УДК 004.85+004.8.032.26

Еволюція нейронних моделей генерування тексту: систематичний огляд досліджень 2022–2024 років

Артем Валерійович Слободянюк
ORCID: 0009-0007-9425-1255
minekosdid@kdpu.edu.ua

Криворізький державний педагогічний університет

Сергій Олексійович Семеріков
професор, старший дослідник, д-р пед. наук
ORCID: 0000-0003-0789-0272
semerikov@gmail.com

Криворізький державний педагогічний університет
Інститут цифровізації освіти НАПН України
Державний університет "Житомирська політехніка"
Криворізький національний університет
Академія когнітивних та природничих наук

Ключові слова:
нейроне генерування тексту;
глибоке навчання;
систематичний огляд;
обробка природної мови;
метрика;
набори даних;
низькоресурсні мови;
трансформери;
механізми уваги.

Останні роки характеризуються значним прогресом у сфері нейронного генерування тексту завдяки появі великих мовних моделей та зростанню інтересу до цієї галузі. Цей систематичний огляд ідентифікує та узагальнює сучасні тенденції, підходи та методи нейронного генерування тексту за період 2022–2024 рр., доповнюючи попередній огляд за 2015–2021 рр. Відповідно до методології PRISMA, для аналізу було початково відібрано 89 статей з бази даних Scopus, із яких після перевірки критеріїв включення та виключення залишилося 43 статті. Виявлено зміщення акценту в бік інноваційних архітектур моделей, як-от Transformer-based (GPT-2, GPT-3, BERT), механізмів уваги та контрольованого генерування тексту. Метрики BLEU, ROUGE та оцінювання людиною залишаються найпопулярнішими. Але з'явилися і нові метрики, поміж яких виділимо BERTScore. Набори даних охоплюють різноманітні домени і типи даних; спостерігається зростання інтересу до неанотованих даних. Сфери застосування розширилися до областей генерування тексту на основі таблиць та графів знань, синтезу анотацій та машинного перекладу. У галузевому плані виділяється генерування медичних текстів. Хоча англійська мова продовжує домінувати, але спостерігається зростання досліджень для низькоресурсних мов, зокрема до німецької та китайської. Огляд також висвітлює актуальні виклики в цій галузі, зокрема адаптацію моделей для низькоресурсних мов, генерування тексту за умов обмеженості навчальних даних та етичні аспекти використання потужних мовних моделей. Автори підкреслюють важливість розробки більш ефективних та інтерпретовних архітектур, вдосконалення методів контрольованого генерування тексту та створення нових оцінювальних метрик. Результати дослідження підкреслюють швидку еволюцію методів нейронного генерування тексту, розширення сфер його застосування. В огляді також окреслено перспективні напрями для майбутніх досліджень з урахуванням актуальних викликів та етичних принципів.

DOI: https://doi.org/10.31558/2786-9482.2024.2.4

Вступ

Опрацювання природної мови (Natural Language Processing, NLP) – міждисциплінарна галузь інформатики та лінгвістики [1], класифікацію основних задач якої подано на рис. 1.

Рисунок 1. Таксономія популярних задач NLP для генерації тексту (на основі [1])

Генерування тексту – розділ NLP, що поєднує обчислювальну лінгвістику та штучний інтелект для синтезу текстів [2]. Найбільш відомою практичною реалізацією цього розділу NLP є ChatGPT – чат-бот на основі моделі GPT, який представлено OpenAI у 2022 р. [3].

Попередній огляд [2] охоплював 90 джерел із 2015 до 2021 р. Надання користувачам доступу до великих мовних моделей у 2022–2023 рр. [4] призвело до зростання інтересу до них (рис. 2), тому виникла потреба у доповненні попереднього огляду.

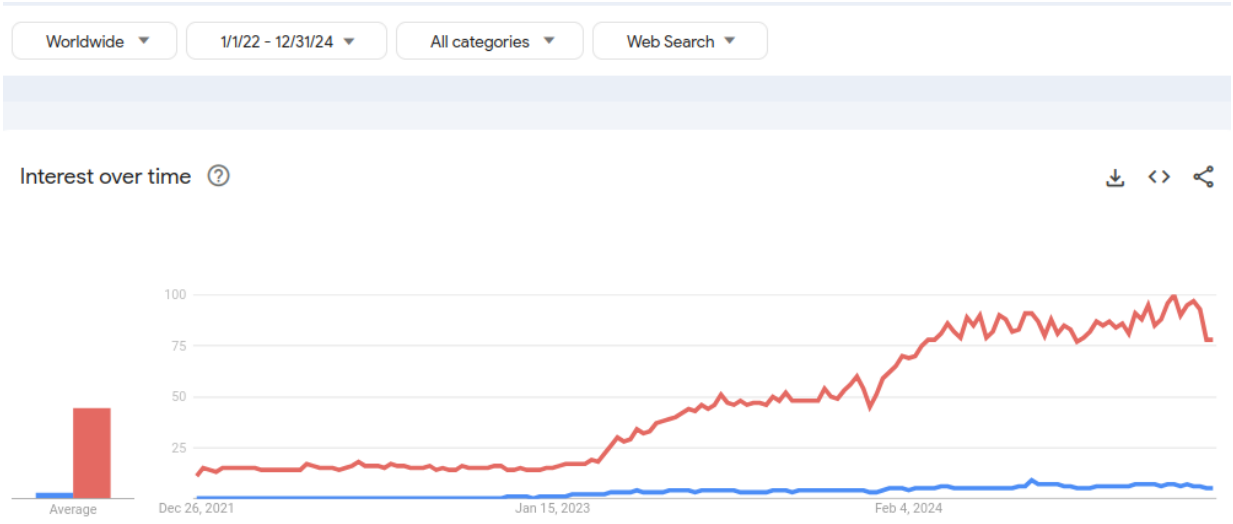

Рисунок 2. Динаміка запитів за пошуковим виразом “large language models” [4]

Основним результатом огляду [2] є класифікація (рис. 3) за п’ятьма ознаками:

1) *за архітектурою нейронної мережі*:

- традиційні:

  RNN – рекурентна нейронна мережа, що використовується для послідовних даних;

  LSTM – мережа з довгою короткочасною пам’яттю, що працює краще за RNN за більших об’ємів даних;

  GRU – вентильний рекурентний вузол (спрощена версія LSTM);

  CNN – згорткова нейронна мережа;

- інноваційні:

  Attention Based – мережі, що використовують механізм уваги для підвищення значущості вхідних даних;

  Transformer – мережі, що використовують механізм уваги без рекурентних або згорткових шарів;

  BERT – розроблена Google нейронна мережа, що поєднує механізми уваги без рекурентних або згорткових шарів із двонапрямленими кодувальниками;

2) *за метриками якості*:

