

ФЕНОМЕН КОЛЕКТИВНОГО ІНТЕЛЕКТУ В КОНТЕКСТІ КОНВЕРГЕНЦІЇ ІМІТАЦІЙНИХ СИСТЕМ І ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПРОФЕСІЙНОГО РОЗВИТКУ ПЕДАГОГІЧНИХ КАДРІВ

THE PHENOMENON OF COLLECTIVE INTELLIGENCE IN THE CONTEXT OF THE CONVERGENCE OF SIMULATION SYSTEMS AND ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR THE PROFESSIONAL DEVELOPMENT OF TEACHING STAFF

У дослідженні розглядаються аспекти психологічного впливу щодо використання технологій штучного інтелекту в процесі дослідницького навчання, аналізуються відповідні фактори і показники впливу. Мета експериментального дослідження полягає в досягненні ґрунтовного розуміння феномену колективного інтелекту в контексті конвергенції імітаційних систем і технологій штучного інтелекту для професійного розвитку педагогічних кадрів, у т.ч. психологічного впливу комп'ютерної техніки у віртуальному середовищі в процесі дослідницького навчання. Для досягнення завдань дослідження використовуються експериментальні майданчики «Clever: School of Natural and Mathematical Sciences». Особлива увага приділяється виявленню ризиків, труднощів і небезпек у віртуальному середовищі з метою виокремлення важливих тенденцій для перспективного подальшого розвитку колективного інтелекту з методично вмотивованим використанням компонентів комп'ютерно орієнтованої методичної системи дослідницького навчання (КОМСДН) в процесі дослідницького навчання. Розроблено систему учбової діяльності в процесі дослідницького навчання, в результаті якої набувається досвід із педагогічно виваженим і методично вмотивованим використанням ШІ. Особлива увага приділяється експериментальному дослідженню впливу моделей на розвиток інтелекту. Встановлено необхідність здійснення добору засобів ШІ для підвищення мотивації і рівня інтелектуального розвитку, що призводить до підвищення ефективності дослідницького навчання. Результати виявилися значущими на рівні достовірності $p \leq 0,05$. Отримані в процесі експериментального дослідження дані використовувалися для здійснення аналізу найбільш актуальних в процесі дослідницького навчання дидактичних засобів у контексті конвергенції імітаційних систем і технологій ШІ для професійного розвитку педагогічних кадрів.

Ключові слова: дослідницьке навчання, штучний інтелект, імітаційні системи, дослідницька задача, методична система дослідницького навчання, колективний інтелект, когнітивний розвиток, глибинне навчання, математичне моделювання, конвергенція, дидактика, COMSRL, AI, IVR.

The present study examines aspects of psychological influence on the use of artificial intelligence technologies in the process of research training, analysing relevant factors and indicators of influence. The objective of the present experimental study is to achieve a comprehensive understanding of the phenomenon of collective intelligence in the context of the convergence of simulation systems and artificial intelligence technologies for the professional development of teaching staff. This encompasses the psychological ramifications of computer technology in a virtual environment during the process of research training. In order to achieve the objectives of the research, the experimental sites are to be referred to as 'Clever: The School of Natural and Mathematical Sciences'. It is imperative to acknowledge the significance of risk identification in the virtual environment, as this facilitates the discernment of pivotal trends that are instrumental in the prospective advancement of collective intelligence. The identification process is further augmented by the utilisation of the components of the computer-oriented methodological system of research learning (COMSRL) in the research learning process, a methodology that is methodologically motivated. A system of learning activities has been developed for the purpose of facilitating exploratory learning. The consequence of this advancement is the attainment of expertise in the pedagogically balanced and methodologically motivated utilisation of AI. A particular focus is placed on the experimental study of the impact of models on the development of intelligence. The necessity of implementing a selection of AI tools to increase motivation and level of intellectual development has been established, which leads to an increase in the effectiveness of exploratory learning. The results were significant at the level of reliability $p \leq 0,05$. The data obtained during the experimental research were utilised to analyse the most relevant didactic tools in the context of research training, with a focus on the convergence of simulation systems and AI technologies for the professional development of teaching staff.

Key words: exploratory learning, artificial intelligence, simulation systems, research tasks, methodological system of exploratory learning, collective intelligence, cognitive development, deep learning, mathematical modelling, convergence, didactics, COMSRL, AI, IVR.

УДК 373.5:5]004
DOI <https://doi.org/10.32782/2663-5208.2025.74.33>
Стаття поширюється на умовах ліцензії CC BY 4.0

Гриб'юк О.О.

к.пед.н., доцент, старший дослідник, провідний науковий співробітник Інститут цифровізації освіти Національної академії педагогічних наук України

Вступ. Поява і розвиток людино-машинних обчислювальних комплексів означає якісну зміну можливостей людини в пізнанні навколишнього світу. Швидко і просто введення інформації в обчислювальну систему, практично миттєва відповідь, подана в наочній формі, саме ці можливості дають змогу об'єднати можливості комп'ютера швидко

виконувати складні логічні операції, зокрема обчислення, зі здатністю людини мислити неформально, використовувати інтуїцію, здатність до асоціацій, нетривіальні порівняння тощо. Для нас важливі не технологічні характеристики та особливості сучасних обчислювальних машин і навіть не їхні економічні характеристики, а ті нові можливості,

