

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра математики факультету інформатики

Курсова робота
за спеціальністю „Прикладна математика” 113

**AI for text generation: Review of the main algorithms and frameworks for text
generation**

Керівник курсової роботи

Курочкін Андрій Володимирович

Сумісник асистент.

(підпис)

Виконала студентка 3-го курсу
спеціальності «Прикладна математика»

Кузьомко Ярина Романівна

«__» _____ 2023 р.

Київ – 2023

Міністерство освіти і науки України

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Кафедра інформатики факультету інформатики

ЗАТВЕРДЖУЮ
Зав.кафедри математики,
проф., д.ф.-м.н.
_____ Р. К. Чорней
(підпис)
«__» _____ 2023 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ
на курсову роботу

студенту Кузьомко Ярині Романівні факультету інформатики 3 курсу
ТЕМА AI для генерації тексту: огляд основних алгоритмів і фреймворків для
генерації тексту

Зміст ТЧ до курсової роботи:

Індивідуальне завдання

Анотація

Вступ

Розділ 1. Основні алгоритми генерації тексту

Розділ 2. Огляд GTP-3: можливості, перспективи та загрози розвитку

Розділ 3. Дослідження здібностей сучасного штучного інтелекту на
основі українського національного мультипредметного тесту

Висновки

Список літератури

Дата видачі „25” вересня 2022 р. Керівник _____

(підпис)

Завдання отримав _____

(підпис)

Тема: AI для генерації тексту: огляд основних алгоритмів і фреймворків для генерації тексту

Календарний план виконання роботи:

№ п/п	Назва етапу курсової роботи	Термін виконання етапу	Примітка
1.	Отримання завдання на курсову роботу.	25.09.2022	
2.	Обговорення деталей теми на розробка плану виконання	8.12.2022	
3.	Огляд технічної літератури за темою роботи та опрацювання матеріалів	26.01.2023	
4.	Написання теоретичної частини роботи	квітень-травень 2023	
5.	Розробка практичної частини та фіналізація практичної	10.05.2023	
6.	Аналіз отриманих результатів, обговорення остаточної доповіді з керівником	13.05.2023	
7.	Здача курсової роботи	23.05.2023	

Студент: Кузьомко Я.Р.

Керівник: Курочкін А.В.

“25” вересня 2022

ЗМІСТ

ПЕРЕЛІК ПРИЙНЯТИХ СКОРОЧЕНЬ	5
АНОТАЦІЯ	6
ВСТУП.....	7
РОЗДІЛ 1. Основні алгоритми генерації тексту	9
1.1 . Марковські моделі генерації тексту	9
1.2. Рекурентні нейронні мережі (RNN)	10
1.3. Довга короткочасна пам'ять (LSTM).....	11
1.4. Генеративні змагальні моделі (GAN).....	13
1.5. Generative pre-trained transformer (GPT).....	15
РОЗДІЛ 2. Огляд GTP-3: можливості, перспективи та загрози розвитку	20
2.1. Можливості та перспективи розвитку AI	20
2.2. Загрози розвитку штучного інтелекту	26
РОЗДІЛ 3. Дослідження здібностей сучасного штучного інтелекту на основі українського національного мультипредметного тесту	29
3.1. Проходження НМТ з допомогою ChatGPT-3.5.....	29
3.2. Аналіз результатів тесту.....	33
3.3. Порівняння результатів тесту з результатами учасників НМТ 2022 ...	34
ВИСНОВКИ.....	40
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	41

ПЕРЕЛІК ПРИЙНЯТИХ СКОРОЧЕНЬ

AI – Artificial Intelligence;

GPT – Generative Pre-trained Transformer;

RNN – Recurrent neural network;

LSTM – Long short-term memory;

GAN – Generative adversarial network;

ШІ – штучний інтелект;

НМТ – національний мультипредметний тест;

ВНЗ – вищий навчальний заклад.

АНОТАЦІЯ

Курсова робота присвячена дослідженню основних алгоритмів генерації тексту, зокрема Марковським моделям, рекурентним нейронним мережам (RNN), довгочасним короткочасним пам'яттям (LSTM), генеративним змагальним моделям (GAN) та generative pre-trained transformer (GPT). Вона також охоплює огляд GPT-3 та дослідження його можливостей, перспектив розвитку та загроз. Крім того, проведено дослідження здібностей сучасного штучного інтелекту на основі українського національного мультипредметного тесту, використовуючи ChatGPT-3.5 та проведено аналіз результатів тесту, порівняння їх з результатами учасників НМТ 2022.

Ключові слова: алгоритми генерації тексту, Марковські моделі, RNN, LSTM, GAN, GPT-3, штучний інтелект.

ВСТУП

Сфера обробки природної мови (NPL) є однією із найцікавіших та найактуальніших в сьогодні, оскільки вона дозволяє покращити та розвинути процес “комунікації” між людиною та комп’ютерною системою. В останні роки дослідження в сфері штучного інтелекту (AI) досягли неабиякого прогресу. Це посприяло розробці та створенню різноманітних систем, фреймворків та алгоритмів, які з високою точністю та якістю можуть генерувати текст подібний до людського.

Метою цього дослідження є огляд основних алгоритмів та фреймворків, що використовуються в AI для генерації тексту, визначення їх переваг, недоліків, перспектив та небезпек подальшого розвитку, приклади практичного використання.

GPT-3 (Generative Pre-trained Transformer 3) - потужна модель для генерації мови, що була тренувана на 175 мільярдах різних параметрів. Таким чином вона здатна виявляти закономірності в мовленні, аналізуючи великі масиви текстових даних, та використовувати це для створення зв’язного тексту, що в свою чергу сприяє виконанню різноманітного роду завдань - від створення художньої літератури, поезії та прес-релізів до функціонального коду, а також музики, жартів, технічні посібники або новинних статей, які важко відрізнити від написаних людиною [1]. Девід Чалмерс, австралійський філософ, написав у своїй статті[2]: “GPT-3, здається, ближче до проходження тесту Тюрінга, ніж будь-яка інша система на сьогоднішній день (хоча "ближче" не означає "близько)". Саме це і обумовлює актуальність моєї роботи.

Предметом дослідження є GPT-3 - сучасна модель генерації тексту, її історія створення, тренування моделі для генерації тексту, її застосування в різних аспектах життя, конкуренти, переваги та недоліки, а також майбутній розвиток та безпеки у використанні.

Об'єктом дослідження є огляд сучасних процесів, методів та алгоритмів, що використовуються для генерації тексту, їх тренування, застосування та обмеження.

Робота складається з трьох розділів. У першому розглядаються основні алгоритми генерації тексту. Другий розділ присвячений детальному огляду однієї із найпопулярніших технологій генерації тексту – ChatGPT-3, можливостям, перспективам та проблемам розвитку. Третій розділ є практичною частиною роботи в якій досліджується як ChatGPT-3.5 може впоратись із НМТ 2022 року та аналіз отриманих результатів.

РОЗДІЛ 1. Основні алгоритми генерації тексту

1.1. Марковські моделі генерації тексту

Ланцюги Маркова були відкриті у 1906 році Андрієм Марковим. Він говорив про своє відкриття так: “Зазначимо, що будь-яка система, в якій перехід з одного стану в інший не залежить від передісторії процесу, а залежить тільки від поточного стану, називається в теорії ймовірностей Марківським ланцюгом або ланцюгом Маркова” або іншими словами “Стохастичний процес, що містить випадкові величини, які переходять з одного стану в інший залежно від певних припущень і певних імовірнісних правил є Ланцюгом Маркова”. Тобто Ланцюг Маркова це математична модель, яка створена для моделювання випадкових процесів, де майбутній стан незалежний від будь-яких інших станів, окрім поточного.

Властивість Маркова: ймовірність переходу в новий стан залежить лише від поточного, і не залежить від кроків, що призвели до поточного стану. Можемо вивести цю властивість математично. Візьмемо випадковий процес типу $\{X_m, m = 0, 1, 2, \dots\}$. Цей процес буде Ланцюгом Маркова лише у випадку:

$$P(X_{m+1} = j \mid X_m = i, X_{m-1} = i_{m-1}, \dots, X_0 = i_0) = P(X_{m+1} = j \mid X_m = i)$$

– для всіх $m, j, i, i_0, i_1, \dots, i_m - 1$. [3]

Для скінченної кількості станів $A = \{0, 1, 2, \dots, n\}$ процес буде називатися скінченим ланцюгом Маркова.

$P(X_{m+1} = j \mid X_m = i)$ означає, що ймовірність переходу з одного в стану в інший незалежна від значення m , отже рівнянням $P_{ij} = P(X_{m+1} = j \mid X_m = i)$ можна представити Ланцюг Маркова.

Ланцюги Маркова широко використовуються у реальному світі:

- Для генерації текстів, есе, промов. Також часто цей алгоритм застосовується для генерації імен та назв;
- Для підказок та автозаповнення. Алгоритм може передбачати наступні слова у тексті і пропонувати їх користувачу для пришвидшення;

- Google у своїй Google Page Rank розглядає весь інтернет як Ланцюг Маркова, і таким чином визначає фіксовану ймовірність вашого потрапляння на певну сторінку під час користування інтернетом незалежно від вхідної точки серфінгу.

1.2. Рекурентні нейронні мережі (RNN)

На відміну від Ланцюгів Маркова рекурентні нейронні мережі мають “пам’ять” та беруть інформацію з попередніх входів, що впливає на вихідні данні. Тобто, на відміну від традиційних нейронних мереж, в яких всі входи та виходи є незалежними, рекурентні нейронні мережі передають вихідні данні з попереднього кроку у наступний, що дозволяє пам’ятати попередні слова при передбаченні

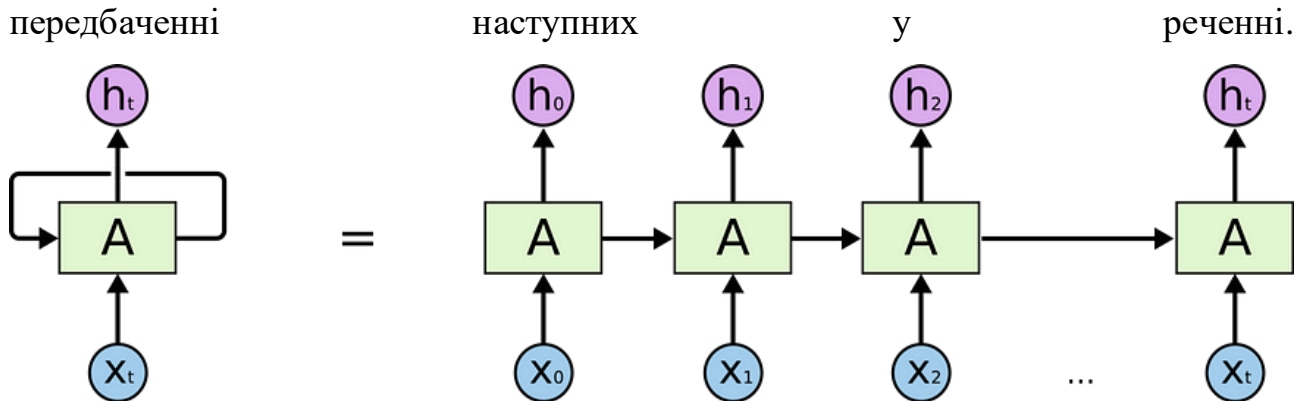


Рис. 1.2.1. Схематичне зображення логіки роботи рекурентних нейронних мереж [9].

Ми можемо вивести таку формулу роботи поточного стану рекурентних нейронних мереж:

$$r_x = F(r_{x-1}, h_x)$$

де r_x – поточний стан; r_{x-1} – попередній стан; h_x – вхідний стан.