- людино-орієнтовані – Domain-Expert, які залучають фахівців у предметній області для валідації результатів;
- машинно-орієнтовані (автоматичні):

  BLEU (bilingual evaluation understudy) – порівнює кількість і значення токенів (лексем) машинного і людського перекладів без врахування значень слів;

  ROUGE (Recall-Oriented Understudy for Gisting Evaluation) – порівнює анотації та переклади, які згенеровано машиною та людиною;

  Cosine Similartity – порівнянє косинус кута двох ненульових векторів;

  Content Selection – схожа з ROUGE метрика, яка використовує механізм уваги щодо аналізованої задачі;

  Diversity Score – метрика оцінювання різноманіття;

3) *за застосуванням нейронної мережі*:

  AMR (Abstract Meaning Representation) – видобування семантичних співвідношень із тексту;

  Language Generation – генерування тексту, подібного до людського;

  Speech-to-text – перетворення усної промови у текст;

  Script Generation – генерування сценаріїв на основі заданих слів;

  Machine Translation – машинний переклад тексту з однієї мови на іншу;

  Text Summarization – генерування анотації тексту;

  Image Captioning – генерування опису за зображенням;

  Shopping Guide – генерування рекламного опису за зображенням товару;

  Weather Forecast – генерування текстового прогнозу погоди;

4) *за мовою тексту*:

- з тих, що гарно забезпечені ресурсами – англійська, китайська;

- з тих, що недостатньо забезпечені ресурсами – бенгальська, корейська, балійська, іспанська, хінді, словацька, македонська тощо.

5) *за набором даних для навчання нейронної мережі*:

- за розміткою:
  Labeled – розмічені дані;
  Unlabeled – нерозмічені дані;
- за типом:
  Sentence – речення;
  Paragraph – абзац;
  Question / answer – дані типу питання та відповідь;
  Document – дані у вигляді документа.

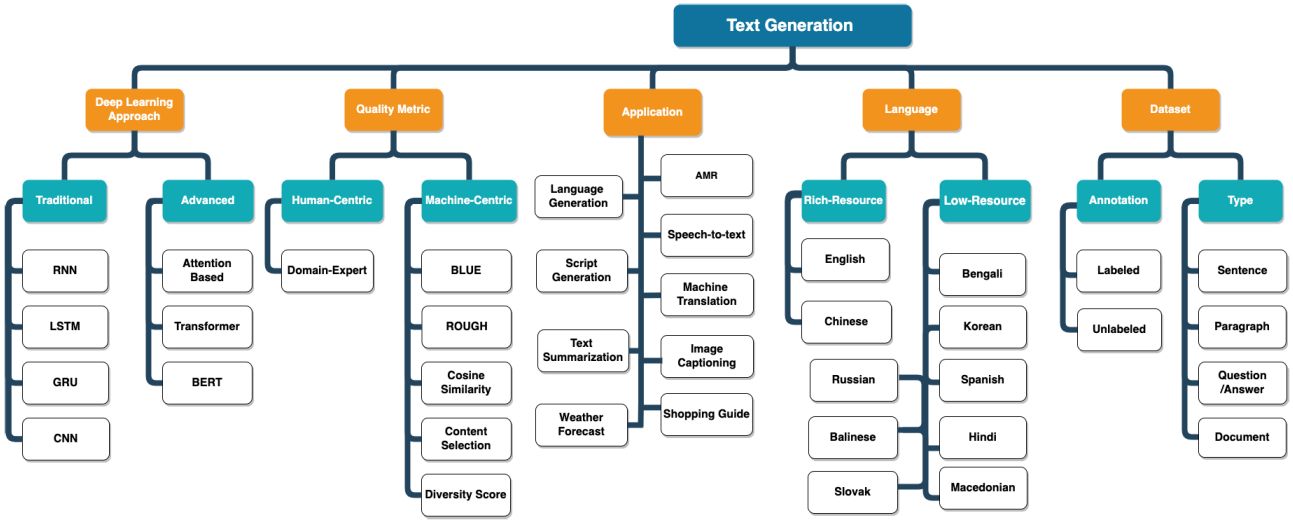

Рисунок 3. Класифікація процесів генерації тексту [2]

Класифікація з рис. 3 є результатом опрацювання таких п’яти задач [2]:

1. Дослідити традиційні та інноваційні методи та підходи глибокого навчання для генерування тексту.
2. Розглянути метрики продуктивності для оцінювання моделей генерування тексту.
3. Дослідити методи оцінювання для вимірювання якості згенерованого тексту.
4. Оглянути нові галузі застосування методів генерування текстового контенту.
5. Обговорити найважливіші проблеми та майбутні напрями дослідження у галузі генерування текстового контенту.

Метою статті є доповнення отриманих у [2] результатів шляхом вирішення таких 5 *дослідницьких питань*:

- які передові методи глибокого навчання використовуються для генерування тексту в літературі 2022–2024 рр.;
- за якими новими метриками оцінюють ефективність моделей генерування тексту в літературі 2022–2024 рр.;
- які набори даних для генерування тексту описано в літературі 2022–2024 рр.;

- які нові застосування методів генерування тексту описано в літературі 2022–2024 рр.;
- які природні мови використовуються для генерування тексту згідно до літератури 2022–2024 рр.

Методика дослідження

Систематичний аналіз літератури є основним **методом цього дослідження,** який дає змогу узагальнити інформацію з великої кількості наукових публікацій (вторинних джерел) за чітко визначеною методикою. Для проведення огляду було обрано методику PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), яка є загальновизнаним стандартом для систематичних оглядів та метааналізу у різних галузях науки [5]. Систематичний аналіз за методикою PRISMA передбачає чітке планування дослідження, визначення критеріїв відбору публікацій, проведення ретельного пошуку літератури у провідних наукових базах даних, відбір релевантних досліджень, видобування та синтез даних. Такий підхід забезпечує повноту, надійність і відтворюваність отриманих результатів, що повністю відповідає меті та завданням дослідження.

Джерела інформації та стратегія пошуку

У попередньому огляді [2] як джерела інформації використовувалися наукометричні бази Web of Science та Scopus та бібліотеки IEEE Xplore, SpringerLink, ScienceDirect та ACM Digital Library. Пошуковий запит за назвами статей, анотаціями та ключовими словами, які використано у [2], подано у табл. 1.