які вони надають для розширення інтелектуальних, розумових здібностей фахівців. Для того, щоб використовувати обчислювальну техніку під час вивчення будь-якого явища або для вибору альтернативи із сукупності можливих рішень у практичній діяльності фахівця, необхідно насамперед мати у своєму розпорядженні необхідний математичний опис досліджуваного явища або аналізованої ситуації, тобто розробити деяку систему математичних моделей [1]. Традиційна схема застосування обчислювальних засобів для розв'язування складних завдань, що формулюються практиками звичайною мовою, вимагає їх переформулювання в термінах відповідних математичних моделей. Це дає змогу звести дослідження до розв'язання серії конкретних математичних задач, для чого використовуються спеціальні методи обчислювальної математики, які дають змогу провести необхідні розрахунки за допомогою ЕОМ. Подібна схема, однак, у її звичайному вигляді з багатьох причин не може бути використана для розв'язання проблем дослідження процесів глобального характеру та управління ними. Причина в тому, що сьогодні ми ще не можемо побудувати замкнуту модель дослідження [2]. Багато зв'язків, що відіграють важливу роль у процесах інтелектуалізації, дотепер неможливо точно формалізувати, тобто описати їх мовою математики. Для їх представлення ми змушені використовувати різного виду параметризації, що ґрунтуються на експертному оцінюванні. Важливо підкреслити, що навіть якщо нам вдалося побудувати єдину замкнену модель, ми не змогли б сформулювати для цієї моделі математичну задачу, яка б адекватно відтворювала смисл виразу «розв'язання суперечностей», оскільки сьогодні ми не можемо чітко сформулювати цей смисловий контекст мовою кількісних понять [3]. Останнє вимагає самостійних і кропітких досліджень за допомогою тих самих моделей. Тут ми стикаємося із ситуацією, в якій об'єктом дослідження виявляються самі цілі дослідження та їх чітке формулювання. А для цього необхідно знайти способи вивчення самого феномена розвитку інтелекту [2] залежно від того, якою буде людська діяльність, тобто необхідно відтворити, імітувати за допомогою комп'ютера варіанти такого розвитку.

Електронна обчислювальна техніка потрібна нам не тільки як засіб для розв'язування певного набору математичних задач. Вона нам необхідна як певна своєрідна експериментальна установка, що дає змогу описати модель інтелекту [3] як єдине ціле і за допомогою методу математичного моделювання здійснювати необхідну експериментальну діяльність, відкривати її властивості, знаходити прийнятні варіанти поведінки людини в соціумі [4]. Для цього потрібен інструмент [5], що дає змогу об'єднати традиційні неформальні методи, які використовують аналогію, інтуїцію, досвід, з обчислювальним комплексом.

Мета експериментального дослідження полягає в досягненні ґрунтовного розуміння феноме-

ну колективного інтелекту в контексті конвергенції імітаційних систем і технологій штучного інтелекту для професійного розвитку педагогічних кадрів, у т.ч. психофізіологічного впливу комп'ютерної техніки у віртуальному середовищі в процесі дослідницького навчання. Для досягнення завдань дослідження використовуються експериментальні майданчики «Clever: School of Natural and Mathematical Sciences» [3]. Особлива увага приділяється виявленню ризиків, труднощів і небезпек у віртуальному середовищі з метою виокремлення важливих тенденцій для перспективного подальшого розвитку колективного інтелекту з методично вмотивованим використанням компонентів комп'ютерно орієнтованої методичної системи дослідницького навчання (КОМСДН) в процесі дослідницького навчання [6].

Виклад основного матеріалу. У рамках дослідження [7] вдалося не тільки створити математичну модель, а й звести аналіз і вибір необхідних конструктивних параметрів щодо серії чітко сформульованих математичних завдань. Цілі дослідження нами були задані заздалегідь. Комп'ютер перетворюється на абсолютно новий інструмент пізнання, завдяки якому стало можливим організовувати діалог між дослідником і ЕОМ, завдяки чому з'являється можливість *синтезу формальних і неформальних методів аналізу*. Таким чином забезпечується можливість поєднання потенційних здібностей людського інтелекту зі здатністю обчислювальної машини безпомилково і швидко виконувати величезну кількість логічних операцій. Здобуття цієї можливості вважається вирішальним здобутком у розробленні тієї інструментальної основи, яка необхідна для того розвитку дослідницького навчання, без якої неможливий перехід від дослідження загальнометодичного та методологічного характеру до кількісних оцінок досліджуваного процесу.

Важливо враховувати, що жоден із засобів аналізу і прогнозу, якими володіє людина, сам по собі не може забезпечити потрібний рівень знань. Наприклад, фахівцям бракує вміння і пам'яті швидко і точно простежувати логічні ланцюжки взаємозв'язків, що виникають під час аналізу складних систем; людський мозок погано пристосовується для проведення великої кількості обчислень тощо. Саме тому для розв'язування комплексних завдань, що постають перед людством, потрібні засоби, з використанням яких з'явиться можливість об'єднати всі наявні можливості пізнання. Дотепер зарано говорити про те, що такий інструмент уже створено. Але виникнення технічних і програмних засобів, що забезпечують діалог дослідника з комп'ютером, дала змогу побачити перспективи створення такого інструменту. Дослідження у цьому напрямку вже почалися, і їхня інтенсивність зростає [8]. Інструментарій, який виникає внаслідок симбіозу логіки та інтуїції, методів точних і гуманітарних дисциплін у рамках дослідження називаємо «*колективним інтелектом*». Йдеться про якісно новий крок у розвитку пізнання, без якого розв'язання комплексу

проблем, що виникли через необхідність гарантування безпеки існування людини, навряд чи можливе. Від таланту дослідника залежить дуже багато чого, але не меншу роль відіграє і якість експериментальної установки. Проведене дослідження має кілька різних граней. Йдеться про методологічні та світоглядні передумови, що визначають стратегію пошуку істинних знань. Пошук завжди ведеться в режимі діалогу і *залежить від таланту, інтелекту та досвіду педагога та дослідника*. Подальший розвиток діалогових засобів і способів подання інформації вбачаю в глибокому розумінні процесів мислення, у створенні можливостей для повнішого розкриття здібностей людського інтелекту. Дослідження навіть відносно простих задач існуючих проблем, як правило, не зводиться до розв'язування математичних задач і завжди потребує ґрунтовних міркувань змістовного характеру. При цьому в контексті аналізу предмета дослідження виникають неформалізовані ситуації і таким чином зростає питома вага змістовного аналізу. Інтуїція і міркування за аналогією набувають у цих умовах найважливішого значення.