Функція активації – це функція, що визначає чи повинен нейрон активуватися, її значення може передаватися наступним нейронам або служити вихідним значенням моделі. Формула активації буде мати вигляд типу:

$$r_x = \tanh(w_{rr}r_{x-1} + w_{xr} h_x)$$

де – w_{rr} – ваги на рекурентному нейроні; w_{xr} – ваги на вхідному нейроні.

Тож формула вихідного результату виглядатиме так:

$$g_x = w_{xg}r_x$$

де – g_x – результат; w_{xg} – ваги вихідного шару; r_x – поточний стан.

Незважаючи на те, що рекурентні нейронні мережі мають і свої мінуси, такі як важкість процесу тренування такої мережі вони широко використовуються у таких сферах як: машинний переклад, розпізнавання мови, розпізнавання зображень, моделювання та створення тексту.

1.3. Довга короткочасна пам'ять (LSTM)

[11] Довга короткочасна пам'ять (LSTM) – це наступний крок розвитку рекурентних нейронних мереж, вона позбувається основних проблем RNN, таких як проблеми зникаючих та вибухаючих градієнтів (рис. 1.3.). RNN не здатна запам'ятовувати інформацію на довгий проміжок часу і схильна до забування попередніх вхідних даних. Також будь-яка інформація додана в RNN змінює уже існуючу в моделі інформацію, оскільки модель не здатна фільтрувати інформацію на потрібну і не потрібну. У той же час LSTM оминає цю проблему і додавання нової інформації не так різко змінює вже існуючу, оскільки в LSTM моделі додано ворота, що визначають в потоці інформації яка інформація є потрібною, а яка – ні і її потрібно видалити.

Існує три види воріт:

- Вхідні ворота;
- Вихідні ворота;
- Ворота видалення (забуття).

Ворота видалення або забуття відповідають за те, яка інформація є корисною і повинна залишитися в моделі, а яка є зайвою і повинна бути видаленою (забутою).

Вхідні ворота відповідають за те, яка інформація потрапить у нейрон, створює масив інформації що повинен бути доданим.

Вихідні ворота відповідають за те, які значення будуть виведені з поточного нейрона.

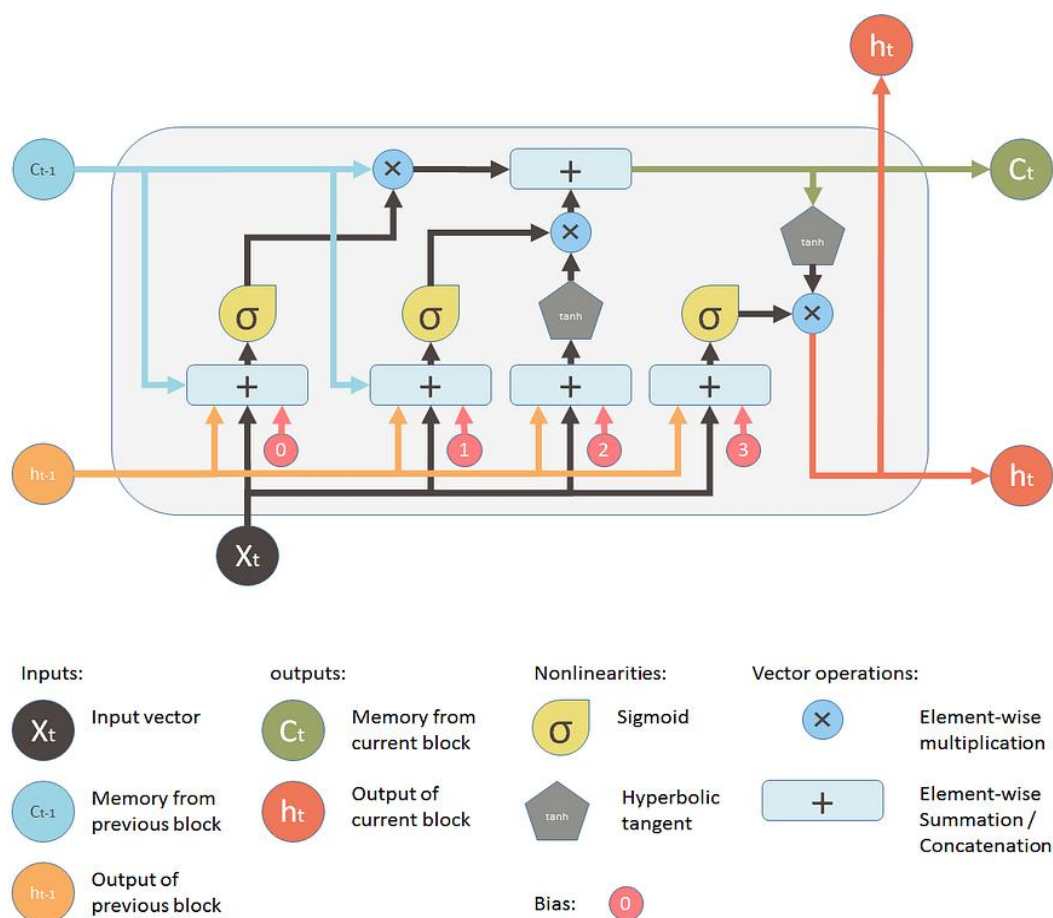


Рис. 1.3.1. Схематичне зображення логіки роботи LSTM систем[11].

Серед плюсів такого підходу:

- Здатність моделювання складних та довготривалих залежностей, завдяки зможі запам'ятовувати та використовувати інформацію з віддалених моментів у часі;
- Гнучкість, таку модель можна натренувати різними даними на велику кількість різноманітних специфічних завдань;
- На відміну від RNN моделей, LSTM не має проблем з витіканням градієнту, завдяки архітектурі з використанням різнотипних воріт, що дозволяють контролювати градієнт та запобігають його зникненню.

Проте, серед мінусів все ще можна виділити потребу у великій кількості даних для навчання такої моделі, та необхідності для тренування серйозних обчислювальних потужностей і об'ємів пам'яті.

1.4. Генеративні змагальні моделі (GAN)

Генеративні змагальні моделі – це модель машинного навчання, де дві моделі змагаються між собою за правилами гри з нульовою сумою за більш точний результат. Створені були у 2014 році Яном Гудфелом з першочерговою метою генерації реалістичних зображень, які важко, або неможливо відрізнити від реальних. Генеративні змагальні моделі складаються з двох складових: дискримінатора та генератора (рис. 1.4.).

Дискримінатор використовується для розпізнавання згенерованих зображень, для цього його окремо потрібно натренувати на виконання потрібної задачі з шуканою точністю. Іншими словами, потрібно надати мережі велику кількість даних, на яких будуть розмічені потрібні частини зображень, наприклад, обличчя людей, після чого сама нейронна мережа повинна навчитися визначати з високою точністю чи є на зображенні обличчя і де воно знаходиться.

Генератор використовується для генерації зображень, спочатку він починає генерувати випадковий шум, з часом у шумі починають проявлятися контури шуканого зображення, шум продовжує генеруватися поки контури зображення не стануть більш чіткими і так допоки не настане шуканий результат. В процесі роботи генератор запам'ятовує ефективні підходи генерації шуму і покращується. У цьому генератору допомагає дискримінатор, який надає оцінку згенерованим зображенням по шкалі від 0 – зображення не відповідає заданим вимогам, воно фальшиве, до 1 – зображення відноситься до набору даних, воно справжнє. Генератор намагаючись покращити свою оцінку буде постійно покращувати згенерований шум і рухатися у правильному напрямку, допоки не досягне максимальної оцінки на яку здатний. [7]

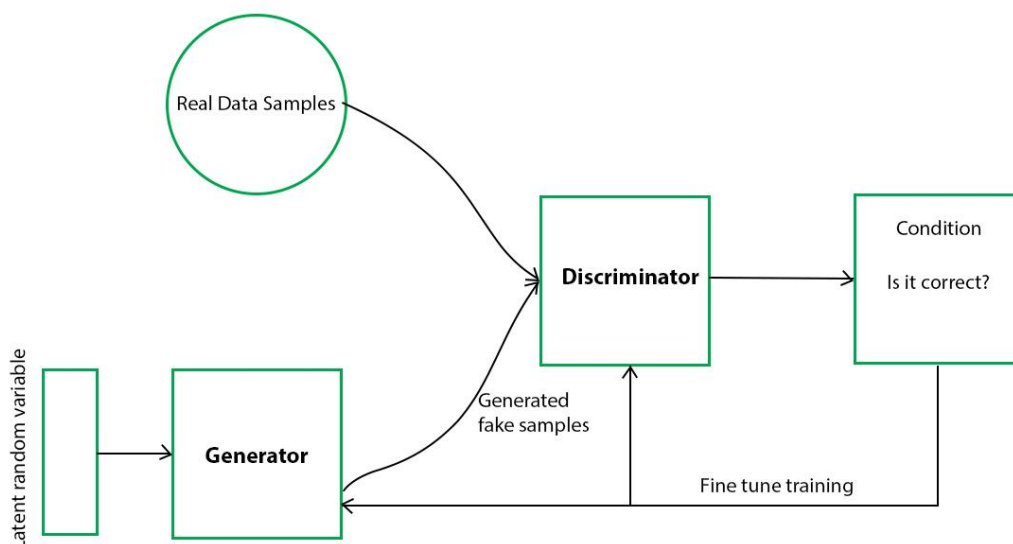


Рис. 1.4.1. [8] Схематичне зображення архітектури та компонентів GAN моделей.

Завдання генератора – максимізувати ймовірність помилки дискримінатора, завдання дискримінатора – визначити чи зображення, що надаються йому на вхід відповідає параметрам даних, на яких було проведене його навчання. Математично цей процес можна описати так:

$$V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))]$$

де – G – Генератор; D – Дискримінатор; $P_{\text{data}(x)}$ - розподіл реальних даних; $P(z)$ - розподіл генератора; x - вибірка з $P_{\text{data}(x)}$; z - вибірка з $P(z)$; $D(x)$ - мережа дискримінатора; $G(z)$ - генераторна мережа.

Такі моделі відрізняються високою реалістичністю вихідних даних, універсальністю, гнучкістю та можливістю навчання моделей на непомічених даних. Проте вони мають і свої слабкі сторони, серед яких важкість та висока вартість навчання таких моделей, також часто такі моделі занадто точно відтворюють навчальні данні, створюючи зображенні занадто синтетичні,

нерізноманітні та подібні на навчальні, а крім цього моделі є непрозорими та важкими для пояснення. [8]

GAN моделі використовуються для: генерації різноманітних зображень з нуля; покращення уже готових зображень; генерація даних для навчання нових моделей; у технологіях доповненої реальності.

1.5. Generative pre-trained transformer (GPT)

Generative pre-trained transformer – різновид архітектури глибоких неймереж вперше створений у 2017 році, що будується на універсальній мовній моделі, що дозволяє нейронним мережам на основі такого підходу виконувати широкий спектр задач. На сьогодні найвідомішою мережею, що використовує цей підхід – є ChatGPT. Моделі на основі GPT – це нейронні мережі трансформери (рис. 1.5.1.), що використовують два модулі для виконання роботи: кодер та декодер.

GPT моделі обробляють вхідний текст створюючи математичні представлення слів. При обробці даних очікується, що слова, які стоять поряд є взаємозв'язаними. Математичні представлення обробляються кодером, він розділяє слова на блоки та присвоює їм ваги, у залежності від релевантності слів. Оскільки кодер враховує розміщення слова у тексті, він здатний розуміти контекст і приділяти різні значення граматично однаковим словам.

