

УДК 004.89:002.6

БРАЙЧЕВСЬКИЙ С.М., кандидат фізико-математичних наук**ШТУЧНИЙ ІНТЕЛЕКТ В СФЕРІ ІНФОРМАЦІЙНОГО ПОШУКУ**

DOI..

***Анотація.** В роботі розглядаються питання використання систем зі штучним інтелектом як альтернативи традиційним інформаційно-пошуковим системам. Аналізуються якісні відмінності методології пошуку, зумовлені застосуванням штучного інтелекту, а також викликані ними ризики.*

***Ключеві слова:** інформаційні технології, інформаційний пошук, штучний інтелект.*

***Summary.** The paper examines the use of artificial intelligence systems as an alternative to traditional information search systems. The qualitative differences in search methodology caused by the use of artificial intelligence, as well as the risks caused by them, are analyzed.*

***Keywords:** information technologies, information search, artificial intelligence.*

Постановка проблеми. Протягом останніх років наше суспільство досягло неабияких успіхів в сфері практичного застосування різноманітних інформаційних технологій. Цьому значною мірою сприяла програма цифровізації [1], що продемонструвала високу ефективність. Але поряд з безперечними досягненнями виникають специфічні проблеми, які необхідно так чи інакше вирішувати. Одна з типових проблем, яка фактично існує, полягає в тому, що активне використання автоматизованих та автоматичних інформаційних систем призводить до надмірного зростання обсягів даних, доступних для подальшої обробки. В результаті виникають складнощі в пошуку потрібної інформації. Сутність цього у тому, що її можна отримати, але далеко не завжди очевидно, як це зробити. Власне, ця проблема сама по собі не є новою, проте специфіка її поточного стану полягає в тому, що широке використання інформаційних технологій особами низької кваліфікації зумовлює наповнення інформаційного простору “сміттям”, тобто відомостями, які не мають змістової цінності і тому завідомо не потрібні користувачу.

Отже, проблема інформаційного пошуку зберігає свою актуальність. Одним із напрямків вирішення її вважається використання штучного інтелекту (далі – ШІ), і цей підхід, безперечно, має сенс. Справжньої популярності ШІ набув останнім часом у зв'язку з тим, що він став доступним широкому колу споживачів. Окрім простоти у використанні, він підкупає своєю ефективністю – користувач отримує саме те, що йому потрібно. Принаймні, так йому здається. Але насправді ситуація набагато складніша, і результати роботи ШІ можуть виявитись далеко не такими, як очікувалося. Зокрема, вони можуть виявитись просто хибними, і найгірше те, що у користувача відсутні надійні засоби верифікації отриманих даних. Існує ціла низка специфічних особливостей інформаційного пошуку з використанням ШІ, які насправді не спрощують, а ускладнюють вирішення конкретних проблем.

Разом з тим, неможливо заперечувати той очевидний факт, що доступ широких верств населення до ШІ є одним з найвидатніших досягнень першої половини ХХІ ст. і немає сумнівів в тому, що його використання набуде характеру загальноприйнятої практики. Тому, на нашу думку, важливим і цікавим є вивчення специфічних особливостей роботи штучного інтелекту, зокрема в сфері інформаційного пошуку. Це

дозволить свідомо використовувати цю потужну технологію з мінімальними ризиками для вирішення основної проблеми.

Нижче проаналізуємо деякі суттєві особливості інформаційного пошуку, порівнюючи технологію традиційного пошуку (далі – ТП) з використанням ШІ. Причому зосередимо увагу не на технологічному, а на змістовному аспекті застосування ШІ. Тобто нас цікавитиме не сама процедура пошуку, а подальше опрацювання отриманих результатів з точки зору інформації, яка в них міститиметься.

Результати аналізу наукових публікацій. Темі ШІ присвячено надзвичайно широке коло публікацій – від фахових статей технічного характеру до ґрунтовних монографій філософського характеру. Основні відомості, важливі для нашої мети, можна знайти, наприклад в [4 – 8].

Метою статті є виявлення позитивних та негативних характерних особливостей процесів традиційного інформаційного пошуку та пошуку з використанням штучного інтелекту на підставі їх порівняльного аналізу.