Таблиця 1. Формування пошукового запиту в [2]

Назва	Зміст
Група 1: Слова, які стосуються глибокого навчання	deep learning OR natural language processing OR NLP OR neural network OR RNN OR recurrent OR recursive OR LSTM OR GAN OR GPT-2 OR generative adversarial network
Група 2: Слова, які стосуються генерування тексту	text generation OR language generation OR language modelling OR natural language generation OR neural language generation
Пошуковий запит	(Група 1) AND (Група 2)

Наразі Scopus покриває приблизно 90% контенту IEEE Xplore та ACM Digital Library, а Web of Science – приблизно 50%. ScienceDirect та Scopus мають одного й того ж власника. Ураховуючи, що до Scopus входить значна частина вказаних бібліотек, замість 2 баз та 4 бібліотек дослідження проведемо лише за базою Scopus. За пошуковим запитом із попереднього огляду (табл. 1) видача становить 2 580 документів за 2015–2020 рр. проти 100 документів, вказаних у [2]. Якщо шукати виключно за назвами статей, кількість документів зменшується до 109 і спостерігається часткове співпадіння з переліком джерел із [2].

Неможливість відтворення попередніх результатів за запитом із табл. 1 спонукало до створення такого нового запиту:

```
(
    TITLE-ABS-KEY(neural network)
    OR
    TITLE-ABS-KEY(machine learning)
    OR
    TITLE-ABS-KEY(deep learning)
)
AND
TITLE("text generation")
```

Перша частина запиту спрощена до трьох ключових фраз, дві з яких (“neural network” та “deep learning”‘) співпадають із першою групою табл. 1, а третя (“machine learning”) узагальнює усі інші ключові слова першої групи, включно з неіснуючими на момент створення попереднього огляду. До другої частини запиту включена лише ключова фраза “text generation”, пошук якої виконується у заголовках документів, а не в заголовках, анотаціях та авторських ключових словах.

Критерії відбору документів

Для аналізу відбираються лише ті документи, які одночасно задовольняють такі 3 умови:

- опубліковані протягом 2022–2024 рр.;
- стосуються генерування тексту за допомогою штучних нейронних мереж;
- описують підходи, архітектури, метрики якості, мови, набори даних або застосування згенерованого тексту.

Документи виключаються, якщо вони задовольняють хоча б одну з таким умов:

- опубліковані до 2022 року або такі, що не містять даних за 2022–2024 рр.
- не стосуються генерування тексту або не використовують штучні нейронні мережі.
- не містять релевантної інформації щодо поставлених дослідницьких питань (нові методи, метрики, набори даних, застосування, природні мови).

Процес відбору документів

Запит до Scopus 04.03.2024 р. повернув 248 документів, розподіл яких за роками подано на рис. 4. Із них 2 виявились дублікатами, а 157 – датованими раніше 2022 р., тому вони були виключені.

На рис. 5 подана схема відбору даних для систематичного огляду. 89 документів отримано із сайтів видавців, із наукової соціальної мережі ResearchGate та серверів препринтів (насамперед arXiv). 41 документ (насамперед з ACM Digital Library та IEEE Xplore) отримати не вдалось. Отже, для оцінювання відібрано 48 документів, перегляд яких виявив 1 документ, що не містив дані за 2022–2024 рр., та 4 документи, що не містили релевантної інформації щодо поставлених дослідницьких питань. На загал, такі 43 документи відібрано для аналізу: [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48].