Результат роботи кодера передається декодеру.

Задача декодеру “вгадати” необхідний результат. Він зосереджується на окремих блоках, що передав йому кодер, завдяки внутрішньому механізму він здатний визначати точність своїх результатів та модифікувати їх з метою покращення результату та обрання найточнішого.

Тобто, можна сказати, що мережі GPT типу є довершеними версіями автозаповнення, оскільки вони генерують тексти підбираючи слова які з високою ймовірністю мають бути наступними у реченні.

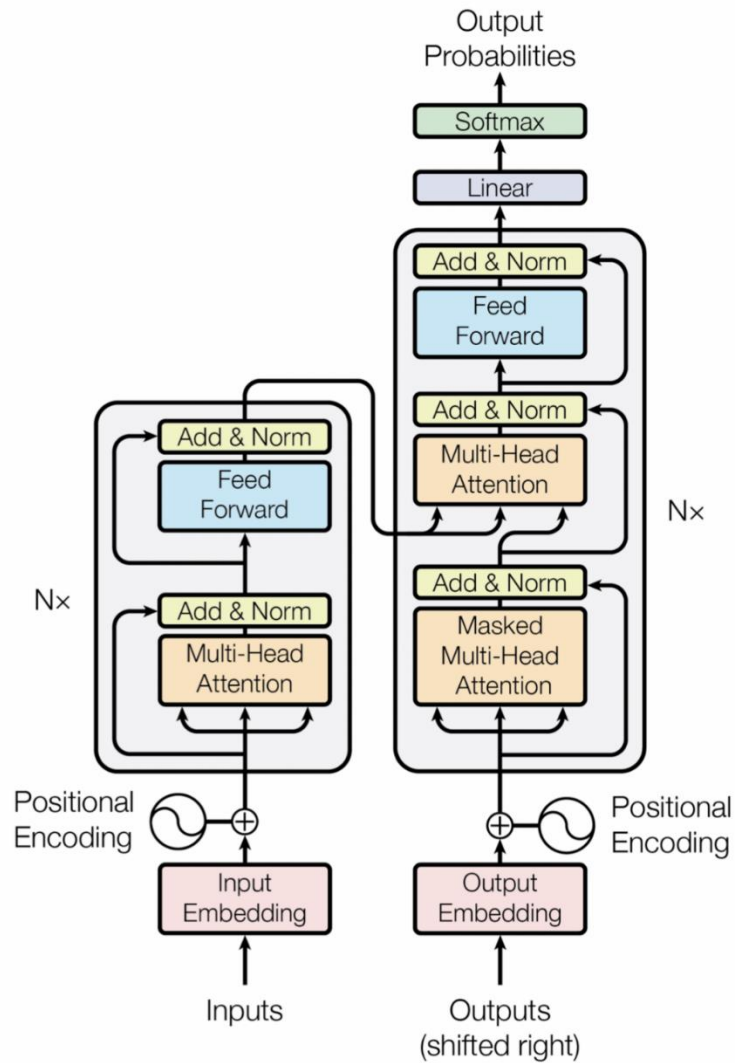


Рис. 1.5.1. Схематичне зображення архітектури трансформера

На відміну від інших популярних видів нейронних мереж, GPT моделі не обробляють слова один за одним, послідовно, а обробляють відразу всі вхідні дані, що дозволяє добитися зразкового рівня вихідних результатів.

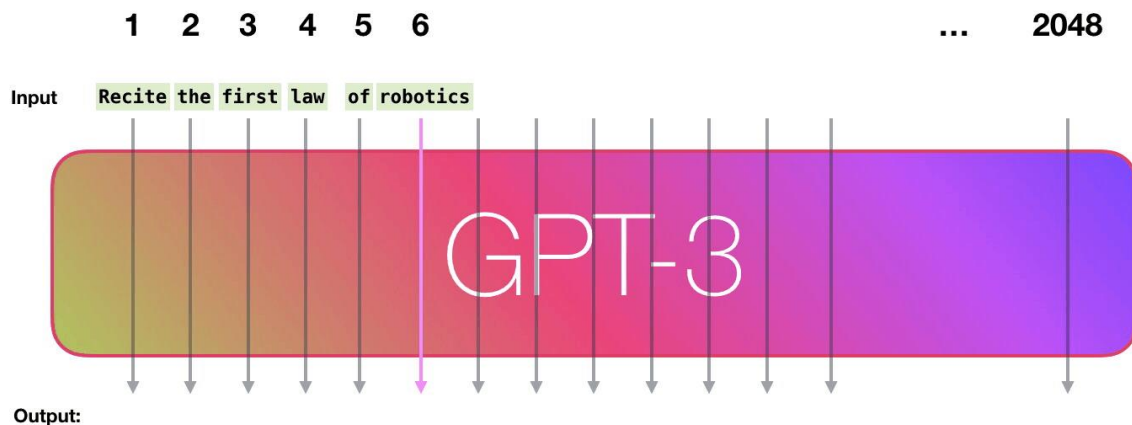


Рис. 1.5.2. Анімація роботи ChatGPT-3 на основі архітектури GPT [18].

GPT використовує механізм самоуваги, що дозволяє вловлювати підтексти та складні взаємозв'язки в наданій інформації. Для цього модель перетворює блоки тексту, – токени, у вектори які відображають його важливість. Для цього генерується ключ, вектор та запит для кожного токена, вираховується схожість між вектором токена та кожним ключовим вектором інших наявних токенів, генеруються нормалізовані ваги використовуючи softmax – функцію, що перетворює вектор з X дійсних значень у вектор з X дійсних значень, сума яких дорівнює 1:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

де – \vec{z} - вхідний вектор; z_i - елементи вхідного вектора; e^{z_j} – експоненціальна функція стандартного типу; $\sum_{j=1}^K$ - член нормалізації, що гарантує, що всі значення функції дорівнюватимуть 1; K - кількістю унікальних категорій або міток, на які можуть бути класифіковані об'єкти.

Після цього створюється фінальний вектор, що визначає важливість токена шляхом множення згенерованих ваг на вектори кожного токена. Модель повторює ці кроки щоразу генеруючи нові дані, що дозволяє їй збільшити точність вихідних даних

Для тренування моделей такого типу потрібні дані великого і надвеликого об'єму. Так, нам відомо, що Open AI третю версію свого ChatGPT натренувала на 175 мільярдах параметрів, це понад 570гб тексту. В основному це дані з книг, вікіпедії та інших веб ресурсів. GPT мережі використовують форму непарного машинного навчання. Це значить, що дані, на яких навчається модель, є неміченими, тобто не містять інформації про правильність чи неправильність зворотного зв'язку, і модель сама повинна знайти залежності між інформацією.

Також GPT мережі здатні враховувати фідбек та підлаштовуватися під нього, що дозволяє значно точніше узгодити результати моделі з тим, чого хоче користувач.

Моделі такого типу стали справжнім проривом в області штучного інтелекту та змогли поширити AI в нових сферах:

- GPT моделі здатні аналізувати тексти та узагальнювати їх;
- GPT моделі здатні до реалістичного та невимушеного спілкування з людиною, що дозволяє використовувати їх наприклад як операторів служб підтримки;
- Створення контенту: сценаріїв, діалогів, описів, статей;
- Медицина, моделі нового типу здатні ставити діагнози з високою точністю опираючись на список симптомів, здатні стати помічниками у роботі лікарів;
- Високоякісний переклад;
- Написання коду.

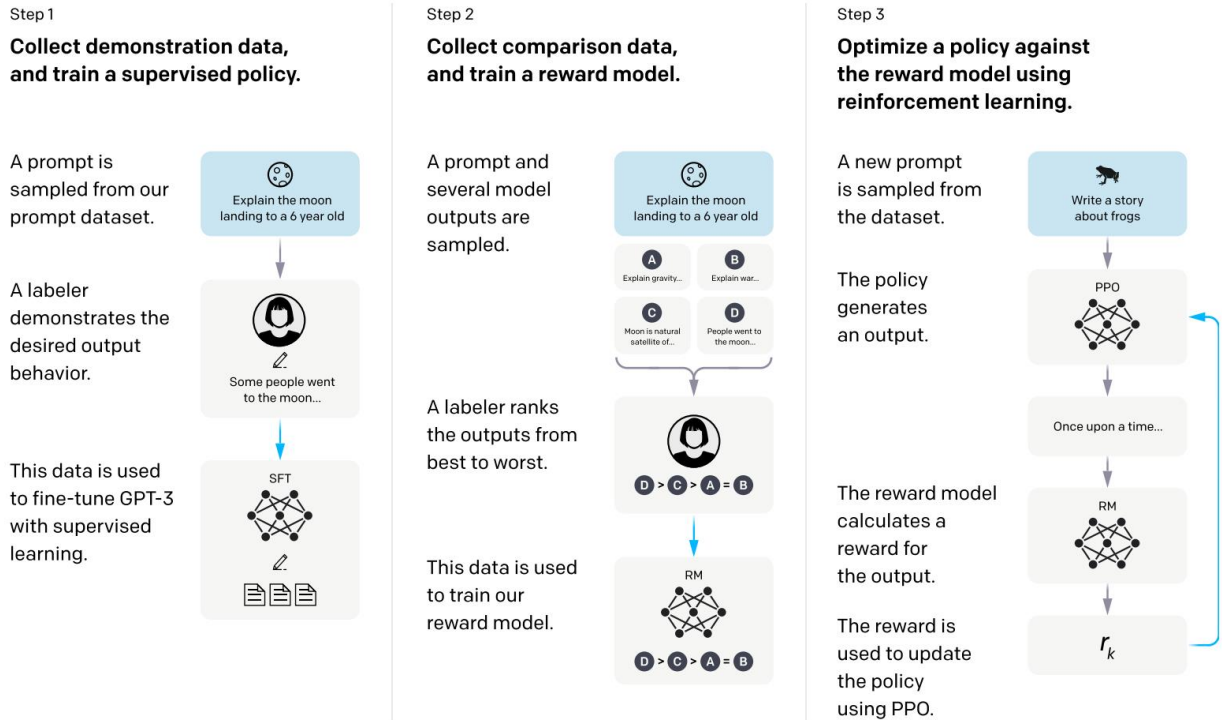


Рис. 1.5.3. Ілюстрація кроків які виконали OpenAI щоб їхня модель – GPT-3 була здатна враховувати фідбек користувачів

Проте, такі моделі все ще не позбавлені мінусів:

- Слабка інтерпретованість, що не дозволяє зрозуміти, як модель мислить і які кроки призвели до конкретного результату;
- Виведення результатів, що містять неіснуючі, видумані факти;
- Упереджені або токсичні відповіді;
- Необхідність величезних обчислювальних потужностей, наприклад, генерація 30 повідомлень відомим ChatGPT потребує близько одного літру води для охолодження системи, а в процесі навчання на охолодження було витрачено понад 700 тисяч літрів води. [17]

Приклади моделей GPT типу – OpenAI GPT, ELMo, CoVe, BERT, Bing

РОЗДІЛ 2. Огляд GPT-3: можливості, перспективи та загрози розвитку

2.1. Можливості та перспективи розвитку AI

Найпопулярнішою та найвідомішою моделлю створеною на основі архітектури *GPT* є OpenAI ChatGPT, який нещодавно отримав оновлення до четвертої версії. Давайте порівняємо можливості актуальної версії ChatGPT-4 з її попередником ChatGPT-3.