Виклад основного матеріалу. Відразу зробимо суттєве застереження. В роботі не будемо торкатись загальних проблем ШІ. Предметом обговорення буде не ШІ в науковому розумінні, а конкретні продукти, призначені для використання кінцевим споживачем. Останнім часом саме вони формують в повсякденному житті поняття “штучний інтелект” і саме вони швидко входять в практичну діяльність в найрізноманітніших сферах нашого суспільства. Зазначимо також, що ці продукти фактично реалізують порівняно незначний обсяг функціональних можливостей, доступних ШІ в тому розумінні, в якому цей термін використовується в інформаційних технологіях. На цей час, наскільки можемо судити, найбільш вживаними є два таких продукти: ChatGPT [2] та Microsoft Copilot [3].

Що стосується ТП, то в рамках пропонованої роботи також будемо говорити не про інформаційний пошук взагалі як такий, а про стандартний набір процедур, який використовується в процесі роботи з мережею Інтернет. І робитимемо це на прикладі звичного для нас всіх використання інформаційно-пошукової системи Google.

Спочатку зробимо стислий огляд методології ТП. Хоча ми займаємось пошуком мало не щодня, методологічні аспекти того, що робимо, як правило залишаються неусвідомленими. Іншими словами, ми далеко не завжди усвідомлюємо, що власне робимо, звертаючись до пошукової системи. На практиці ця обставина не має жодного значення, оскільки ми виробили, якщо можна так сказати, режим отримання потрібних для роботи відомостей. Не завжди отримуємо все, що треба, але в будь-якому випадку знаємо, як працювати з отриманою інформацією. Але якщо ми хочемо порівняти ТП з використанням ШІ, необхідно усвідомити відмінність, яка існує між ними.

Ключовою характерною рисою ТП (яка часто не помічається) є те, що ми як правило не отримуємо відповіді на питання, заради якого вдаємось до пошуку. Ми отримуємо “сирі” відомості, які для розуміння слід певним чином опрацювати. Якщо цікавить, наприклад, яка столиця Гондурасу, то Google дасть нам конкретну остаточну відповідь (Тегусигальпа). Але для того, щоб розібратися в поточній економічній ситуації в Гондурасі, потрібно буде опрацювати значний обсяг матеріалів, які надасть Google. На запит “економічна ситуація в Гондурасі” маємо список посилань на різноманітні документи, які, можливо, мають відношення до економіки Гондурасу. А можливо, й ні. Нам доведеться переглядати ці документи і вилучати з них все, що нам необхідно. Наведемо перші три заголовки зі списку, який видає нам Google:

- Економіка Гондурасу (Вікіпедія);
- Парадокс Гондурасу: більшість населення не вважає... (Global Voices);

- Гондурас (Укрінформ).

Також пропонуються додаткові запити, що можуть бути корисними в нашому пошуку:

- Що експортує Гондурас?

- Що вирощує Гондурас?

Наданих документів явно недостатньо для повноцінного огляду економічної ситуації. Тому доведеться продовжувати пошук, доки не отримаємо все, що можливо в межах реальних можливостей. Можливо, нам пощастить, і пошукова система швидко запропонує нам ґрунтовний огляд, який міститиме всі основні відомості, потрібні нам для роботи. Тоді залишиться лише доопрацювання щодо деталей і додаткового фактичного матеріалу. Але може статись і так, що доведеться багато разів повторювати пошук з різними запитами (“торгівля статистика Гондурас” і таке інше), вдосконалюючи та конкретизуючи їх. З отриманих матеріалів ми вилучатимемо окремі фрагментарні відомості, а потім, коли фактичний матеріал буде зібрано, почнеться процес систематизації фрагментів та компіляції їх в цілісний текст. І, підкреслимо це, даний процес ми здійснюватимемо за допомогою нашого природного інтелекту. Без нього інформаційний пошук, як правило, просто позбавлений сенсу.