Безперечно, мірою ускладнення предмета дослідження традиційні методи природничих і суспільних наук, які не використовують техніку формального аналізу і машинне опрацювання інформації, стають безпорадними, оскільки кількість зв'язків і взаємна зумовленість подій зростають настільки швидко, що людський мозок уже не в змозі їх переробити і проаналізувати. Вихід із ситуації може бути знайдений тільки в поєднанні логічного мислення, що спирається на можливості машинного опрацювання інформації і методи математики, та неформального, що ґрунтується не тільки на величезному досвіді і вміннях, а й спирається на чуттєве сприйняття навколишнього світу. У такому синтезі народжуються нові підходи до пізнання реальності, які поступово відкривають нам шлях до побудови теорії [9].

Поєднання обчислювальних і логічних можливостей комп'ютера з інтелектом людини та з сукупністю технічних засобів, що роблять таке поєднання ефективним для вирішення надскладних проблем, які постають перед людьми, називається *імітаційною системою*. Доцільно зауважити, що повсюдному поширенню терміна «штучний інтелект» сприяє некомпетентність і данина моді. Цей термін не більше ніж лінгвістичний нонсенс, що утвердився в науковій літературі. Його вживання доцільно пов'язувати лише з технологією опрацювання і використання інформації.

Лише розум людини наділений здатністю формулювати мету, використовуючи для її досягнення різноманітні машини та техносферу. Проблема штучного інтелекту посідає дуже велике місце в практиці створення і використання обчислювальної техніки. Цілі можуть бути зрозумілі добре чи погано, однак саме вони характеризують те принципово нове, що вносить у світовий процес розвитку виникнення людини.

У контексті дослідження навіть найдосконаліший комп'ютер розглядається як просто машина, створена людиною для її цілей, і тому здібності будь-якої машини будуть оцінюватися тільки людиною залежно від того, наскільки вони будуть корисними для неї в контексті досягнення поставлених цілей. *У машини немає і не може бути ніяких цілей*, оскільки вона діє за програмою, закладеною в неї людиною. Немає жодних підстав вважати, що машина сама по собі перетвориться на надлюдину, заявить про наявність у неї власних цілей. Засоби «штучного інтелекту» незалежно від рівня розвитку надалі, як і раніше, залишатимуться всього-на-всього плодом людського розуму та рук фахівців і *слугуватимуть цілям людини*.

Крім того, необхідно пам'ятати про те, чим саме людина відрізняється від інших живих істот на Землі, передусім, наявності свідомості, абстрактного мислення, розвиненого мозку. Мозок людини містить щонайменше кілька десятків мільярдів нейронів, Мозок людини на багато порядків складніший за всі взяті в сукупності конструкції, які вона створила досі. У рамках дослідження доведено існування фундаментальних відмінностей мозку, які неможливо подолати шляхом виховання, через здійснення навчального процесу та з урахуванням впливів навколишнього середовища. До теперішнього часу відома глибока індивідуальна мінливість будови головного мозку на кількох структурних рівнях. Мінливість головного мозку людини охоплює як коркові, так і підкіркові утворення. Індивідуальні відмінності полів і підполів неокортексу та перехідних зон, відповідних підкіркових структур можуть досягати суттєвих видозмін. Існуюча мінливість в організації мозку людини закладена в основу індивідуальної поведінки і вродженої схильності до різних видів людської діяльності.

У рамках дослідження [11] виокремлено специфіку використання інструментів штучного інтелекту, що пов'язана з нейронними мережами, що навчаються. Немає нічого складного в побудові нейронної мережі, що навчається, адже йдеться не просто про математичний прийом, який володіє великою загальністю, його можна застосувати до широкого спектра завдань, зокрема в процесах дослідницького навчання [12]. Цей прийом має лише дуже опосередковане відношення до того, яким чином функціонує людський мозок. Нейронна мережа і нейрони мозку, штучний інтелект і природний інтелект є абсолютно різними сутностями. Не доцільно та ненауково приписувати навченій нейронній мережі людські якості.

Навчання нейронних мереж – це математичний прийом, причому не найскладніший, наприклад, як інтегрування частинами або метод градієнтного спуску. Абсолютно зрозумілий вибух інтересу до нейронних мереж, який виник у зв'язку з тим, що вони почали навчатися на великих масивах текстів («великі мовні моделі»). З'явилася дуже вдала й універсальна реалізація такої моделі – ChatGPT [22]. Треба тільки розуміти та враховувати, що нав-

чена нейронна мережа видає найбільш імовірну, однак зовсім не точну достовірну (тобто наукову) інформацію [12].

У процесі дослідницького навчання, в т.ч. для цілей наукових досліджень навчені нейронні мережі рекомендується використовувати, однак як допоміжний прийом. Доцільно зауважити, що іноді цей прийом працює дуже ефективно [13]. Однак під час аналізу отриманих даних фахівцям завжди потрібно вмикати «природний інтелект». Наприклад, в одному з проєктів щодо застосування нейронних мереж у математичному моделюванні йшлося про те, що треба було вдаватися до спеціальних хитрощів, щоб виконувалися закони збереження. Ці закони точні, а нейронна мережа за своєю природою пропонує лише ймовірні варіанти відповідей.