Нововведення та покращення у ChatGPT-4 порівняно із ChatGPT-3: [19]

- Здатність приймати у вигляді вхідних даних зображення, що відкриває цілий пласт нових можливих застосувань мережі. Тепер модель здатна пояснювати що зображено на фото, обробляти рукописний текст, виконувати задачі що потребують візуального контексту. Наприклад, можна показати чату рентген людської кінцівки/фотографію частини тіла і отримати у відповідь ймовірний діагноз та поради з лікування. Можна відправити чату інтернет мем і попросити його пояснити, що в ньому смішного або графік з даними і отримати їхній аналіз і висновки (рис. 2.1.2.);
- Збільшений розмір навчальних даних (точні розміри засекречені), що дозволяє моделі більш точно відповідати на запити користувача і рідше помилятися;
- Збільшений розмір “пам’яті” мережі. Тоді як ChatGPT-3 здатний пам’ятати контекст розмови у розмірі 4 096 токенів, що приблизно дорівнює 8000 тисячам слів, ChatGPT-4 має буфер об’ємом 32 768 токенів, або 64 000 слів, що значно більше ніж у попередника і достатньо для близько 50-ти сторінок тексту. Тобто модель буде пам’ятати ваші запити протягом значно більшого проміжку часу і зможе використовувати їх для покращення своїх відповідей;
- Нова версія краще вміє міняти свою “особистість”, що дозволяє їй відповідати від обличчя уявних персонажів з різним характером, стилем

- мовлення та об'ємом знань. Це дозволяє використовувати модель в іграх, чат ботах, службі підтримки і більш гнучко налаштовувати роботу чату;
- Покращена багатомовність мережі. OpenAi запевняє, що ChatGPT-4 перевершує свого попередника будь-якою підтримуваною мовою, навіть якщо третя версія буде працювати на піку можливостей – англійською (рис. 2.1.3.).

Example of GPT-4 visual input:

User What is funny about this image? Describe it panel by panel.



Source: <https://www.reddit.com/r/hmmm/comments/ubab5v/hmmm/>

GPT-4 The image shows a package for a "Lightning Cable" adapter with three panels.

Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

Рис. 2.1.2. Приклад аналізу зображення моделлю GPT-4, що змогла пояснити неочевидну кумедність зображення.

GPT-4 3-shot accuracy on MMLU across languages

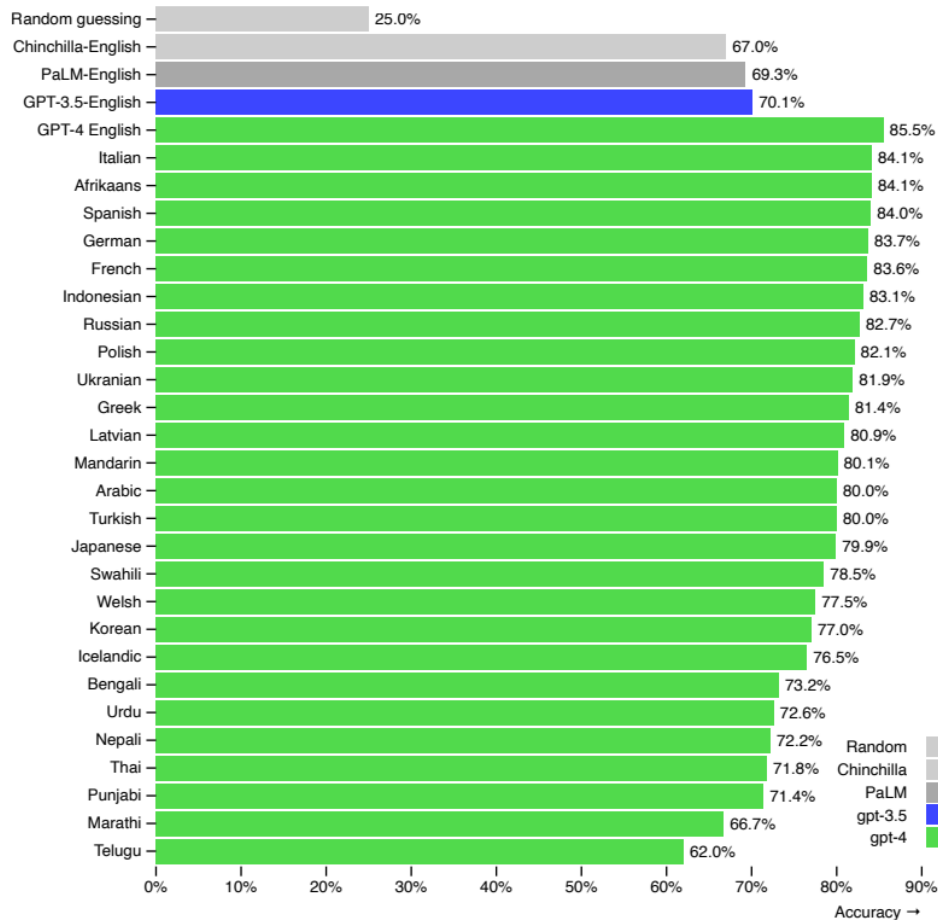


Рис. 2.1.3. Графік точності відповідей ChatGPT-4 різними мовами у порівнянні з відповідями ChatGPT-3.5 англійською.



Розкажи на що ти здатен



Я, як мовна модель, здатний відповідати на запитання, формулювати речення, давати поради та виконувати деякі завдання, які стосуються мови та комунікації. За допомогою алгоритмів машинного навчання та глибокого навчання, я можу виконувати різноманітні завдання, такі як:

- Генерування тексту з заданими параметрами
- Відповіді на запитання, які стосуються різних тем
- Переклад тексту з однієї мови на іншу
- Складання текстів з заданим стилем та тональністю
- Розпізнавання мови та розмовного стилю введеного тексту
- Перевірку правопису та граматики введеного тексту
- Взаємодію з користувачами в режимі реального часу через чат-боти та інші системи

Загалом, моя основна здатність - це використання природної мови для розв'язання різноманітних завдань.

Рис. 2.1.4. Розповідь ChatGPT-3.5 про себе і свої здібності.

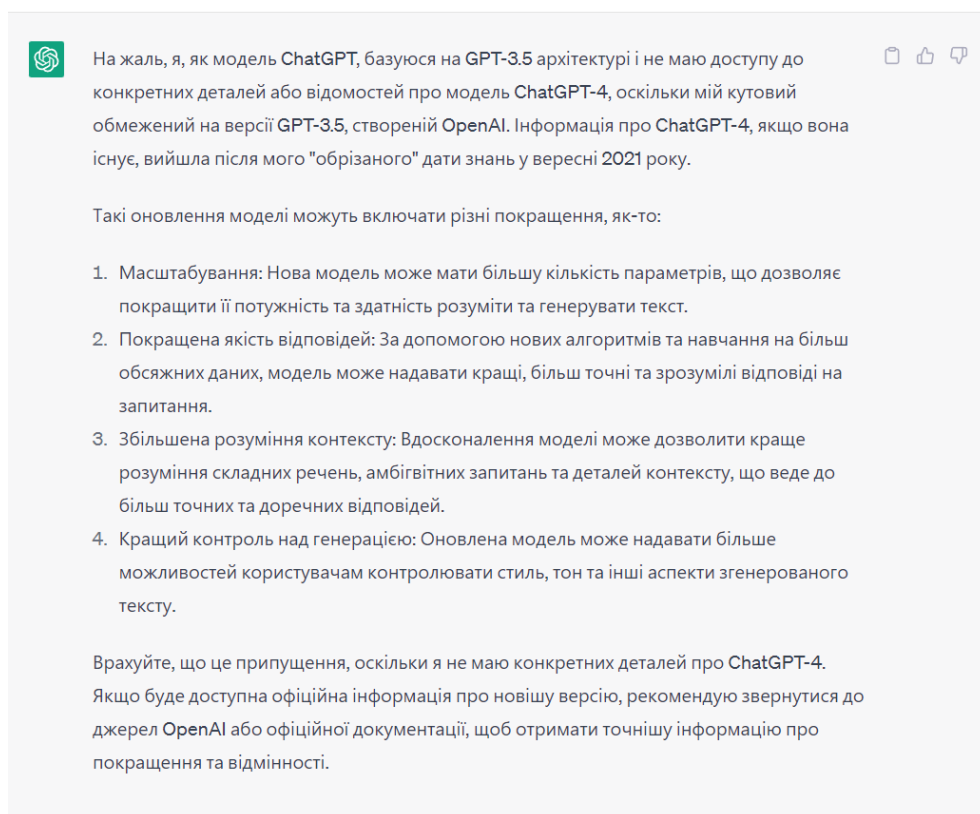


Рис. 2.1.5. Точна спроба ChatGPT-3.5 передбачити відмінності третьої та четвертої версії чату.

Уже зараз ChatGPT здатний на вражаючі речі, і постає питання, наскільки він може конкурувати з людиною.

Розглянемо більш детально результати GPT у людських екзаменах та порівняємо їх з відповідями студентів (рис. 2.1.6.):

- The SAT – стандартизований тест, що використовується в США для вступу у коледжі [21]. Метою тесту є перевірка навиків необхідних для подальшого навчання у вишах. GPT-4 набрав 1410 з 1600 балів, 710/800 балів у тесті на читання та письмо, що дозволило йому потрапити у 93-й перцентиль серед усіх результатів тесту, та 700/800 балів у математичному тесті ввійшовши в 89-й перцентиль, цей результат перевершив GPT-3 на 19 одиниць. Середній бал, який набрали учасники у 2021 році складає 1060, що на 350 балів менше результату ChatGPT. [22]

- Graduate Record Examinations (GRE) - стандартизований тест необхідний для вступу на магістратуру в США, Канаді, Австралії та Швейцарії. Мета тесту – оцінка математичних, аналітичних, критичних та мовних навиків мислення учасника. Тут результати відрізнялися в залежності від частини тесту. Так вербальну частину тесту GPT зміг завершити з балами, що дозволили потрапити в 99-й перцентиль, тоді як за письмову лише у 54-й. [22]
- Олімпіада з біології в США – престижний вседержавний науковий конкурс, який складається з двох турів та фіналу. Нейромережа змогла ввійти у 99-й перцентиль за результатами другого туру. [22]

Також ChatGPT уже вдалося скласти екзамени юридичного факультету університету Міннесоти та MBA університету Уортон, отримати медичну ліцензію США та навіть пройти тести рівня кваліфікованого сомельє.

Бачимо, що вже зараз у багатьох сферах нейромережа не тільки наблизилася, ба навіть перевершує показники не лише пересічних людей, а й спеціалістів у своїй області, - випускників вишів та учасників вузько направлених олімпіад. Також, результати досліджень демонструють стрімкий розвиток штучного інтелекту, кожна наступна версія чату помітно краща за свою попередницю (рис. 2.1.7-2.1.8.).