Надзвичайно важливою є та обставина, що в процесі обробки відібраних пошуковою системою даних ми застосовуємо критичне мислення, навіть якщо не фіксуємо на цьому увагу. Наприклад, відомості, отримані в процесі пошуку з різних джерел, можуть виявитись суперечливими, і доведеться з'ясувати, що з них є правильним, а що – хибним. Інтелектуальне опрацювання результатів ТП містить багато різних складових, але перше місце, безперечно, займає верифікація того, що отримуємо. Тому що, на жаль, в мережі міститься багато сумнівних та хибних відомостей, які не так просто ідентифікувати за формальними ознаками. На кожному кроці доводиться перевіряти, чи відповідає та чи інша інформація дійсності. Іноді цей процес виявляється більш складним і витратним, ніж власне пошук матеріалів, при чому нерідко виникає необхідність виходити за межі роботи з мережними даними (консультації фахівців, використання друкованих джерел тощо).

Позитивний бік справи полягає в тому, що здебільшого, доклавши належних зусиль, можна зрештою отримати все необхідне для вирішення початкової проблеми. Адже наш інтелект допомагає нам ефективно керувати процесом пошуку, розширюючи коло джерел і критично оцінюючи їхній вміст.

Сама ж технологія пошуку доволі проста. Система відбирає релевантні документи і надає їх користувачеві. Нагадаємо, що релевантним називають документ, який містить пошукові терміни запиту (наскільки він корисний споживачеві значення не має). Що з ними робитиме користувач – його особиста справа. Ця специфіка ТП зумовлює структуру пошукового запиту – він формується не на природній мові, а являє собою певним чином організований набір пошукових термінів. Побудова ефективних запитів – це свого роду мистецтво, що вимагає фахових знань і досвіду. Тому часто споживачі звертаються по допомогу до кваліфікованих експертів. Вони лише окреслюють загальну тему пошуку, а експерт будує набір запитів і працює з ними, модифікуючи і розширюючи їх.

Отже, в цілому традиційний пошук включає в себе три фази (етапи):

- побудова пошукового запиту;
- отримання вибірки релевантних документів;
- інтелектуальна обробка документів.

І, як впливає зі сказаного, ці операції часто доводиться повторювати багато разів.

Ще раз підкреслимо принциповий момент: виконуючи ТП, ми отримуємо лише матеріал, з якого за допомогою власного інтелекту будуємо остаточний інформаційний блок, що є відповіддю на початкове запитання. Про саме запитання комп'ютер нічого "не знає", він лише знаходить тексти, які містять пошукові терміни з нашого запиту. Подальша робота – наша компетенція і наша відповідальність. Ми очікуємо, що отримані дані міститимуть певні відомості, на підставі яких ми зможемо зробити належні висновки і так дізнатися те, що нас цікавить.

Отже, ТП не вирішує проблему, а лише надає нам матеріал, з яким нам належить працювати самим. Це, як ми побачимо, має свої недоліки, але й свої переваги.

Пошук з використанням ШІ принципово відрізняється тим, що ми безпосередньо (в явному вигляді) задаємо йому основне запитання, що цікавить, і робимо це природною мовою. Відповідно, результат його роботи являє собою готову відповідь на наше запитання, і вона також представлена природною мовою. В ідеальному випадку після цього взагалі не треба нічого робити, оскільки ми відразу отримали те, що нам було потрібно. Відбувається все дуже швидко. Головні витрати часу – це формулювання та набір на клавіатурі тексту запитання.

На перший погляд, сказане виглядає банальним, але насправді має глибокий зміст, який більшість користувачів тією чи іншою мірою ігнорує. *Головний аспект проблеми, на нашу думку, полягає в тому, що такий режим роботи значно зменшує, а часто просто виключає критичне осмислення результатів пошуку.* Особливо це стосується тих, хто схильний переоцінювати можливості ШІ і повністю довіряє результатам його роботи. Наслідком може стати використання цілком осмислених і правдоподібних, але при цьому необґрунтованих, часто просто хибних тверджень, які споживачем сприймаються як достовірна інформація.

Подивимось, як ШІ здійснює пошук, звісно в гранично спрощеному вигляді.