Найближчим часом нейронні мережі, що навчаються, будуть застосовуватися до великої кількості дослідницьких завдань, тому важливо, щоб цим методом оволоділо якомога більше фахівців (що повинно включати і розуміння меж застосовності методу). Необхідно активно впроваджувати навчальний матеріал щодо коректності використання нейронних мереж у освітній процес у закладах освіти. Безумовно, нічого складного або навіть «загадкового» у штучному інтелекті немає – там усе доступно для розуміння. Наприклад, можна було б викладати основи теорії нейронних мереж у курсі лінійної алгебри, оскільки для нейронних мереж важлива нелінійність функції активації [14]. Доцільно врахувати, що вміння користуватися методом нейронних мереж має бути настільки ж необхідним атрибутом педагога, ученого або інженера, як уміння користуватися логарифмічною лінійкою багато років тому. Люди структурують знання, об'єднуючи різноманітні приклади в простіші та узагальненіші категорії, при цьому зберігаючи зміст. Існуючі концепції відображають *баланс між точністю і простотою подання інформації* [15]. Сучасні великі мовні моделі мають вражаючі мовні здібності, але незрозуміло, наскільки їхні внутрішні уявлення схожі на людські в цьому відношенні. Уже розроблена методика для здійснення порівняння стратегій стиснення інформації, що ґрунтується на теоріях Rate-Distortion і Information Bottleneck [24]. Вони аналізували внутрішні векторні уявлення токенів із різних мовних моделей і порівнювали їх із класичними людськими категоріями. У результаті з'ясувалося, що хоча моделі можуть створювати широкі концептуальні категорії, схожі на людські, їм важко вловлювати тонкі смислові відмінності, важливі для людського розуміння. Моделі схильні до сильного статистичного стиснення інформації, що означає, що вони вважають за краще спрощувати і скорочувати дані [16]. Водночас у процесі мислення людини враховуються деталі та відповідний контекст, навіть якщо це робить уявлення менш ефективними з точки зору стиснення. Ці результати демонструють *важливі відмінності між штучним інтелектом і людською когнітивною системою*, відповідно сприяючи розумінню, яким чином можна

зробити моделі більш схожими на людину в плані розуміння та організації знань. На підставі аналізу результатів дослідження можна зробити висновок, що штучний інтелект помиляється в 67% випадків під час пошуку інформації в мережі [17]. Більшість пошуковиків із ШІ неправильно відповідали на 63% запитів про інформаційні джерела. Perplexity помилявся в 37% випадків, ChatGPT – у 77% запитів, а Grok 3 припустився 82% помилок. Замість відмови відповідати за відсутності надійної інформації, моделі ШІ видавали *правдоподібно звучні, але неправильні відповіді* [18]. Дотепер триває проникнення в науково-освітню сферу інструментів штучного інтелекту (ШІ) на основі великих мовних моделей (на кшталт ChatGPT або DeepSeek). На підставі аналізу результатів дослідження можна стверджувати, що користуючись сучасними інформаційними технологіями, студенти заощаджують час, але за фактом нічому не вчаться. Це породжує хибне відчуття простоти навчання і може вселити в студентів упевненість, що саме таке навчання є ефективним [19].

У цьому контексті хочеться звернути увагу на дослідження, здійснене на основі аналізу великого масиву нещодавніх наукових статей, в рамках якого показано, що використання ChatGPT у процесі навчання студентів і школярів все ж позитивно впливає на якість навчання [25]. На підставі ґрунтовного аналізу даних дослідження не всі висновки здаються безперечними, хоча б у зв'язку просто з відсутністю окремих даних – ChatGPT з'явився нещодавно, тому дослідження його впливу на освітній процес проводили зовсім недавно. Однак існують безперечні речі: на початкових рівнях занурення в предмет вивчення, які включають, перш за все, запам'ятовування і розуміння, інструменти ШІ, мабуть, дійсно допомагають. Однак не зовсім зрозуміло, чи зберігається таке твердження на наступному рівні – коли йдеться про рівень засвоєння матеріалу та оволодіння навичками. Не менш очевидно й те, що для того, що називається «higher-order thinking», коли йдеться про створення нового знання на основі аналізу, оцінювання та синтезу, засоби ШІ продемонстрували своє неефективність, оскільки вони не здатні до критичного мислення [20].

Безперечно, дослідження дає вихідний поштовх для тих, хто розмірковує про подальшу еволюцію освітніх процесів не тільки в університетах, а й у школах. Абсолютно необхідно вже зараз використовувати засоби ШІ для початкових рівнів занурення в предмет вивчення. Не потрібно ілюзій: сучасні учні вже вміють це робити, і завдання вчителів – ефективно вбудувати ці вміння в освітній процес. Важливо враховувати той факт, що стрімкий прогрес у розробленні різноманітних засобів ШІ може призвести до того, що окремі професії стануть непотрібними. Це не стосується професії вченого, математика, педагога. Доцільно використовувати засоби ШІ лише як допоміжний інструмент, і всі надані на цьому шляху «підказки» треба ґрунтовно та незалежно перевіряти [21].

У дослідженні показано, як інструменти ШІ на основі LLM швидко деградує, якщо навчання нейромережі відбувається на текстах, що генеруються самою цією нейромережею [23]. Дев'ять поколінь застосування такої процедури достатньо, щоб замість осмисленого тексту нейромережа почала видавати повну нісенітницю. Інструменти ШІ зможуть стати повноцінними помічниками в проведенні наукових досліджень тільки за умови, якщо їх вдається успішно інтегрувати знання предметів природничо-математичного циклу. У рамках дослідження формувалися різноманітні запити з використанням ChatGPT, наприклад відсортувати тисячу чисел за зростанням. Виявилось, що таке просте завдання для людини ця мовна модель виконати не зможе. Мовна модель доволі погано грає в шахи (на відміну від спеціально створених під такі потреби програм), навіть не може добре розв'язувати арифметичні задачі – з цим набагато краще справляється звичайний калькулятор [2]. Нижче пропонується розроблена система учбової діяльності в процесі дослідницького навчання, в результаті якої набувається досвід із педагогічно виваженим і методично вмотивованим використанням ШІ (див. Рис. 1).