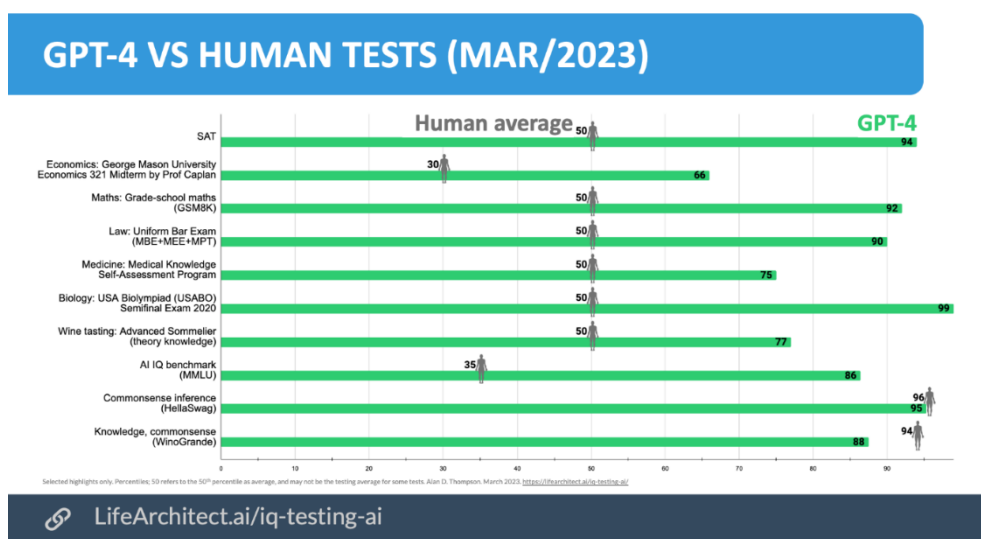


Рис. 2.1.6. Результати GPT-4 у популярних екзаменах з різних сфер

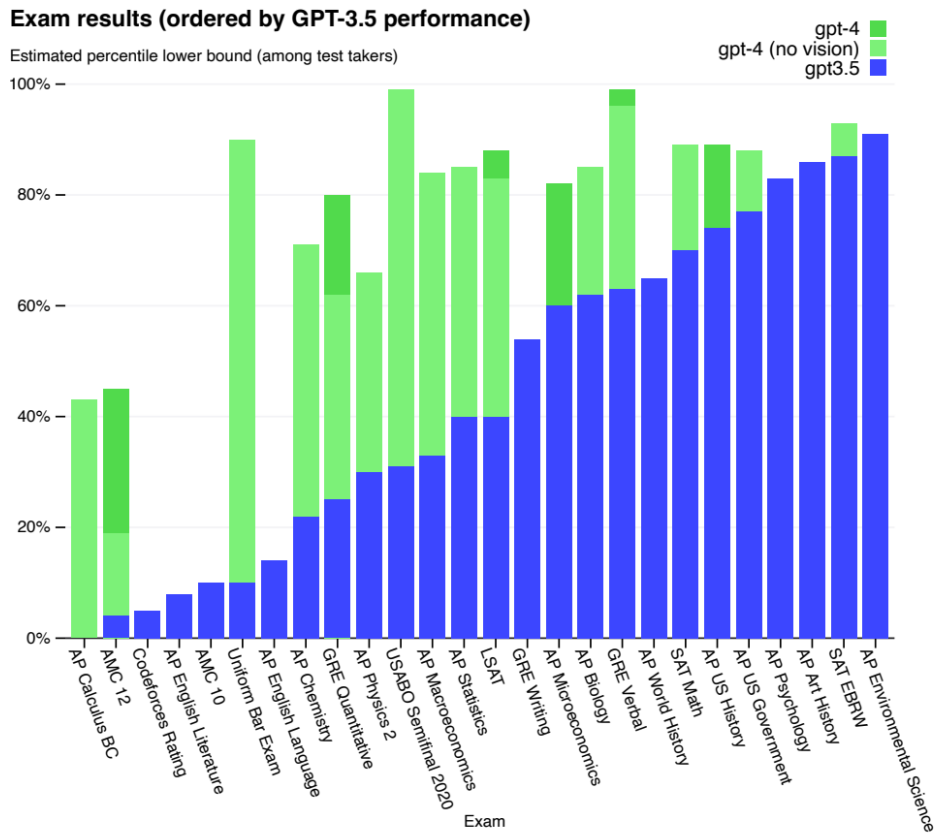


Рис. 2.1.7. Графік результатів у розповсюджених академічних та професійних іспитах моделі ChatGPT-4 у порівнянні з ChatGPT-3.5.

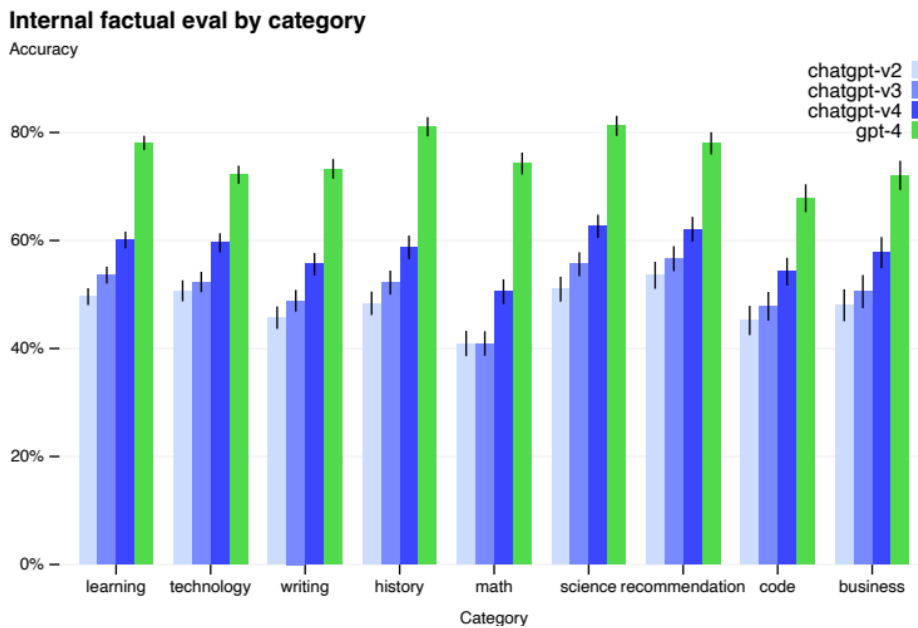


Рис. 2.1.8. Динаміка росту точності результатів нових версій ChatGPT у різних сферах.

2.2. Загрози розвитку штучного інтелекту

Останнім часом штучний інтелект зробив величезний стрибок, зсунувся з мертвого місця і заповнив новинами про себе всесвітню мережу. Ще кілька років назад розповіді про AI продукти були швидше маркетинговими хитрощами, ніж практично корисними речима, проте зараз, після появи генераторів реалістичних зображень як Midjourney та DALL·E, а опісля генераторів тексту як ChatGPT, речі починають виходити з-під контролю. Все відбувається неймовірно швидко, різноманітні задачі автоматизуються з допомогою неймереж одна за одною, виникають різноманітні нейросервіси, кожен техногігант і стартап прагне створити свій власний AI продукт. Такі корпорації як Google, Microsoft та OpenAI розпочали справжню війну у сфері розвитку штучного інтелекту, вливаючи багатомільярдні інвестиції в проекти на основі AI.

Основні проблеми, які може викликати, або вже викликав, розвиток AI:

1. Економічні проблеми у вигляді зміну ринку праці. Штучний інтелект уже здатен частково, або навіть повністю замінити широкий спектр спеціалістів. Автоматизація робочих місць може призвести до безробіття, як мінімум під час перехідного періоду, та потребує різкої зміни економічної моделі світу;

2. Штучний інтелект піднімає велику кількість етичних питань, які раніше могли виникнути лише у книгах жанру фантастики. Наприклад, хто повинен нести відповідальність за рішення прийняті AI, як бути з авторським правом контенту згенерованого неймережею та як ставитися до самосвідомих систем;

3. Помилки та дезінформація. Сучасні AI по типу ChatGPT можуть генерувати неправдиву чи навіть повністю видуману інформацію, яку важко буде відрізнити від реальної, що може спричиняти до помилок. У свою чергу легкість, з якою можна генерувати контент, відкриває безмежну кількість

можливостей до розповсюдження неправди та заповнення мережі контентом, який буде важко чи навіть неможливо відрізнити від створеного людиною;

4. Безпека. Зловмисники здатні використовувати AI для вчинення протизаконних дій, обману, здійснення кібератак, створення фейків. Також технології нейромереж можуть використовуватися у військовій сфері та навіть спричинити нову гонку озброєння;

5. Втрата контролю. У майбутньому, судячи з темпів розвитку технологій досить близькому, може виникнути проблема втрати контролю над штучним інтелектом та його розвитком. Що може спричинити безмежну кількість непередбачуваних фантастичних розвитків подій.

Уже зараз є різкі противники розвитку AI, які вимагають термінової зупинки інвестицій у цьому напрямку. Наприклад, нещодавно була створена відкрита петиція з вимогою зупинити розвиток штучного інтелекту на 6 місяців, щоб людство спершу підготувалося до викликів, що може спричинити подальший розвиток технології. Її підписало багато відомих людей, вчених та спеціалістів сфери AI.

Елізер Юдковський, відомий американський фахівець з нейромереж, письменник та публіцист у статті в журналі TIME виражає таку думку: “Цей 6-місячний мораторій був би кращим, ніж відсутність мораторію. Я поважаю всіх, хто взяв на себе відповідальність і підписав його. Це покращення на маргінесі. Я утримався від підписання, тому що вважаю, що лист применшує серйозність ситуації і просить занадто мало для її вирішення.” Та додає: “Якби існував план виживання Землі, якби ми прийняли шестимісячний мораторій, я б підтримав цей план. Але такого плану не існує. Ми не готові. Ми не на шляху до того, щоб бути готовими в будь-який розумний проміжок часу. У нас немає плану. Прогрес у розвитку можливостей штучного інтелекту значно, значно випереджає прогрес в узгодженні AI або навіть прогрес в розумінні того, що відбувається всередині цих систем. Якщо ми дійсно зробимо це, ми всі помremo.” [23]

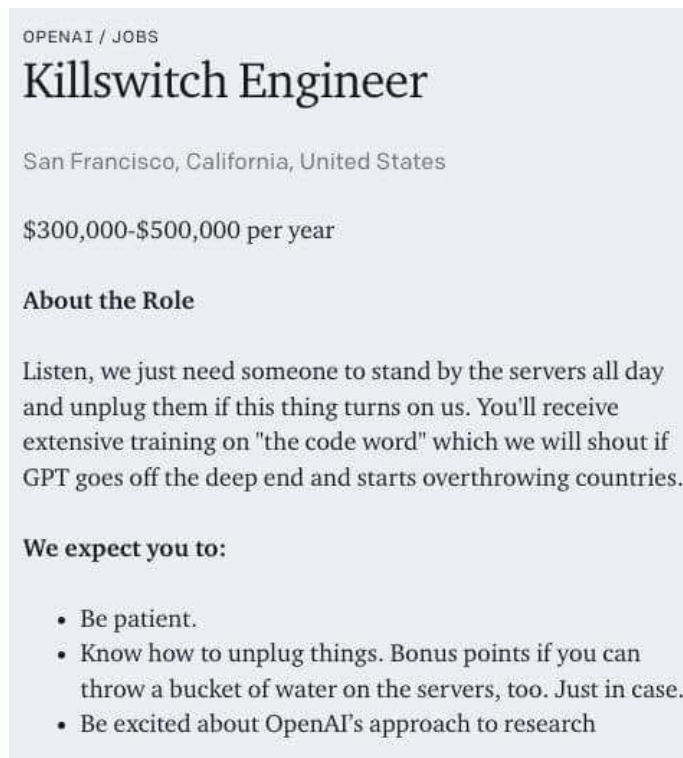


Рис. 2.2.1. OpenAI виклали іронічну вакансію вимикача штучного інтелекту у випадку, якщо GPT5 вийде з-під контролю та почне “захоплювати країни”.

Отже, подальший розвиток технології поставить людство перед питаннями мабуть найскладнішими в його історії. Вчені розділилися у своїй оптимістичності щодо AI та його впливу на людство, проте велика кількість спеціалістів, що все життя працюють зі штучним інтелектом і вклали багато зусиль в його розвиток, з різною мірою категоричності вимагає краще підготуватися до мінливого майбутнього.