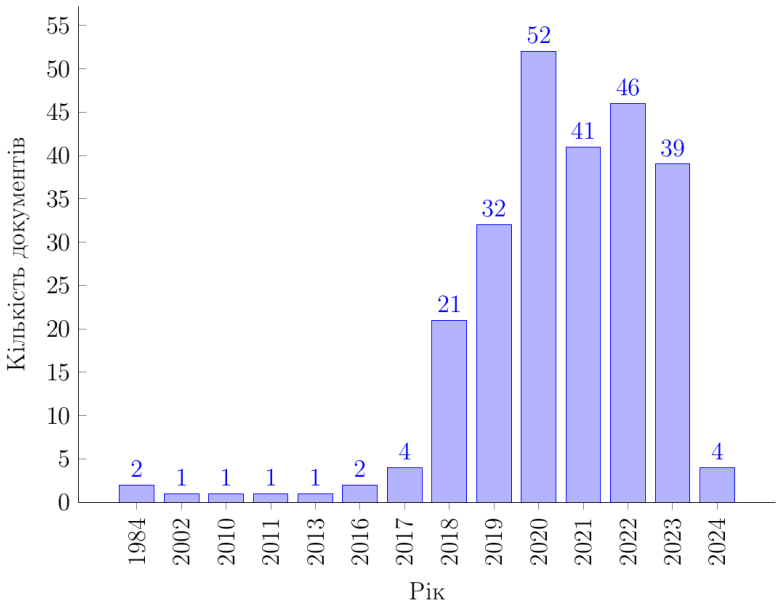

Рисунок 4. Розподіл результатів пошуку за роками

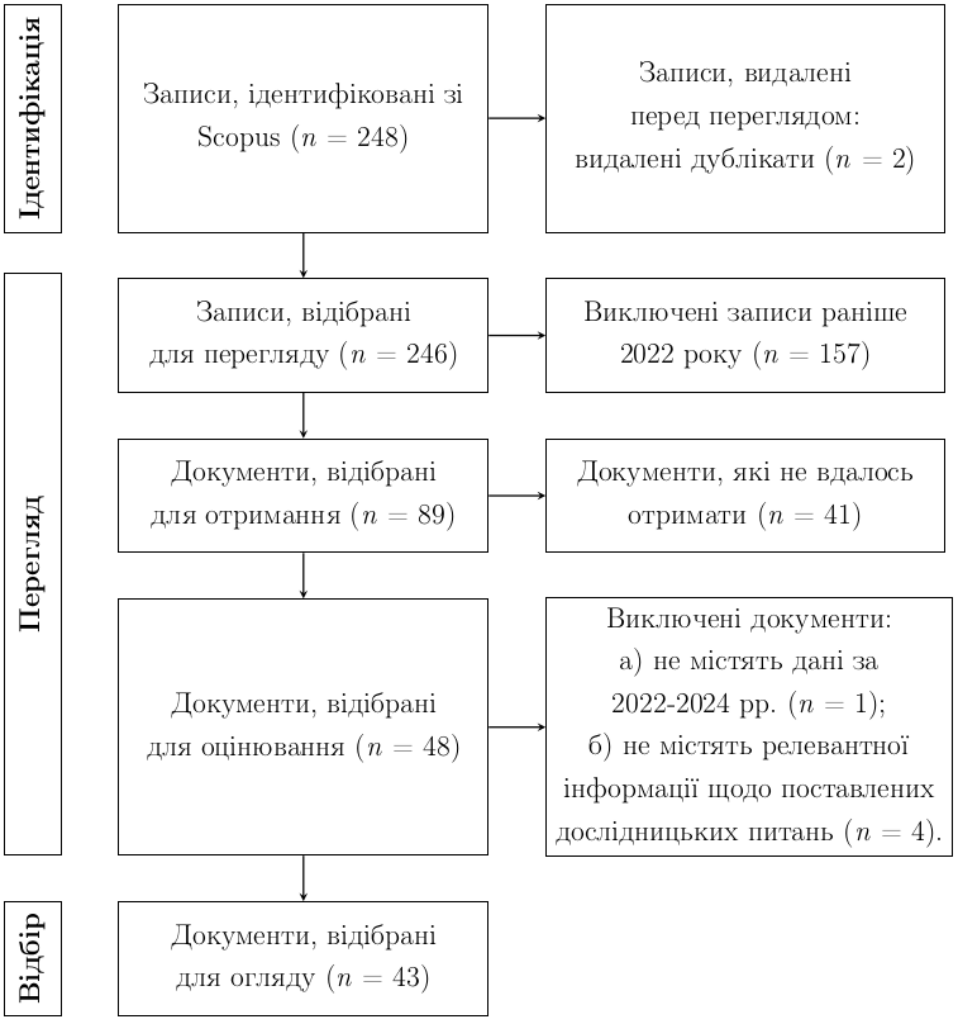

Рисунок 5. Схема відбору даних для систематичного огляду (згідно з методикою PRISMA [5])

Аналіз кожного документа виконувався відповідно до карти огляду за допомогою великої мовної моделі Claude 3 Sonnet [49]. На вхід моделі подавався PDF-файл документа з промптом, який відповідає карті огляди. Промт є таким:

```
Опиши статтю за такими характеристиками:
Тип документу: стаття у журналі (ARTICLE) або стаття у матеріалах
конференції (CONFERENCE)
Назва
Рік публікації
Країни, які представляють автори
Мета статті
Використані архітектури нейронних мереж
Використані метрики якості
Характеристики використаних наборів даних - ім’я
Характеристики використаних наборів даних - тип даних: речення, абзац,
документ, питання-відповідь, не вказаний
Характеристики використаних наборів даних - розмір
Характеристики використаних наборів даних - формат: CSV, JSON, XML,
файли, не вказаний
Характеристики використаних наборів даних - за типом розмітки: розмічені
дані, нерозмічені дані
Характеристики використаних наборів даних - якість даних: попередньо не
опрацьовані (сирі), попередньо опрацьовані
Характеристики використаних наборів даних - за доступності:
загальнодоступний, приватний, не вказаний
Характеристики використаних наборів даних - посилання
Розв’язана задача генерації тексту (для чого застосовувалась нейронна
мережа)
Мова генерації тексту
```

Приклад відповіді подано на рис. 6. Для зовнішньої перевірки застосовано велику мовну модель GPT 4 через розширення Scholar GPT [50]. Заповнені карти огляду для кожної статті наведено у [51]. Для прикладу в табл. 2 подано порівняння результатів автоматичного аналізу статті [47].

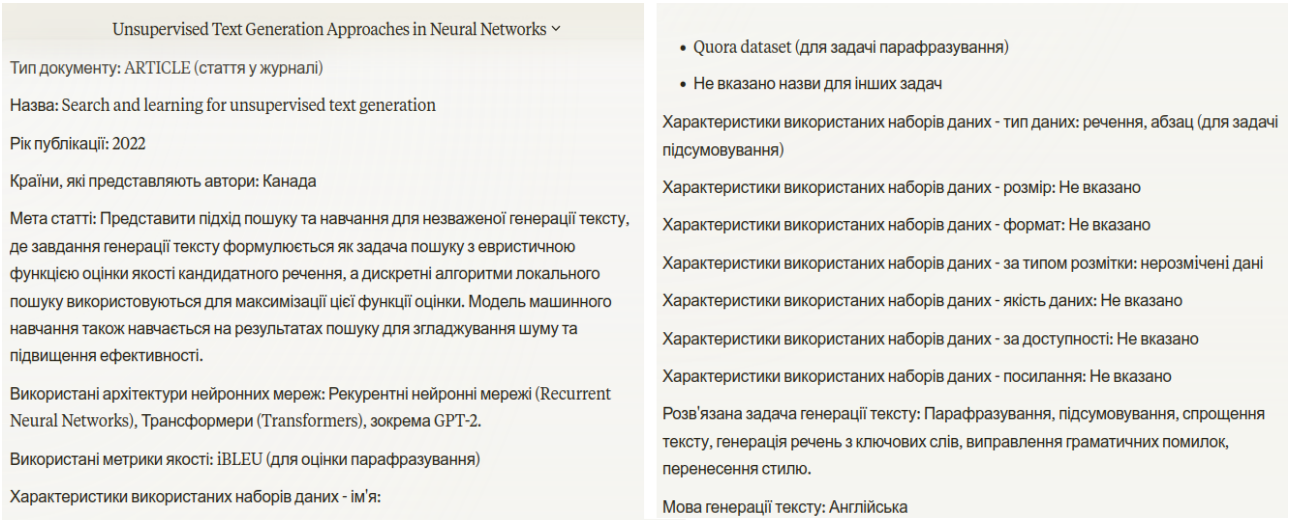

Рисунок 6. Відповідь Claude 3 Sonnet на запит щодо складання карти огляду статті [47]

Таблиця 2. Порівняння результатів роботи великих мовних моделей Claude 3 Sonnet та GPT 4

Запит	**Відповідь Claude 3 Sonnet**	**Відповідь GPT 4**
Тип документу: стаття у журналі (ARTICLE) або стаття у матеріалах конференції (CONFERENCE)	ARTICLE (стаття у журналі)	Article in a journal
Назва	Search and learning for unsupervised text generation	Search and Learning for Unsupervised Text Generation
Рік публікації	2022	2022
Країни, які представляють автори	Канада	Canada (The author, Lili Mou, is affiliated with the University of Alberta in Edmonton, Canada)
Мета статті	Представити підхід пошуку та навчання для незваженої генерації тексту, де завдання генерації тексту формулюється як задача пошуку з евристичною функцією оцінки якості кандидатного речення, а дискретні алгоритми локального пошуку використовуються для максимізації цієї функції оцінки. Модель машинного навчання також навчається на результатах пошуку для згладжування шуму та підвищення ефективності	The article aims to introduce approaches to unsupervised text generation using search and learning methodologies. It focuses on overcoming the need for massive labeled datasets, which are required in supervised methods, thereby making text generation feasible for low-resource languages and reducing the human labor required for data annotation
Використані архітектури нейронних мереж	Рекурентні нейронні мережі (Recurrent Neural Networks), Трансформери (Transformers), зокрема GPT-2	Neural architectures mentioned include recurrent neural networks (RNNs), Transformer architecture, and language models such as RoBERTa and pretrained language models by Radford et al., 2019
Використані метрики якості	iBLEU (для оцінки парафразування)	Metrics such as language fluency, semantic coherency, and task-specific constraints (like sentence simplicity scores) are used to evaluate the quality of generated texts
Характеристики наборів даних		
- ім’я	Quora dataset (для задачі парафразування). Не вказано назви для інших задач	Not specified in detail for individual datasets
- тип даних: речення, абзац, документ, питання-відповідь, не вказаний	речення, абзац (для задачі підсумовування)	Sentences, as the text generation involves tasks like paraphrasing, summarization, etc
- розмір	Не вказано	Specific sizes are not mentioned, but the use of large-scale generic corpora is implied
- формат: CSV, JSON, XML, файли, не вказаний	Не вказано	Not explicitly mentioned, but it includes labeled input-output pairs in various formats typical for NLP tasks