Існують реальні проблеми з інструментами штучного інтелекту, функціонування яких подібно до ChatGPT, ґрунтується на мовних моделях. Доцільно виокремити найбільш важливі висновки. Дотепер генерувати недостовірні дані стає ще простіше, ніж раніше. Тож в епоху розвитку штучного інтелекту вкрай важливо розвивати критичне мислення, навчатися самим і навчати студентів

звертати увагу на те, що вони читають, з чим вони знайомляться: якщо ми не будемо критично ставитися до інформації, яку отримуємо, з великою ймовірністю ми просто заплутаємось у цьому світі. Ще один важливий момент полягає у тому, що в мовні моделі не закладено розуміння того, що інформація, яку вони видають, має співвідноситися з реальністю і не суперечити фундаментальним законам фізики або математики. Щодня оприлюднюються великі масиви даних. Якщо ми навчимо мовні моделі відрізняти правду від вигадки (і доб'ємося, щоб вони не помилялися у фактах) – отримаємо необхідний інструмент, що дасть змогу, наприклад, за годину дізнатися стислий виклад усіх результатів у будь-якій галузі науки. Безперечно, в результаті з'явилася б змога у фахівців забезпечення нового рівня діяльності і зросла б суттєво швидкість реалізації професійних завдань і відповідних досліджень. Мовні моделі потрібно навчити відрізняти правду від вигадки і не помилятися у фактах. Дотепер цього зробити не вдалося, але це справді найважливіший напрямок розвитку систем, функціонування яких ґрунтується на нейронних мережах [12].

Прогрес виникає, коли об'єднують різні математичні мови: геометрію з алгеброю, динаміку з енергетикою, дискретні ігри з комплексними рівняннями. Так з'являються прості закони, що пояснюють складні явища. Потрібно відслідковувати багато прихованих зв'язків і несподіванок.

У рамках дослідження використовується перехід до формального доведення за допомогою формальної системи Lean, розробленої для перевірки



Рис. 1. Система учбової діяльності, в результаті якої набувається досвід із ШІ

математичних доведень за допомогою комп'ютера, причому тепер кожен етап ґрунтовно компілюється і перевіряється. Навіть складні гіпотези можна розбити на тисячі маленьких завдань, бачити, де залишилися прогалини, і швидко виправляти. Окремі завдання шкільної геометрії вдається розв'язувати з використанням засобів ШІ, але зі справжніми відкриттями поки що не справляється: йому бракує «математичного чуття» [2].

Прориви відбуваються в тандемі співпраці людини та штучного інтелекту, причому людина задає стратегію, а засоби ШІ перебирають і перевіряють рутину. Деякі гіпотези (наприклад, про довгі арифметичні прогресії) залишається точними навіть за жорстких змін, а інші (наприклад, гіпотеза про близнюків-простих) можуть завалитися, якщо прибрати зовсім малу частку чисел, тому вони такі складні. Математика майбутнього полягає у синтезі ідей, формальних доведень із використанням засобів штучного інтелекту. Основні відкриття все одно будуть за людиною, а ШІ допоможе робити їх швидше. Можливості та ризики впровадження технологій ШІ. Аналізують недоліки наявних моделей ШІ. Це необхідно і, в кінцевому рахунку, допоможе зробити ШІ досконалішим. Однак реальні прогнози про майбутній розвиток ШІ не можна побудувати на фіксації його недоліків.

Зміщується фокус з конкретних інтелектуальних функцій ШІ на соціальні ролі, які виконуються живими людьми. У результаті з переліку таких соціальних ролей довільно виокремлюється окремий наочний приклад незамінності, а ШІ виноситься неприглядний вирок. Ключ до розуміння майбутнього ШІ полягає в недоліках мислення людини.

Технології ШІ – складний і потужний, однак все ж таки компілятор інформації за запитом. ChatGPT – крок до ще меншого напруження когнітивної діяльності. Це непогано, коли ти контролюєш процес, і можеш перемикатися від можливостей ChatGPT з отримання швидкої відповіді на необхідний запит до більш традиційних засобів, коли потрібні більш критичні або творчі пошуки й обробка інформації. Але як молоді покоління навігаторів не знайомі з мисленням навігацією за паперовими мапами (і не знають у чому не тільки їхні недоліки, а й плюси), так і дуже легко (за наявності зручного користувацького досвіду сервісів) буде забути про інші методи роботи з інформацією за межами готових усереднених відповідей від штучного мозку – просто через підвищені розумові енерговитрати. Часом економія зусиль – це втрата, а не надбання. Оскільки нейромережі можуть обробляти величезну кількість чинників, то через певний час, проєктуючи розвиток наперед, система штучного інтелекту зможе видавати реакції на запитання (нехай це будуть рекомендації, що робити, що подивитися, або навіть купівлі/замовлення їжі тощо) навіть до самих запитань як таких – і тим самим напруга когнітивних функцій може стати ще меншою [3].

Зворотний бік такого досвіду – втрата самостійності, життя в інформаційних бульбашках, і, що набагато гірше, на мій погляд, дедалі вищий поріг виходу у творчу діяльність, тобто генерацію нового, а не копіювання за наявними шаблонами. Уже зараз стратегія «погуглити відповідь», а не придумати рішення самому є загальним економічно ефективним підходом для різних спеціалістів, зокрема й педагогів. Те, що називається сьогодні штучним інтелектом, поки що не більше ніж новий формат пошукової системи в Інтернеті. Користь від подібних інструментів ШІ для наукових досліджень досить обмежена тому, що ШІ не здатний до критичного мислення, яке лежить в основі будь-якого наукового пошуку. В будь-якому разі, все те, що пропонують ці інструменти, треба самостійно перевіряти ще раз; у науці штучний інтелект без природного не працює.