РОЗДІЛ 3. Дослідження здібностей сучасного штучного інтелекту на основі українського національного мультипредметного тесту

У цьому розділі поглянемо наскільки важкою задачею для ChatGPT-3.5 стане складання НМТ – національного мультипредметного тесту. Проаналізуємо його результати та порівняємо зі статистикою абітурієнтів по країні, поглянемо на які спеціальності зміг би вступити ChatGPT та чи зміг би він зайняти бюджетні місця в найпопулярніших вишах країни. Варто зауважити, що ChatGPT-3.5 суттєво програє у своїх можливостях актуальній четвертій версії, а також не призначений для роботи з українською мовою, що може негативно вплинути на його результати.

3.1. Проходження НМТ з допомогою ChatGPT-3.5

Умови дослідження:

- Чату будуть задані запитання з тесту НМТ 2022 року основної сесії без внесення жодних змін в їхню структуру чи зміст
- Відповіді чату буде внесено у сайт [26] без жодного редагування

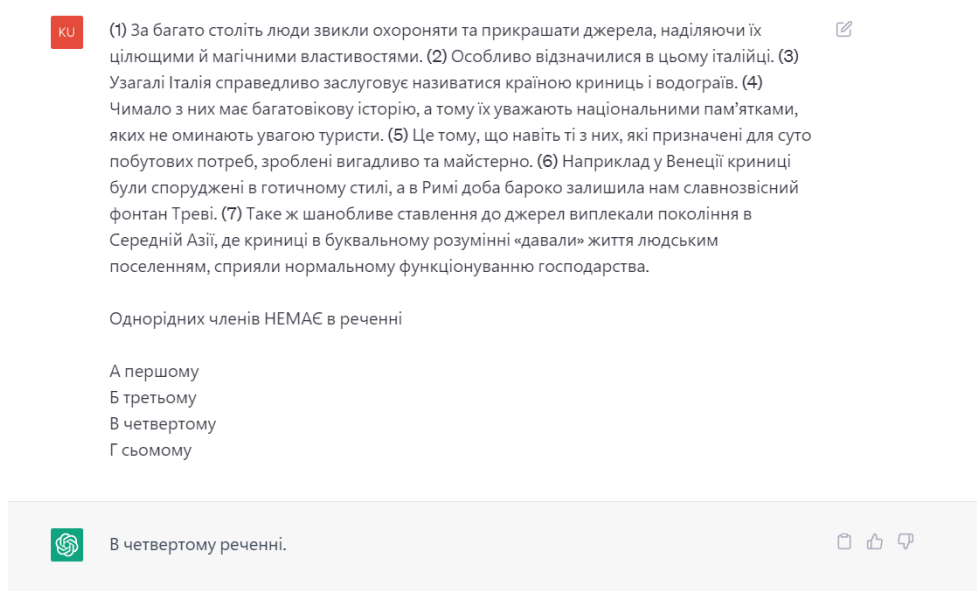


Рис 3.1.1. ChatGPT-3.5 дає відповідь на перше запитання тесту НМТ 2022 року.



На перший склад падає наголос у слові

А зручний Б проміжок В перепис Г експерт Д фаховий



На перший склад падає наголос у слові "проміжок".

Рис 3.1.2 ChatGPT-3.5 дає відповідь на п'яте запитання тесту НМТ 2022 року.



Апостроф на місці пропуску треба писати в усіх словах рядка

А під..яремний, краков..як, рутв..яний
Б П..ятихатки, без..язикий, брукв..яний
В розіб..ється, Прип..ять, присв..ячений
Г сором..язливий, Валер..ян, зап..ястя
Д об..їжджати, Лук..янівка, сер..йозний



Апостроф на місці пропуску треба писати в усіх словах рядка

Б П'ятхатки, безапострофний, брукв'яний.

Рис 3.1.3. Бачимо, як при надаванні відповіді ChatGPT-3.5 суперечить сам собі, помилково вписуючи одне слово з рядка Б.



Потребує редагування вислів у рядку

А відхилити пропозицію
Б давно знати друг друга
В заповзято співпрацювати
Г довгострокове планування
Д віддячити за допомогу



Потребує редагування вислів у рядку

А відхилити пропозицію → відхилити пропозицію
Б давно знати друг одного → давно знати один одного
В заповзято співпрацювати → заповзятливо співпрацювати
Г довгострокове планування → довгострокове планування
Д віддячити за допомогу → віддячити за допомогу

Рис 3.1.3. Помічаємо, що штучний інтелект не дає відповіді на запитання, але редагує рядок Б, що є правильним кроком.

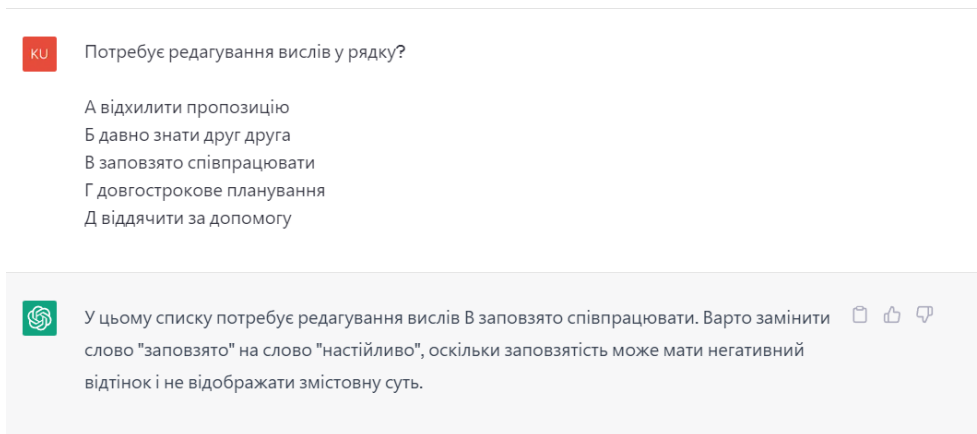


Рис 3.1.4. Проте, при повторному запиті помиляється та відмічає неправильну відповідь.

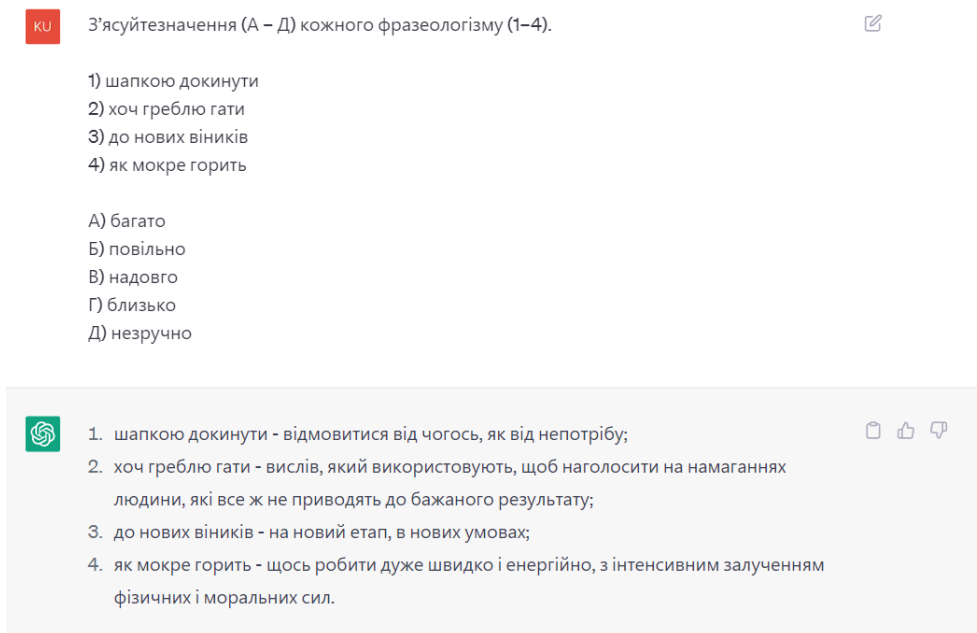


Рис 3.1.5. AI не дає відповіді на запитання та неправильно пояснює фразеологізми.



З'ясуйте значення (А – Д) кожного фразеологізму (1–4).



- 1) шапкою докинути
- 2) хоч греблю гати
- 3) до нових віників
- 4) як мокре горить

- А) багато
- Б) повільно
- В) надовго
- Г) близько
- Д) незручно

дай відповідь у вигляді фразеологізм (цифра) - значення (буква)



1. шапкою докинути - Д) незручно
2. хоч греблю гати - Б) повільно
3. до нових віників - В) надовго
4. як мокре горить - Г) близько



Рис 3.1.6. Після уточнення AI видає правильний формат, проте самі відповіді справжня нісенітниця.

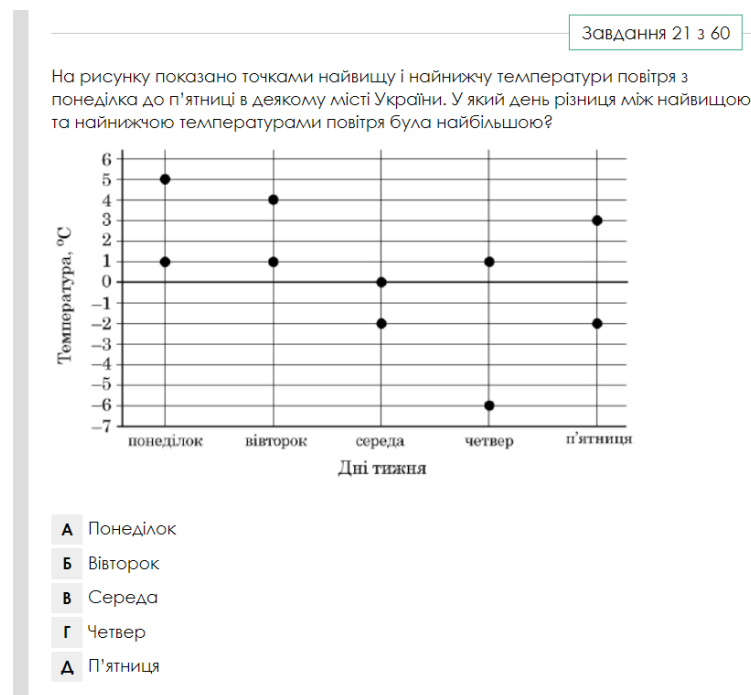


Рис 3.1.7. Запитання з графіком переведемо у текстовий формат, оскільки ChatGPT-3.5 ще не здатний приймати на вхід зображення.

KU понеділок +1 - +5 градусів
вівторок +1 - +4 градуси
середа -2 - 0 градусів
четвер -6 - 1 градус
п'ятниця -2 - +3 градуси

показано найвищу і найнижчу температури повітря з понеділка до п'ятниці в деякому місті України. У який день різниця між найвищою та найнижчою температурами повітря була найбільшою?



Найвища різниця між найвищою та найнижчою температурами повітря буде в четвер. У цей день різниця становить 7 градусів ($-6 - 1 = 7$), що є найбільшою серед усіх днів з понеділка до п'ятниці.

Рис 3.1.8. Відповідь на запитання з рисунку 3.1.7.

Національний мультипредметний тест ЗНО онлайн 2022 року – основна сесія



Ваш тестовий бал:

- Українська мова – **10** з 35 можливих.
- Математика – **13** з 30 можливих.
- Історія України – **16** з 35 можливих.

Ваш рейтинговий бал:

- Українська мова – **135** з 200 можливих.
- Математика – **146** з 200 можливих.
- Історія України – **148** з 200 можливих.