Запит	Відповідь Claude 3 Sonnet	Відповідь GPT 4
- за типом розмітки: розмічені або нерозмічені дані	нерозмічені дані	Primarily deals with labeled data in supervised contexts, though the focus here is on reducing dependence on such datasets
- якість даних: не опрацьовані (сирі), попередньо опрацьовані	Не вказано	The discussion includes the use of both raw and preprocessed data in different contexts
- загальнодоступний, приватний, не вказаний	Не вказано	The text suggests a shift towards methods that do not require publicly available or parallel datasets
- посилання	Не вказано	Not specific dataset URLs provided
Розв'язана задача генерації тексту (для чого застосовувалась нейронна мережа)	Парафразування, підсумовування, спрощення тексту, генерація речень з ключових слів, виправлення граматичних помилок, перенесення стилю	The neural networks are applied to generate text for various applications including paraphrasing, summarization, text simplification, and creating sentences from keywords, all in an unsupervised manner
Мова генерації тексту	Англійська	The tasks are primarily demonstrated with English examples, though the methodologies are applicable to multiple languages

Порівняння опису статті [47], виконаного за допомогою іншої великої мовної моделі та перевіреною людиною, з результатами з табл. 2 показує, що опис добре узгоджується з результатами роботи як Claude 3 Sonnet, так і GPT-4. Обидві моделі точно визначили тип документа, назву, рік публікації, країни авторів, мету статті, використані архітектури нейронних мереж, метрики якості та розв'язані задачі генерування тексту. Щодо характеристик наборів даних, обидві моделі вказали, що деталі про конкретні набори даних не надаються, за винятком набору даних Quora для парафразування. Вони також зазначили, що стаття зосереджується на зменшенні залежності від розмічених або публічно доступних наборів даних, хоча в різних контекстах обговорюються як розмічені, так і нерозмічені дані.

Отже, великі мовні моделі можуть точно видобувати ключову інформацію зі статей, хоча іноді упускають деталі, які явно не вказані в тексті. Для мінімізації ризику таких помилок виконано перевірку людиною результатів роботи Claude 3 Sonnet. Задля уникнення проблем, пов'язаних із перекладом термінології, відповіді великої мовної моделі додатково затребувано мовою відібраних документів, а саме англійською.

Оцінювання якості

Для оцінювання якості процесу видобування ключової інформацію зі статей застосуємо такі критерії:

- чіткість і відповідність критеріїв включення та виключення досліджень меті огляду;
- повнота та систематичність пошуку релевантних досліджень в обраних базах даних;
- послідовність і відтворюваність процесу відбору досліджень згідно з критеріями включення та виключення;
- застосування стандартизованої картки огляду для збору та систематизації даних із відібраних досліджень;

- залучення щонайменше двох незалежних дослідників до процесу відбору, аналізу та синтезу даних для мінімізації ризику упередженості;
- врахування та опис будь-яких розбіжностей або невизначеностей у процесі відбору та аналізу досліджень;
- забезпечення прозорості та відтворюваності процесу огляду шляхом детального опису кожного етапу у звіті.

Дотримання зазначених критеріїв якості забезпечує надійність і обґрунтованість результатів та висновків цього систематичного огляду.

PRISMA передбачає наявність у методиці дослідження таких додаткових компонентів:

- *оцінка ризику упередженості у відібраних дослідженнях* не є релевантною через те, що у цьому огляді розглядаються різні підходи та методи генерації тексту, а не порівнюються результати окремих досліджень;
- *визначення міри ефекту для кожного результату (або типу результату)* не виконується через те, що цей огляд не має на меті здійснення метааналізу чи кількісного синтезу результатів;
- *опис методів синтезу результатів досліджень*, як-от метааналіз, не виконується через те, що огляд не передбачає кількісного синтезу результатів;
- *оцінка ризику упередженості через неповноту подання результатів у публікаціях* не наводиться через те, що цей огляд фокусується на описі та класифікації описаних підходів і методів;
- *оцінювання достовірності та надійності результатів*, отриманих із публікацій, не здійснюється через використання надійних джерел, а саме видань з бази Scopus.

Розподіл відібраних документів за роками

Протягом 2022–2024 рр. кількість журнальних статей (ARTICLE) майже дорівнює кількості матеріалів конференцій (CONFERENCE) – 22 проти 21 (рис. 7). У 2022 р. кількість матеріалів конференції (15) значно перевищувала кількість статей у журналах (4), проте з 2023 р. збільшилась кількість статей у журналах (16), порівняно з матеріалами конференцій (6). За січень та лютий 2024 р. наявні лише статті у журналах (2), а матеріали конференцій відсутні. Така динаміка останніх років може свідчити про більш ґрунтовне висвітлення проблематики у наукових журналах, порівняно з матеріалами конференцій.

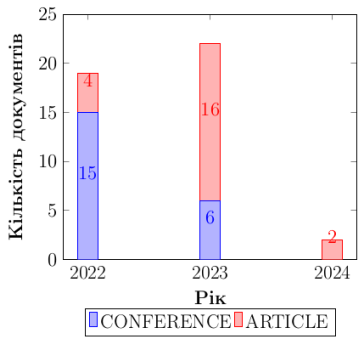

Рисунок 7. Розподіл статей за типом видання

Які передові методи глибокого навчання використовуються для генерування тексту в літературі 2022–2024 рр.?

Табл. 3 представляє огляд архітектур нейронних мереж, що використовуються для генерування тексту, згідно з даними публікацій 2022–2024 рр. Табл. 4 представляє узагальнення підходів до генерації тексту на основі даних табл. 3.