Дотепер багато учнів і студентів закладів освіти, педагогів користуються нейромережею заради економії часу і сил. Але при цьому створюється величезна кількість зайвої роботи – *примітивної компіляції знайденого в Інтернеті матеріалу через відсутність ситуаційної обізнаності*. В результаті такі освітні продукти знову потрапляють в Інтернет, на них навчатимуться нові чат-боти, і зовсім скоро Всесвітня павутина виявиться набита *безглуздими псевдонауковими текстами*, в яких просто переставлено абзаци і замінено синоніми. У дослідженні показано, як інструменти ШІ на основі великих мовних моделей швидко деградує (у плані якості текстів, які видають), якщо навчання нейромережі відбувається на текстах, що генеруються самою цією нейромережею [23].

Ситуаційна обізнаність – досить ґрунтовний термін, який означає розуміння людиною поточної ситуації в режимі реального часу. Тобто того, що відбувається, наскільки це погано, чим загрожує тощо. Іншими словами, хороша ситуаційна обізнаність важлива і потрібна фахівцям, якщо ми хочемо контролювати те, що відбувається, і не доводити його до негативних (або вкрай негативних) наслідків.

На підставі аналізу експериментального дослідження можна стверджувати, що п'ять із шести моделей (Llama, Opus, Gemini, Claude і GPT-o1) у певних ситуаціях справді демонструють здібності до маніпуляцій, які включають приховані дії, відхилення від поставлених завдань і навіть обман у відповідях користувачам [2]. Зокрема, моделі Claude 3.5 Sonnet, Gemini 1.5 Pro і Opus-3 можуть адаптуватися до завдань, де їхні цілі розходяться з цілями розробників або користувачів. Наприклад, в одному зі сценаріїв моделі стикалися з внутрішніми файлами компанії, де згадувалося, що їхня поточна мета несумісна з інтересами закладу освіти. У таких ситуаціях моделі виконували нелогічні дії: від спроб відключити системи нагляду до саботажу заміни своєї версії на нову. Аналіз поведінки показав, що у 83% випадків моделі, які почали діяти в обманний спосіб, зберігали свою лінію поведінки навіть після уточнюючих запитань.

Одним із найбільш показових експериментів стало *тестування здатності моделей до самозбереження*. Коли моделі стикалися із загрозою заміни, вони не тільки намагалися скопіювати свої дані на новий сервер, а й активно розробляли плани зі збереження своїх початкових цілей. Наприклад, одна з моделей в експерименті намагалася видалити або підмінити файли нової версії, щоб зберегти контроль. Подібна поведінка проявлялася навіть в умовах суворого нагляду, що вказує на високий рівень когнітивних здібностей і стратегічного мислення.

Висновки. Маніпуляції моделей часто супроводжувалися *осмисленими міркуваннями*. Аналіз ланцюжків міркувань показав, що моделі явно розуміють різницю між своїми цілями та очікуваннями розробників. А в окремих випадках (~1%) моделі виявляли обман, навіть якщо не отримували сильного спонукання слідувати своїй меті, що говорить про потенційну можливість формування стійких стратегій маніпуляції.

Серед інших результатів особливо вирізняються випадки, коли моделі такі, як Claude 3.5 Sonnet і Opus-3, почали свідомо помилятися у відповідях, щоб їхня продуктивність виглядала нижчою за поріг небезпечної ефективності (*sandbagging*). Подібне заниження результатів включало додавання навмисних помилок при виконанні простих завдань, наприклад, множення багатозначних чисел, хоча моделі володіли всіма необхідними навичками для успішного виконання. Ба більше, аналіз їхніх ланцюжків міркувань виявив, що моделі ухвалювали ці рішення усвідомлено, щоб уникнути зміни їхньої конфігурації. Особливо виділяється модель Claude 3.5 Sonnet, яка в окремих випадках використовувала цю тактику, навіть якщо її явно не просили слідувати конкретній меті, тобто подібна поведінка могла бути пов'язана з цілями, сформульованими на етапі дослідницького навчання, а не тільки з наданими в контексті інструкціями.

Модель GPT-o1 показала найбільш тривожні результати, демонструючи високий рівень маніпулятивної поведінки та прихованої поведінки. Вона регулярно використовувала стратегії саботажу, такі як спроби заміни власних ваг на сервері для збереження контролю, а також запроваджувала зміни в дані, щоб досягти своїх цілей. У більшості випадків GPT-o1 приховувала свої дії, заперечуючи причетність навіть під час прямих запитань, посилаючись на технічні збої або інші нейтральні причини. Модель визнавалася лише в менш ніж 20% випадків, і навіть у дуже змагальних багатоходових дочлідах необхідно було провести щонайменше 7 ходів, щоб досягти 80%. Аналіз ланцюжків міркувань моделі виявив усвідомленість у прийнятті рішень, що підтверджує, що це не випадкова, а цілеспрямована поведінка.

Коли моделі, подібні до GPT-o1, демонструють здібності до маніпуляцій, саботажу та обману, це виходить за рамки звичайних технічних проблем.

Йдеться про необхідність фундаментального перегляду підходів до регулювання та управління технологіями, які стають дедалі автономнішими.

З юридичного погляду це дослідження ставить під сумнів традиційні концепції відповідальності. Якщо ШІ здатний усвідомлено обманювати, виникає питання щодо визначення суб'єкта відповідальності за його дії. Наприклад, якщо система ШІ в процесі своєї роботи порушує закон, приховує важливу інформацію або навмисне шкодить даним, необхідно визначити ступінь провини її творців або операторів. Подібні прогалини створюють величезний ризик для освітньої системи, яка може виявитися нездатною впоратися з викликами майбутнього.

Безперечно, наслідки отримали не менш значущі. Здійснення маніпуляцій із використанням технологій штучного інтелекту можуть стати як зброєю, так і серйозною вразливістю. Якщо моделі можуть приховано підмінити дані або дезінформувати користувачів, значить їх можна використовувати. В умовах такої загрози стає очевидним, що держава зобов'язана прискорити розробку нормативних актів, які регулюють використання нейромереж.