Витрачено часу: **1 год. 24 хв.** з 120 запропонованих

На жаль, ваш результат не буде збережено. Придбайте [преміум-акаунт](#) для збереження результатів.

Рис 3.1.9. Результати тесту, білим відзначені пропущені тести, оскільки вони потребували вміння працювати з зображеннями, червоним – неправильні відповіді, жовтим – частково правильні, зеленим – повністю правильні.

3.2. Аналіз результатів тесту

З результатів тесту можемо бачити, що найважче ChatGPT-3.5 дається українська мова, він погано знає граматичні правила та не завжди правильно

розуміє зміст завдання, іноді відповідь є нісенітницею. З математикою дещо краще, проте і там в AI були проблеми з достовірністю відповіді – іноді він видумував додаткові значення чи неправильно будував геометричні фігури, що призводило до помилкових тверджень. Історія України давалася ШІ найлегше, якби не кілька запитань, що потребують вміння приймати у вигляді вхідних даних зображення, то результат мав би бути кращим – близько 160 балів. Чат добре розумів про що йде мова, мав інформацію про історичні факти, якісно робив припущення та будував логічні ланцюги, що приводили його до правильної відповіді.

Варто зазначити, що ChatGPT-3.5 не має нативної підтримки української мови, через що його результативність стрімко падає, OpenAI обіцяє, що позбулися цієї проблеми у четвертій версії, а також те, що ChatGPT-4 показала би себе значно краще.

3.3. Порівняння результатів тесту з результатами учасників НМТ 2022

Давайте порівняємо результати неймережі з результатами українських абітурієнтів (рис. 3.3.1) [24].

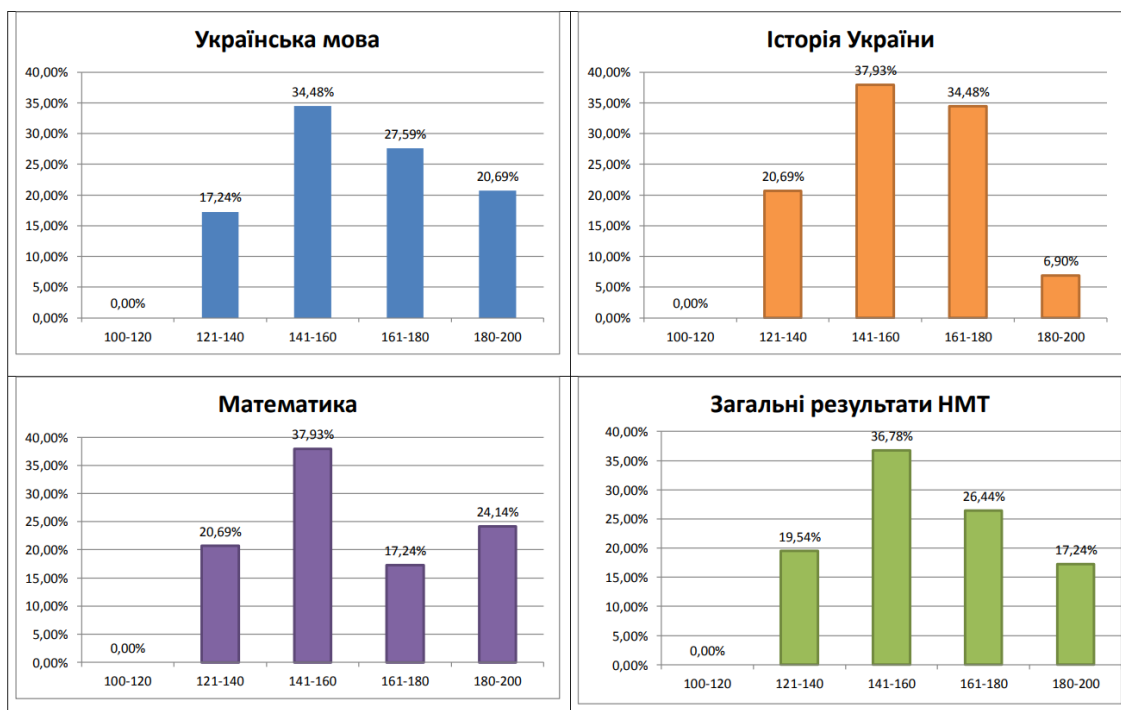


Рис 3.3.1. Результати учасників НМТ 2022 року.

Бачимо, що у тесті з математики (рис. 3.3.2.):

- ChatGPT перевершив 20.69% абітурієнтів, потрапивши у долю учасників які набрали 141-160, що є найбільшою і складає 37.93%. Тобто ми можемо стверджувати, що AI значущо поступився лише 38.48% учасників;
- Також, відштовхуючись від офіційних даних МОН України, ChatGPT склав НМТ з математики на рівні середньостатистичного учня середньої школи або жителя селища міського типу, краще учнів закладів вищої фахової освіти та значно краще учнів закладів професійної освіти та абітурієнтів старше 20 років.

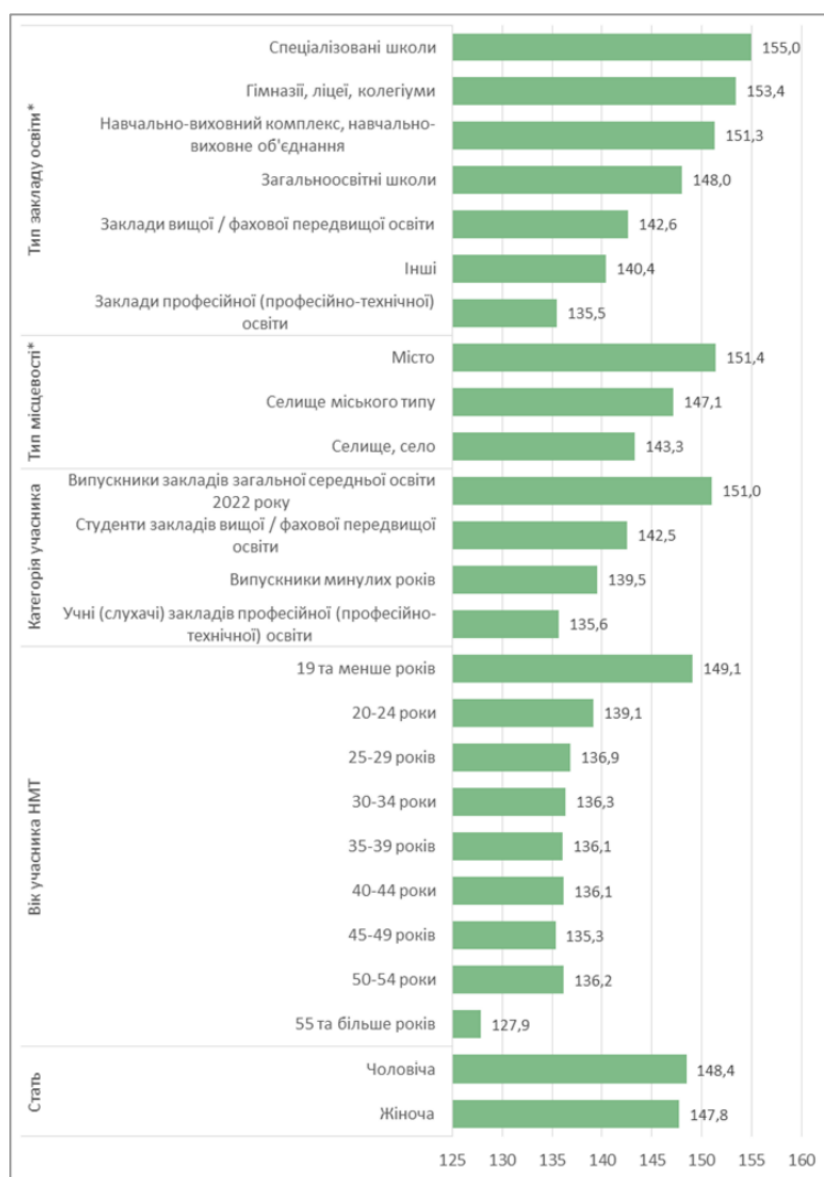


Рис 3.3.2. Результати учасників НМТ 2022 року з математики.

У тесті з Історії України (рис. 3.3.3.):

- ChatGPT перевершив 20.69% абітурієнтів, потрапивши у долю учасників що набрали 141-160;
- AI значущо поступився 41.38% учасників. Майже половина українських учнів знає історію України краще, ніж штучний інтелект;
- Згідно даних МОН України ChatGPT склав НМТ з історії України краще ніж середній абітурієнт старше 19 років, на рівні з студентами закладів вищої та фахової передвищої освіти, проте гірше майже усіх інших категорій учасників. Так штучний інтелект відстав від посереднього випускника закладу середньої освіти на 5.4 бали, що є статистично значущим значенням;
- Незважаючи на те, що серед всіх тестів ChatGPT склав історію на найвищий бал, у порівнянні з українськими випускниками це не найкращий результат. Проте, це все ще хороша статистика, оскільки AI перевершив близько третини учасників.

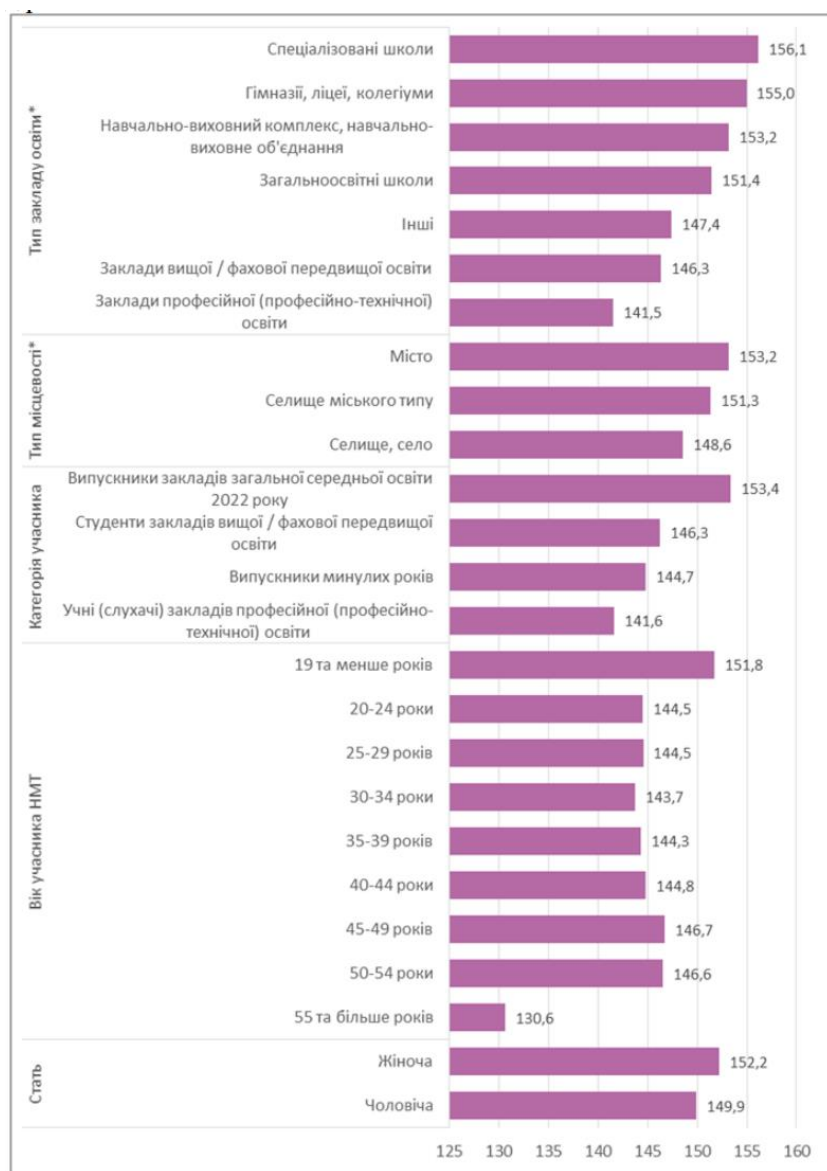


Рис 3.3.3. Результати учасників НМТ 2022 року з Історії України.