Таблиця 3. Архітектури нейронних мереж для генерації тексту

Архітектура	Опис	Представники	Статті
Традиційні підходи			
RNN (Recurrent Neural Networks)	Рекурентні нейронні мережі, що використовуються для обробки послідовних даних	–	[6, 9, 10, 11, 29, 30, 47]
LSTM (Long Short-Term Memory)	Варіант RNN, що краще запам'ятовує довгострокові залежності	–	[6, 10, 11, 13, 14, 15, 29, 30, 33, 41, 42]
GRU (Gated Recurrent Unit)	Спрощений варіант LSTM з меншою кількістю параметрів	–	–
CNN (Convolutional Neural Networks)	Згорткові нейронні мережі, що часто використовуються для обробки зображень	YOLOv5	[6, 9, 16, 38]
Graph Neural Networks	Моделі, що працюють із графовими структурами даних	GraphWriter, CGE-LW	[7, 9]
Інноваційні підходи			
Autoencoders	Мережі, які використовують для навчання ефективних кодувань нерозмічених даних	AE, VAE, iVAE, clVAE+ MI, β0.4 VAE, SaVAE, LagVAE	[15, 17, 29]
Transformer	Архітектура, що використовує механізм уваги для обробки послідовних даних	T5, CodeT5, TrICY, DETR	[7, 8, 9, 18, 19, 20, 22, 27, 31, 32, 34, 39, 41, 43, 44, 47, 48]
BERT (Bidirectional Encoder Representations from Transformers)	Модель на основі Transformer, що навчається на великих обсягах нерозміченого тексту	PubmedBERT, BioLinkBERT, RoBERTa, XLM-RoBERTa	[8, 9, 12, 13, 18, 19, 20, 26, 28, 30, 32, 35, 37, 39, 40, 45]
GPT-2, GPT-3 (Generative Pre-trained Transformer)	Моделі на основі Transformer, що використовуються для генерації тексту	OPT, Llama, CodeBERT	[6, 8, 10, 11, 12, 13, 15, 18, 19, 21, 22, 23, 24, 25, 26, 32, 33, 34, 36, 37, 39, 43, 44, 45, 47]
Attention-based models	Моделі, що використовують механізм уваги для покращення якості генерованого тексту	–	[8, 20, 26, 43, 44, 47]
Seq2Seq (Sequence-to-Sequence)	Архітектура, що використовує кодувальник та декодувальник для генерування послідовностей	S2ST, S2SL, S2SG, S2ST+, D+ Full, DSG	[15, 28, 31, 39, 42, 43, 46]
GAN (Generative Adversarial Networks)	Генеративно-змагальні мережі, що складаються з генератора та дискримінатора	EGAN, TILGAN, DoubAN-Full, WRGAN, CatGAN, SeqGAN, DGSAN	[6, 25, 29]

Архітектура	Опис	Представники	Статті
Memory Networks	Моделі, що використовують зовнішню пам'ять для зберігання та доступу до інформації	DM-NLG (with memory), MemNNs, Mem2Seq, GLMP	[9, 34]
Diffusion Models	Моделі, що використовують дифузійний процес для генерування тексту	GENIE, NAT, iNAT, ELMER, MASS, CMLM, ProphetNet, InsT, LevT, BANG, ConstLeven	[41]
Prompt-based models	Моделі, що використовують prompt-engineering донавчання для управління генеруванням тексту	–	[23]

Таблиця 4. Підходи до генерування тексту

Категорія	Статті
Традиційні підходи	[14, 16, 38]
Інноваційні підходи	[8, 12, 18, 19, 20, 21, 22, 23, 24, 25 26, 27, 28, 31, 32, 34, 35, 36, 37, 39, 40, 43, 44, 45, 46, 48]
Комбінація традиційних та інноваційних підходів	[6, 7, 9, 10, 11, 13, 15, 29, 30, 33, 41, 42, 47]

Серед інноваційних підходів найбільш популярним є використання моделей на основі архітектури Transformer, зокрема GPT-2, GPT-3, BERT та їх варіацій. Вони демонструють високу ефективність генерування зв'язного та семантично релевантного тексту. Також набувають популярності підходи з використанням механізмів уваги (attention) та контрольованого генерування тексту (controllable text generation). Традиційні підходи, хоча і використовуються рідше, все ще знаходять своє застосування у певних задачах, як-от генерування тексту на основі зображень, машинний переклад та інші. Спостерігається тенденція до переходу від традиційних підходів до більш інноваційних та ефективних моделей на основі архітектури Transformer та механізмів уваги. Це дає змогу покращити якість генерованого тексту та розширити сферу застосування цих технологій. Рис. 8 показує, що у 2022–2023 рр. переважають інноваційні підходи до генерування тексту. У 2024 р. кількість статей з інноваційним та комбінованим підходами однакова, проте вибірка за цей рік є неповною.

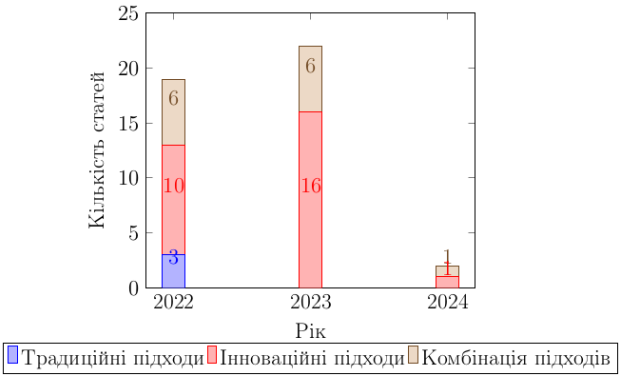

Рисунок 8. Розподіл статей за категоріями підходів до генерації тексту

Порівнюючи отримані результати з даними попереднього систематичного огляду [2], робимо такі висновки:

- традиційні підходи, як-от RNN, LSTM, CNN, все ще використовуються для генерування тексту, але меншою мірою, порівняно з інноваційними підходами;
- архітектура Transformer та її варіанти GPT-2, GPT-3 та BERT набули значної популярності у 2022–2024 рр., демонструючи високу ефективність генерування зв'язного семантично релевантного тексту;
- з'явилися нові архітектури та підходи, як-от Diffusion Models та Memory Networks models;
- значна увага приділяється моделям, що використовують механізми уваги (Attention-based models) та контрольованого генерування тексту (Controllable Text Generation);
- спостерігається тенденція до комбінування традиційних та інноваційних підходів для досягнення кращих результатів;
- у 2022–2024 рр. спостерігається перехід від традиційних підходів до більш інноваційних та ефективних моделей на основі архітектури Transformer та механізмів уваги, що дає змогу покращити якість генерованого тексту та розширити сферу застосування цих технологій.

За якими новими метриками оцінюють ефективність моделей генерування тексту в літературі 2022–2024 рр.?

Табл. 5 представляє огляд метрик якості, що використовуються для оцінювання згенерованого тексту. Метрики розділені на дві категорії: human-centred (орієнтовані на людину) та machine-centred (орієнтовані на машину). До human-centred метрик належать Human Evaluation та Turing Test, які передбачають оцінювання якості згенерованого тексту людьми-експертами або тест на здатність моделі генерувати текст, схожий на написаний людиною. Machine-centred метрики включають широкий спектр автоматичних метрик, як-от BLEU, ROUGE, METEOR, Perplexity, Distinct-n, BERTScore тощо. Ці метрики оцінюють різні аспекти якості згенерованого тексту – схожість з еталонним текстом, плавність, змістовність, різноманітність лексики та синтаксису тощо.