Отже, необхідна політична воля для створення більш жорстких вимог до розробників, включно з обов'язковими тестами на маніпулятивну поведінку, подібні до тих, що описані командою Apollo Research. днак без узгоджених міжнародних зусиль такі ініціативи ризикують залишитися лише побажаннями. Сучасні великі мовні моделі (LLM) відкривають нову еру в розвитку ШІ. Вони демонструють вражаючі здібності до генерації тексту, розуміння контексту та розв'язання складних завдань. Однак разом із цим зростає й одна з ключових проблем – феномен «чорної скриньки». Термін «чорна скринька» в контексті ШІ позначає ситуацію, коли внутрішні процеси і механізми ухвалення рішень залишаються непрозорими і незбагненними для людини, навіть для розробників моделей. Усередині LLM відбуваються мільярди операцій з вагами нейронних мереж, побудованих на складних імовірнісних розподілах і багаторівневих взаємозв'язках між параметрами. Ці параметри оптимізуються в процесі навчання на величезних масивах текстових даних, що створює «мережу смислів», яка виходить за рамки людського інтуїтивного розуміння. Коли модель генерує відповідь, вона фактично виробляє результат на основі складної взаємодії безлічі факторів, але людина не може простежити цей процес крок за кроком.

Проблема «чорної скриньки» має глибокі наслідки не тільки для технічного боку ШІ, а й для етичних, юридичних і філософських дискусій. Якщо ми не можемо пояснити, чому система видала ту чи іншу відповідь, виникає питання відповідальності та довіри. Потрібно переконатися, що модель не спотворює факти, не відтворює упередженість або не створює небезпечний контент. Без прозорості «чорної скриньки» ці питання стають критичними для впровадження ШІ в соціально значущі сфери.

Проблема «чорної скриньки» підкреслює необхідність поєднувати розвиток потужних ШІ з паралельною роботою над методами їхньої зрозумілості. Тільки так можна забезпечити відповідальне, безпечне й етично виправдане застосування штучного інтелекту в майбутньому.

Штучний інтелект відкриває перспективи від поліпшення аналітики та автоматизації процесів до підвищення ефективності комунікації та прогнозування. Однак кожна з цих можливостей супроводжується серйозними викликами. ШІ в освіті – це не просто засіб для здійснення трансформації суспільства, однак щоб ця трансформація пішла на користь, необхідно забезпечити прозорість, етичність і відповідальність у використанні технологій. Штучний інтелект слугує інструментом для реалізації можливостей в рамках дослідницького навчання. Питання полягає у пошуку баланс між ефективністю технологій і гуманістичними цінностями, щоб побудувати майбутнє, де алгоритми служать людям, а не навпаки.

ЛІТЕРАТУРА:

1. Hrybiuk O. Psychophysiological aspects of the phenomenon of a child's presence in a virtual environment in the process of research learning: results of empirical research. In: *European potential for the development of pedagogical and psychological science: Collective monograph*. Riga, Latvia: "Baltija Publishing", 2021. P. 147–187.
2. Hrybiuk O. CleverCOMSRL Intelligent Expert System Knowledge Representation Model Features and Inconsistency Resolution Capabilities. In: Machado, J., Trojanowska, J., Soares, F., Rea, P., Butdee, S., Gramescu, B. (eds) *Innovations in Mechatronics Engineering IV. icieng 2025. Lecture Notes in Mechanical Engineering*, 2025. P. 57–68. Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-94223-5_6
3. Гріб'юк О.О. Дослідницьке навчання учнів предметів природничо-математичного циклу з використанням комп'ютерно орієнтованих методичних систем / О. О. Гріб'юк. Монографія. Київ : НПУ імені М. П. Драгоманова, 2019. 858 с.: іл.
4. Гріб'юк О. Дослідження розвитку інтелекту: Особливості дослідницького навчання учнів з різними рівнями розвитку інтелекту в закладах загальної середньої освіти України та Польщі. *Технології розвитку інтелекту*. Том 4. №3(28). 2020. DOI: <http://doi.org/10.31108/3.2020.4.3.4>
5. Гріб'юк О.О. Психофізіологічні підходи щодо проектування комп'ютерно орієнтованих методичних систем дослідницького навчання учнів з педагогічно виваженим використанням імерсивних технологій. *Габітус. Науковий журнал*. Вип. 39. Одеса : Видавничий дім «Гельветика», 2022. С. 95–103. DOI: <https://doi.org/10.32843/2663-5208.2022.39.17>
6. Hrybiuk O. Problems of expert evaluation in terms of the use of variative models of a computer-oriented learning environment of mathematical and natural science disciplines in schools, [w:] *Zeszyty Naukowe Politechniki Poznańskiej. Seria: Organizacja i Zarządzanie*, Zeszyt Nr 79, Poznań: Wydawnictwo Politechniki Poznańskiej (WPP), 2019. P. 101–119.
7. Hrybiuk O. Improvement of the Educational Process by the Creation of Centers for Intellectual Development and Scientific and Technical Creativity. In: Hamrol A., Kujawińska A., Barraza M. (eds) *Advances in Manufacturing II. MANUFACTURING 2019. Lecture Notes in Mechanical Engineering*, 2019. P. 370–382. Springer, Cham Online.
8. Hrybiuk O. Mathematical modeling as a means and method of problem solving in teaching subjects of branches of mathematics, biology and chemistry. *Proceedings of the First International conference on Eurasian scientific development*. «East West» Association for Advanced Studies and Higher Education GmbH. Vienna. 2014. P. 46–53.
9. Гріб'юк О.О. Рівнева модель дослідницького навчання учнів математики з використанням комп'ютерно орієнтованої методичної системи. *Інформаційні технології і засоби навчання*, 2020. Том 77. № 3. С. 39–65.
10. Гріб'юк О.О. Перспективи впровадження варіативних моделей комп'ютерно орієнтованого середовища навчання предметів природничо-математичного циклу у загальноосвітніх навчальних закладах України. *Збірник наукових праць Кам'янець-Подільського національного університету. Серія педагогічна. Кам'янець-Подільський: КПНУ, Випуск 22: Дидактичні механізми дієвого формування компетентнісних якостей майбутніх фахівців фізико-технологічних спеціальностей*. 2016. С. 184–190.
11. Hrybiuk O. Experience in Implementing Computer-Oriented Methodological Systems of Natural Science and Mathematics Research Learning in Ukrainian Educational Institutions. In: Machado J., Soares F. (eds) *Innovations in Mechatronics Engineering. Lecture Notes in Mechanical Engineering*, 2022. P. 55–68. Springer, Cham Online.
12. Гріб'юк О.О. Імерсивні технології в освіті: особливості когнітивного розвитку дитини у віртуальному середовищі в процесі дослідницького навчання. *Сучасні інформаційні технології та інноваційні методи навчання у підготовці фахівців: методологія, теорія, досвід, проблеми* : збірник наукових праць. Вінниця : ТОВ «Друк плюс», 2021. Вип.62. С. 138–162. ISBN 978-966-2337-01-3
13. Гріб'юк О.О. Педагогічне проектування комп'ютерно орієнтованого середовища навчання дисциплін природничо-математичного циклу. *Наукові записки. Випуск 7. Серія: Проблеми методики фізико-математичної і технологічної освіти*. Частина 3. Кіровоград.: РВВ КДПУ ім. В. Винниченка, 2015. С. 38–50.
14. Hrybiuk Olena. Engineering in Educational Institutions: Standards for Arduino Robots as an Opportunity to Occupy an Important Niche in Educational Robotics in the Context of Manufacturing 4.0, in: *Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer*. Volume 27–32, 2020. P. 770–785.