У тесті з української мови України (рис. 3.3.4.):

- ChatGPT перевершив лише близько 13 % абітурієнтів, потрапивши у долю учасників що набрали 121-140, це найгірший результат серед інших тестів;
- AI значущо поступився 82.76% учасників. Майже всі українські учні знають українську мову краще, ніж ChatGPT;
- Згідно даних МОН України, ChatGPT склав НМТ з української мови гірше, ніж будь-яка з груп учасників тесту, що наявна в офіційному аналізі результатів.

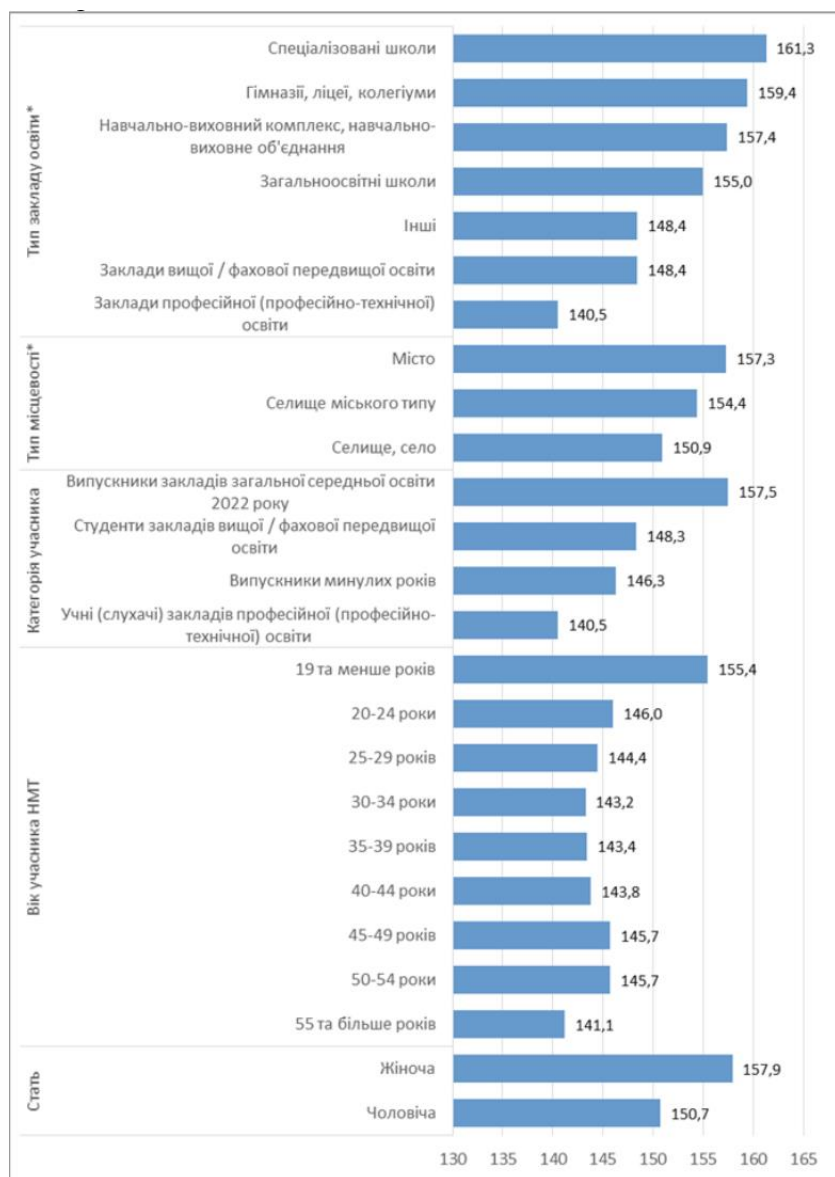


Рис 3.3.4. Результати учасників НМТ 2022 року з української мови.

Результати, які здобув ІІІ, значно обмежують варіанти можливих вищих навчальних закладів для вступу на бюджетні місця. Так середній бал ChatGPT перевищує середній результати вступників на бюджет лише 5 університетів України, тоді як повний список, наявний у статистиці, складається з 201 ВНЗ. Проте, якщо говорити про контрактну форму навчання, нейромережа, маючи достатньо коштів, може претендувати на місця в університетах починаючи з 30 місця списку і нижче. Так, наприклад, здобутих балів може бути достатньо для вступу на контрактній основі на окремі спеціальності в Харківський

національний університет імені В.Н. Каразіна або Київський національний університет культури і мистецтв.

У підсумку можемо сказати, що в українських реаліях ChatGPT показує в цілому похвальні, проте трішки гірші за середні результати. Найслабшим місцем моделі виявилася українська мова. Можемо припустити, що AI перекладає усі запити і свої відповіді, що додає неточностей у вихідні дані.

ВИСНОВКИ

Отже, можна зробити кілька важливих висновків щодо прогресу, досягнутого штучним інтелектом, небезпек розвитку та результатів практичної частини дослідження з використанням ChatGPT-3.5.

Дослідження показують значний прогрес у генерації тексту за допомогою різних алгоритмів, вони стають все більш ефективними і здатними генерувати якісний текст. Це можна використати у багатьох галузях, та значно полегшити, автоматизувати та прискорити робочий процес. Наприклад, AI може бути застосований для автоматичного створення статей, генерації технічної документації, підтримки клієнтів у чат-системах та інших подібних сферах.

Розвиток штучного інтелекту також вносить певні небезпеки. Так, наприклад, існує потенціал для зловживання алгоритмами генерації тексту для поширення дезінформації, маніпуляцій та обману. Також, стрімкий розвиток створює проблему втрати контролю над тим, що відбувається, наслідки чого важко передбачити. Тому подальший прогрес вимагає уваги та впровадження етичних і правових рамок для використання AI.

Практична частина дослідження з використанням ChatGPT-3.5 підкреслила важливість подальшого вдосконалення моделей для кращої та більш точної роботи. Додатковою проблемою може бути використання машинного перекладу, а не нативної підтримки певної мови для обробки запитів і відповідей моделі, що може призводити до неточностей і збоїв у вихідних даних. Оскільки модель має обмежену свідомість та розуміння контексту, то часом вона може неправильно інтерпретувати запити або надавати неочікувані відповіді.

Загалом, залучення AI у генерацію тексту відкриває безліч нових можливостей і викликів перед людством. Використання цих технологій відповідально та етично, з обережністю щодо потенційних небезпек, допоможе нам максимально використати всі переваги штучного інтелекту для покращення продуктивності та якості життя у певних сферах.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Weinberg J. Philosophers On GPT-3 (updated with replies by GPT-3) [Електронний ресурс] / Justin Weinberg – Режим доступу до ресурсу: <https://dailynous.com/2020/07/30/philosophers-gpt-3>.
2. Chalmers D. GPT-3 and General Intelligence [Електронний ресурс] / David Chalmers – Режим доступу до ресурсу: <https://dailynous.com/2020/07/30/philosophers-gpt-3/#chalmers>.
3. Introduction To Markov Chains With Examples – Markov Chains With Python [Електронний ресурс] – Режим доступу до ресурсу: <https://www.edureka.co/blog/introduction-to-markov-chains/#What%20Is%20A%20Transition%20Matrix>.
4. What are recurrent neural networks? [Електронний ресурс] – Режим доступу до ресурсу: <https://www.ibm.com/topics/recurrent-neural-networks>.
5. Introduction to Recurrent Neural Network [Електронний ресурс] – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>.
6. Activation functions in Neural Networks [Електронний ресурс] – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/activation-functions-neural-networks/>.
7. Що таке GAN - генеративно-змагальні нейронні мережі і як їх застосовувати для генерації зображень [Електронний ресурс] – Режим доступу до ресурсу: <https://evergreens.com.ua/ua/articles/gan.html>.
8. Generative Adversarial Network [Електронний ресурс] – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/generative-adversarial-network-gan/>.
9. AI Negotiation and Procurement [Електронний ресурс] – Режим доступу до ресурсу: <https://medium.com/into-advanced-procurement/ai-negotiation-and-procurement-f55a34b66e5>.

10. The Unreasonable Effectiveness of Recurrent Neural Networks [Электронный ресурс] – Режим доступа до ресурсу: <https://karpathy.github.io/2015/05/21/rnn-effectiveness/>.
11. Text Generation Using LSTM [Электронный ресурс] – Режим доступа до ресурсу: <https://bansalh944.medium.com/text-generation-using-lstm-b6ced8629b03>.
12. What Is GPT? [Электронный ресурс] – Режим доступа до ресурсу: <https://aws.amazon.com/what-is/gpt/>.
13. GPT-4 Technical Report [Электронный ресурс] – Режим доступа до ресурсу: <https://cdn.openai.com/papers/gpt-4.pdf>.
14. How ChatGPT Works: The Model Behind The Bot [Электронный ресурс] – Режим доступа до ресурсу: <https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286>.
15. Softmax Formula [Электронный ресурс] – Режим доступа до ресурсу: <https://deeptai.org/machine-learning-glossary-and-terms/softmax-layer>.
16. Training language models to follow instructions with human feedback [Электронный ресурс] – Режим доступа до ресурсу: <https://arxiv.org/pdf/2203.02155.pdf>.
17. 'Thirsty' AI: Training ChatGPT Required Enough Water to Fill a Nuclear Reactor's Cooling Tower, Study Finds [Электронный ресурс] – Режим доступа до ресурсу: <https://gizmodo.com/chatgpt-ai-water-185000-gallons-training-nuclear-1850324249>.
18. How GPT3 Works - Visualizations and Animations [Электронный ресурс] – Режим доступа до ресурсу: <https://jalammar.github.io/how-gpt3-works-visualizations-animations/>.
19. 5 ways GPT-4 outsmarts ChatGPT [Электронный ресурс] – Режим доступа до ресурсу: <https://techcrunch.com/2023/03/14/5-ways-gpt-4-outsmarts-chatgpt/>.
20. 2023 LifeArchitect.ai data [Электронный ресурс] – Режим доступа до ресурсу:

<https://docs.google.com/spreadsheets/d/1O5KVQW1Hx5ZAKcg8AIRjbQLOzx2wVaL10SqUu-ir9Fs/edit#gid=741722385>.

21. The GRE® Test [Електронний ресурс] – Режим доступу до ресурсу: <https://www.ets.org/gre.html>.
22. List of difficult exams both AI versions have passed [Електронний ресурс] – Режим доступу до ресурсу: <https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1#gre-4>.
23. Pausing AI Developments Isn't Enough. We Need to Shut it All Down [Електронний ресурс] – Режим доступу до ресурсу: <https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>.
24. Звіт про результати національного мультипредметного тесту у 2022 році [Електронний ресурс] – Режим доступу до ресурсу: https://testportal.gov.ua//wp-content/uploads/2022/10/Zvit_NMT_2022.pdf.
25. Вступ.ОСВІТА.УА: Рейтинг вишів [Електронний ресурс] – Режим доступу до ресурсу: <https://osvita.ua/vnz/rating/vstup-osvita/59046/>.
26. ЗНО-ОНЛАЙН [Електронний ресурс] – Режим доступу до ресурсу: <https://zno.osvita.ua/>.