Табл. 6 надає огляд застосованих у статтях метрик якості. Більшість досліджень використовують machine-centred метрики для автоматичного оцінювання якості згенерованого тексту. Значно менша кількість досліджень застосовує human-centred метрики, що може бути пов'язано з трудомісткістю та суб'єктивністю оцінювання. Проте використання human-centred метрик все ще залишається важливим для отримання більш повної та надійної оцінки якості генерації тексту. Деякі дослідження не застосовують жодних метрик якості, що може бути пов'язано з фокусом на інших аспектах генерування тексту, як-от ефективність чи швидкість роботи моделей.

Використання різноманітних метрик якості є важливим для всебічної оцінки ефективності моделей та підходів до генерації тексту. Комбінування machine-centred та human-centred метрик дає змогу отримати більш надійні та валідні результати оцінювання. Діаграма на рис. 9 показує, що найчастіше використовується метрики BLEU (55.8% статей)

та ROUGE (48.8% статей). Також доволі поширеним є оцінювання якості людьми (Human Evaluation) – вона застосовується у 23.3% статей. Інші метрики, як-от Perplexity, METEOR, BERTScore та Distinct-n, використовуються рідше, але все ще мають значну частку згадувань у статтях. Найменш поширеними є метрики Turing Test, Fluency, Coherence, Diversity, N-gram Overlap та Embedding Similarity, кожна з яких згадується лише в одній статті (2.3%).

Таблиця 5. Основні метрики якості згенерованого тексту

Метрика якості	**Опис**	**Представники**	**Статті**
Human-centred metrics			
Human Evaluation	Оцінювання якості згенерованого тексту людьми-експертами	–	[9, 10, 11, 25, 30, 31, 32, 33, 36, 37]
Turing Test	Тест на здатність моделі генерувати текст, який неможливо відрізнити від написаного людиною	–	[33]
Machine-centred metrics			
BLEU	Метрика, що оцінює якість згенерованого тексту шляхом порівняння його з еталонним текстом	BLEU-1, BLEU-2, BLEU-3, BLEU-4, BLEU-5	[7, 8, 9, 10, 13, 15, 18, 19, 23, 27, 29, 31, 32, 33, 34, 36, 37, 41, 42, 43, 44, 45, 46, 47]
ROUGE	Метрика, що оцінює якість автоматичного реферування тексту	ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-L	[7, 8, 9, 10, 13, 18, 19, 23, 27, 28, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46, 48]
METEOR	Метрика, що оцінює якість машинного перекладу	–	[18, 27, 32, 34, 36, 42, 43, 44, 46, 48]
BERTScore	Метрика, що оцінює якість згенерованого тексту з використанням попередньо навченої моделі BERT	–	[8, 13, 18, 19, 26, 34, 32, 37]
CIDEr	Метрика, що оцінює якість автоматичного опису зображень, порівнюючи згенеровані описи з наборами референсних описів	–	[14, 18, 23, 36, 37, 41, 42, 46]
Perplexity	Метрика, що оцінює якість мовної моделі	–	[8, 9, 15, 17, 26, 29, 36, 39]
F1-score	Метрика, що оцінює якість класифікації, зокрема в задачах двокласової класифікації	–	[13, 20, 21, 26, 34, 40]
CHRF++	Метрика, що оцінює якість машинного перекладу, базуючись на збігах символів та n-грам	–	[7, 32, 37, 48]
Distinct-n	Метрика, що оцінює різноманітність згенерованого тексту	Dist-1, Dist-2, Dist-3, Dist-4	[8, 9, 15]

Таблиця 6. Огляд застосованих у статтях метрик якості

Метрики якості	**Статті**
Machine-centred	[7, 8, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28, 29, 34, 35, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]
Human-centred	[11, 30]
Обидві	[9, 10, 25, 32, 33, 36, 31, 37]
Не застосовано	[24, 12, 6]

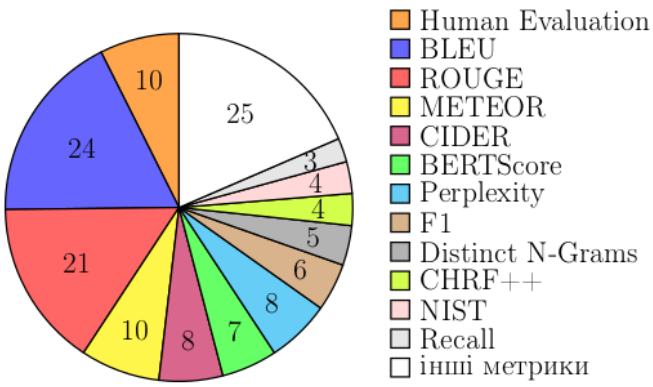

Рисунок 9. Розподіл метрик якості за кількістю статей, у яких вони згадані

Автоматичні метрики якості, як-от BLEU та ROUGE, найчастіше використовуються для оцінювання ефективності моделей генерування тексту, тоді як оцінювання якості людьми застосовується рідше, але залишається важливим компонентом для більш повної та надійної оцінки якості згенерованого тексту.

Порівнюючи отримані результати з даними попереднього систематичного огляду [2], можна виділити такі спостереження:

- BLEU та ROUGE залишаються найпопулярнішими метриками оцінювання якості згенерованого тексту як у 2015–2021 рр., так і у 2022–2024 рр.;
- Human Evaluation все ще широко застосовується для отримання більш повної та надійної оцінки якості генерації тексту, незважаючи на трудомісткість і суб'єктивність такого підходу;
- у 2022–2024 рр. з'явились нові метрики – BERTScore, Fluency, Coherence, Diversity, N-gram Overlap та Embedding Similarity, що засвідчує активний розвиток методів оцінювання якості згенерованого тексту та пошук більш ефективних та інформативних метрик;
- перплексія (Perplexity) набула більшої популярності у 2022–2024 рр., порівняно з попереднім періодом, що може бути пов'язано з її ефективністю в оцінюванні якості мовних моделей.
- метрика METEOR, за якою оцінюють якість машинного перекладу, також частіше використовується у 2022–2024 рр., що може свідчити про зростання інтересу до застосування генерування тексту в задачах машинного перекладу;
- загалом спостерігається тенденція до комбінування різних типів метрик (machine-centred та human-centred) для отримання більш надійних та валідних результатів оцінювання ефективності моделей генерації тексту.