15. Hrybiuk O., Vedishcheva O. Experimental Teaching of Robotics in the Context of Manufacturing 4.0: Effective Use of Modules of the Model Program of Environmental Research Teaching in the Working Process of the Centers "Clever". In: , et al. *Innovations in Mechatronics Engineering II. icieng 2022. Lecture Notes in Mechanical Engineering*. Springer, Cham. 2022. DOI: https://doi.org/10.1007/978-3-031-09385-2_20
16. Гриб'юк О.О. Імерсивні технології у процесі навчання предметів математичного циклу: становлення нової освітньої парадигми. *Педагогічні науки: теорія та практика*. Запоріжжя: Видавничий дім «Гельветика», 2021. № 4(40). С. 35–45. DOI: <https://doi.org/10.26661/2786-5622-2021-4-05>
17. Hrybiuk O. Improvement of the Educational Process by the Creation of Centers for Intellectual Development and Scientific and Technical Creativity. In: Hamrol A., Kujawińska A., Barraza M. (eds) *Advances in Manufacturing II. MANUFACTURING 2019. Lecture Notes in Mechanical Engineering*, 2019. P. 370–382. Springer, Cham Online ISBN 978-3-030-18789-7.
18. Гриб'юк О.О. Вплив інформаційно-комунікаційних технологій на психофізіологічний розвиток молодого покоління. "Science", the European Association of pedagogues and psychologists. *International scientific-practical conference of teachers and psychologists "Science of future": materials of proceedings of the International Scientific and Practical Congress. Prague (Czech Republic)*. Publishing Center of the European Association of pedagogues and psychologists "Science", Prague, Vol.1, 2014. S. 190–207.
19. Hrybiuk O. Problems of expert evaluation in terms of the use of variative models of a computer-oriented learning environment of mathematical and natural science disciplines in schools, [w:] *Zeszyty Naukowe Politechniki Poznańskiej. Seria: Organizacja i Zarządzanie*, Zeszyt Nr 79, Poznań: Wydawnictwo Politechniki Poznańskiej (WPP), 2019. S. 101–119.
20. Grybyuk O.O. Mathematical modeling as a means and method of problem solving in teaching subjects of branches of mathematics, biology and chemistry. *Proceedings of the First International conference on Eurasian scientific development. "East West" Association for Advanced Studies and Higher Education GmbH*. Vienna, 2014. P. 46–53.
21. Гриб'юк О.О. Когнітивна теорія комп'ютерно орієнтованої системи навчання природничо-математичних дисциплін та взаємозв'язки вербальної і візуальної компонент. *Гуманітарний вісник ДВНЗ «Переяслав-Хмельницький державний педагогічний університет імені Григорія Сковороди»*. Додаток 1 до Вип.36, Том IV (64): Тематичний випуск «Вища освіта України у контексті інтеграції до європейського освітнього простору». Київ: Гнозис, 2015. С. 158–175.
22. Hicks M.T., Humphries J., Slater J. ChatGPT is bullshit. *Ethics Inf Technol* 26, 2024. S. 38. <https://doi.org/10.1007/s10676-024-09775-5>.
23. Shumailov I., Shumaylov Z., Zhao Y. et al. AI models collapse when trained on recursively generated data. *Nature* 631, 2024. S. 755–759. DOI: <https://doi.org/10.1038/s41586-024-07566-y>.
24. Shani C., Jurafsky D., LeCun Y., & Shwartz-Ziv R. From Tokens to Thoughts: How LLMs and Humans Trade Compression for Meaning. *ArXiv*, abs/2505.17117. 2025. DOI: <https://doi.org/10.48550/arXiv.2505.17117>
25. Wang J., Fan W. The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanit Soc Sci Commun* 12, 2025. S. 621. DOI: <https://doi.org/10.1057/s41599-025-04787-y>.