Перш за все, слід сказати, що не варто переоцінювати значення слова "інтелект" в даному контексті. В будь-якому випадку комп'ютер не розуміє запитання в тому сенсі, в якому його розуміємо ми. Фундаментальна відмінність в технології пошуку полягає в тому, що ШІ при відборі інформації не обмежується вимогою релевантності документів, а використовує певним чином побудовані набори семантичних зв'язків між структурними елементами текстів. Семантика використовується двічі: при аналізі запиту, тобто початкового запитання (природною мовою) і при побудові кінцевого тексту, тобто відповіді, (теж природною мовою). Власне пошук являє собою маніпуляцію нормалізованими патернами^{*}, доступними ШІ, з використанням сформованого комплексу семантичних зв'язків. В сучасних системах ШІ основним джерелом цих патернів є контент соціальних мереж.

^{*} *Bid red.* Патерн (англ. *pattern* – "зразок, шаблон; форма, модель; схема, діаграма") – схема-образ, що діє як посередник уявлення, або чуттєве поняття, завдяки якому в режимі одночасності сприйняття і мислення виявляються закономірності, як вони існують у природі й суспільстві. Патерн (інформатика) – спосіб вирішення характерних задач проектування, зокрема проектування комп'ютерних програм [URL: <https://uk.wikipedia.org/wiki/%D0%9F%D0%B0%D1%82%D0%B5%D1%80%D0%BD>].

Патерн часто плутають з алгоритмом, адже обидва поняття описують типові рішення відомих проблем. Якщо алгоритм – це чіткий набір дій, то патерн – це високорівневий опис рішення, реалізація якого може відрізнитися у двох різних програмах [URL: <https://refactoring.guru/uk/design-patterns/what-is-pattern>].

Так на запитання – “яка економічна ситуація в Гондурасі”, (в даному випадку використовувався Microsoft Copilot) пропонується нам текст, початок якого наводимо – “економічна ситуація в Гондурасі залишається складною. Країна стикається з високим рівнем безробіття, економічною кризою та злочинністю [1]. Гондурас є аграрною країною, де основні галузі промисловості включають текстильну, цукрову, харчову та нафтопереробну” [2; 3].

Отже, типовий сеанс пошуку за допомогою ШІ складається з двох фаз:

- побудова запитання природною мовою;
- отримання результату у вигляді цілісного тексту теж природною мовою.

Фаза вилучення з результату пошуку потрібних для вирішення основної проблеми інформаційних блоків, подальшої їх обробки та компіляції в кінцевий цілісний текст в цьому випадку відсутня. Таким чином, ШІ звільняє від великого обсягу технічної роботи. І може скластися враження, що цю роботу він бере на себе. Але це хибне уявлення. Насправді ШІ робить зовсім інші речі, причому так, щоб ми сприйняли результат його роботи як змістовну відповідь на поставлене запитання. Взагалі кажучи, те, що отримуємо від ШІ, є формальною відповіддю, тобто нам здається, що вона має реальний зміст лише тому, що містить очікувані слова та відповідає синтаксичним правилам.

В спрощеному викладі можемо сказати, що ШІ на основі аналізу запиту буде набір патернів, за допомогою яких він вибирає з доступного набору даних текстові фрагменти, які відповідають поставленій задачі. Але на відміну від ТП відбір здійснюється за допомогою методів, набагато складніших, ніж пошук релевантних текстів. Так наприклад, враховується не лише відповідність фрагментів заданій темі, але й смислова узгодженість окремих фрагментів. З них, взятих в певній послідовності, формується набір даних, який після лінгвістичної обробки перетворюється на кінцевий текст.

В більшості випадків сформований ШІ текст містить в собі корисні для роботи користувача відомості, але це далеко не завжди так. Також слід враховувати, що такий текст може одночасно містити і достовірні, і сумнівні і повністю хибні відомості. Так, в одному з оглядів, створеному ШІ, стверджувалось, що Крим є регіоном РФ. Причина: ШІ скористався джерелом, яке за передбаченими розробниками формальними ознаками було сприйняте як надійне. На жаль, формальні ознаки не гарантують дійсну надійність мережних джерел.

Найгірше те, що в нашому розпорядженні немає надійних інструментальних засобів, за допомогою яких могли б відрізнити одне від іншого.

