>

9. Mitchell M. Artificial intelligence: A guide for thinking humans. Farrar, Straus and Giroux, 2019. 448 p.

10. Newell A., Simon H. A. Human problem solving. *American psychologist*. 1971. Vol. 26, No. 2. P. 145–159.

11. Peterson J. C., Bourgin D. D., Agrawal M., Reichman D., Griffiths T. L. Using large-

scale experiments and machine learning to discover theories of human decision-making. *Science*. 2021. Vol. 372. № 6547. P. 1209–1214. DOI: <https://doi.org/10.1126/science.abe2629>

12. Tversky A., Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science*. 1974. Vol. 185, № 4157. P. 1124–1131. DOI: <https://doi.org/10.1126/science.185.4157.1124>

Дата першого надходження статті до видання: 18.02.2026

Дата прийняття статті до друку після рецензування: 16.03.2026

Дата публікації (оприлюднення) статті: 11.05.2026