Таким чином, ми ніколи не маємо певності в тому що текст, отриманий від ШІ, є повністю достовірним. Логіку викладу та структурно-лінгвістичні характеристики ми можемо оцінити самі (це доступно кожній освіченій людині). Фактографічну ж достовірність, на жаль, визначити важко. Адже ми отримуємо невідому нам інформацію, інакше немає потреби вдаватися до інформаційного пошуку.

Тому важко не погодитися з тими, хто дивиться на матеріали, які видає ШІ, як на проміжні техніко-технологічні дані, що потребують подальшої специфічної обробки.

На завершення наведемо ще одну можливість ТП, практично відсутню у ШІ.

Традиційна пошукова система видає великий обсяг даних, переважна більшість яких нам може бути не потрібна. Але іноді серед них трапляється інформація, на яку ми не розраховували, але яка виявляється цікавою і важливою. При чому не обов’язково в плані вирішення поставленої проблеми. ТП часто дає нам значно більше, ніж ми планували одержати.

Висновки.

ШІ швидко входить в наше повсякденне життя, в першу чергу в професійну діяльність. І до цього потрібно адаптуватись.

В сфері інформаційного пошуку ШІ поступово займає своє місце, і далеко не завжди за вибором користувача. Багато пошукових систем починають інтегрувати елементи ШІ в свої пошукові алгоритми, при чому користувачі здебільшого про це навіть не здогадуються. Існують вже інтегровані системи пошуку, в яких повноцінно застосовується ШІ. З найвідоміших таких систем назовемо You.com [9], Perplexity [10] та Andi Search [11]. Тому ризики, що несе в собі ШІ, хочемо ми цього чи ні, стають частиною нашої практики. І ми повинні свідомо підходити до використання сучасних інформаційних технологій.

ШІ пропонує нам зручність і швидкість.

Він звільняє нас від значних витрат часу і сил, які у випадку ТП йдуть на обробку отриманого матеріалу. Але при цьому процес обробки “сирих” даних стає неконтрольованим і ви ніколи не маєте певності в достовірності результатів. Тому відповідальне використання ШІ неодмінно повинно включати процедуру верифікації отриманих відомостей, як на фактографічному, так і на семантичному рівні. Що важливіше для конкретного споживача в рамках конкретної задачі, вирішує він сам.

Використана література

1. Цифрова трансформація. URL: <https://www.kmu.gov.ua/diyalnist/mizhnarodna-dopomoga/coordination/cifrova-transformaciya> (дата звернення: 06.12.2024).
2. ChatGPT – Українська версія. URL: <https://chatgpt.com/g/g-kKgesLZ4G-ukrainska-versiia> (дата звернення: 06.12.2024).
3. Особистий помічник із ШІ Microsoft Copilot. URL: <https://www.microsoft.com/uk-ua/microsoft-copilot/personal-ai-assistant> (дата звернення: 15.11.2024)
4. Баранов О.А. Визначення терміну “штучний інтелект”. *Інформація і право*. № 1(44)/2023. С. 32-49.
5. Баранов О.А. Інтернет речей і штучний інтелект: витоки проблеми правового регулювання: збірник матеріалів II-ї Міжнародної науково-практичної конференції *IT-право: проблеми та перспективи розвитку в Україні*, м. Львів, 17 лист. 2017 р. Львів: НУ “Львівська політехніка”, 2017. 318 с.
6. Beyer, H. et al. 2016. Solving conservation planning problems with integer linear programming. *Ecological Modelling* 328: 14-22.
7. Haenlein, Michael and Andreas M. Kaplan. 2019. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review* 61: 14-5.
8. Raj, M. and R. Seamans. 2019. Primer on artificial intelligence and robotics. *J. Org. Design* 8, 11: 1-14.
9. You.com. URL: <https://you.com> (дата звернення: 06.12.2024).
10. Perplexity. URL: <https://www.perplexity.ai> (дата звернення: 06.12.2024).
11. Andi Search. URL: <https://andisearch.com> (дата звернення: 06.12.2024).

~~~~~ \* \* \* ~~~~~