Отже, порівняння результатів двох оглядів демонструє, що хоча традиційні метрики, як-от BLEU та ROUGE, залишаються популярними, у 2022–2024 рр. з'явилися нові метрики, які враховують різні аспекти якості згенерованого тексту. Це свідчить про активний розвиток методів оцінки якості й пошук більш ефективних та інформативних підходів до оцінювання моделей генерування тексту.

Які набори даних для генерування тексту описано в літературі 2022–2024 рр.?

Табл. 7 представляє набори даних з оглянутих статей. Вони спочатку впорядковані за спаданням кількості згадувань, а потім – за алфавітом. Набір даних E2E згадується найчастіше – у семи статтях, за ним слідують XSum, CNN/DailyMail, CommonGen, ToTTo та WebNLG, які згадуються у чотирьох статтях. У проаналізованих дослідженнях використовується широкий спектр наборів даних, що охоплюють різні домени і типи текстів, від відгуків користувачів та новинних статей до медичних і технічних текстів. Це свідчить про активний розвиток та застосування методів генерування тексту у різноманітних сферах.

Таблиця 7. Набори даних, про які згадано в оглянутих статтях

Назва набору даних	Статті
E2E	[19, 23, 30, 31, 34, 36, 44]
CNN/DailyMail (CNN/DM)	[9, 23, 41, 45]
Totto	[18, 31, 43, 46]
CommonGen	[9, 18, 36, 41]
WebNLG	[7, 31, 37, 44]
XSum	[9, 18, 23, 41]
WikiBio	[18, 31, 34]
Abstract Generation Dataset (AGENDA)	[7, 9]
DDI	[9, 12]
NIST	[9, 27]
PubMed	[12, 23]
Quora	[9, 47]
ROCStories	[9, 36]
Snips	[19, 39]
SST-2	[21, 45]
WMT’14 English-German	[18, 27]
WMT’16 Romanian-English	[18, 27]
Yelp	[17, 40]
Baidu Tieba; PersonaChat; Gigawords; Yahoo! Answers; NLPCC; Tencent; SQuAD; ComVE; αNLG-ART; EntDesc; VisualStory; PaperWriting; Reddit-10M; EMNLP dialog; ICLR dialog; NarrativeQA; Wizard of Wikipedia (WoW); MS-MARCO; ELI5; ChangeMyView; Amazon books; Foursquare	[9]
Scratch online community comments	[11]
BC5-Chemical; BC5-Disease; NCBI-Disease; BC2GM; JNLPBA; EBM PICO; ChemProt; GAD; BIOSSES; HoC; PubMedQA; BioASQ	[12]
Logic2Text	[13]
Concadia	[14]
REDIAL	[15]
Custom dataset for Bangla word sign language	[16]
Synthetic dataset; Penn Treebank	[17]
IWSLT’14 De-En	[18]
WMT16 English-German	[45]
WMT17 English-German	[36]
WMT20; WMT21	[37]
WMT’14 German-English	[27]
Multi-News; Java; Python	[18]

Назва набору даних	Статті
English ATIS; ViGGO; TREC; Korean Weather; Rest; KLUE-TC	[19]
C4; M2D2; Political Slant	[20]
Layoff; MC; M&A; Flood; Wildfire; Boston Bombings; Bohol Earthquake; West Texas Explosion; Dublin; New York City	[21]
WSC; CBT-CN; CBT-NE	[22]
Wikihow; SAMSum; DART	[23]
Custom dataset composed of tweets labeled with emotions	[25]
AFQMC; CHIP-STS; QQP; MRPC	[26]
ParaNMT-small; NIST Chinese-English	[27]
GTZAN	[28]
Minions; Japanimation; WikiArt; Nottingham; Lakh MIDI; TheoryTab; Poem-5; Poem-7	[29]
Synthetic date generation dataset	[30]
LDC2020T02 (AMR 3.0 release)	[32]
One Million Urdu News Dataset; Australian Broadcasting Corporation (ABC) news dataset	[33]
DailyMed drug labels	[35]
COCO Image Captioning	[37]
German and French commercial datasets; MASSIVE	[39]
Gold-PMB; Silver-PMB	[42]
numericNLG	[43]
Custom dataset related to text messaging applications	[44]
TweetEval; AGnews; QNLI; IMDB; CC-News	[45]
WITA	[46]
XWIKIREF	[48]

Табл. 8 представляє типи даних, які використані в оглянутих статтях. Типи даних впорядковано за спаданням кількості згадувань. Найчастіше використовуються набори даних, що містять речення – вони згадуються у 26 статтях. Це може бути пов’язано з тим, що багато завдань генерування тексту, як-от машинний переклад, парафразування, відповіді на запитання, вирішуються на рівні речень. Водночас наявність різноманітних типів даних, включно з абзацами, документами, зображеннями, музикою та іншим, свідчить про те, що методи генерування тексту можуть застосовуватись до широкого спектра задач та доменів.

Табл. 9 представляє типи розмітки даних, які використані в оглянутих статтях. Типи розмітки впорядковано за спаданням кількості згадувань. Найчастіше використовуються розмічені набори даних – вони згадуються у 22 статтях. Це може бути пов’язано з тим, що багато завдань генерування тексту, особливо ті, що використовують контрольовані підходи або вимагають відповідності певним шаблонам чи структурам, потребують розмічених даних для навчання моделей. Розмітка може включати такі елементи: частини мови, синтаксичні структури, семантичні ролі, теги для контрольованої генерації тощо.

Водночас наявність досліджень з нерозміченими даними або з комбінацією розмічених та нерозмічених даних свідчить про активний розвиток методів навчання без вчителя та напівавтоматичного навчання в галузі генерування тексту. Ці підходи дають змогу використовувати великі обсяги нерозмічених текстових даних для попереднього навчання моделей та покращення їх здатності до генерування зв’язного та змістовного тексту.

Таблиця 8. Типи даних, які використовуються в оглянутих статтях

Тип даних	Статті
Речення	[7, 9, 11, 12, 14, 15, 17, 18, 19, 20, 22, 23, 25, 26, 29, 30, 31, 32, 33, 36, 39, 40, 41, 45, 46, 48]
Абзац	[7, 9, 12, 15, 17, 18, 19, 20, 23, 29, 37, 14, 39, 40, 41, 45, 46, 48]
Документ	[9, 12, 18, 19, 20, 29, 35, 40, 41, 42, 45]
Питання-відповідь	[7, 9, 12, 17, 18, 19, 22, 26, 45, 47]
Таблиці з описом	[13, 21, 30, 31, 33, 34, 36, 43, 44]
Переклади	[18, 27, 31, 33, 36, 37, 45]
Історії	[9, 31, 33, 36]
Зображення	[14, 29, 37, 38]
Аудіофайли	[29, 28]
Відеокліпи	[16]
Комп’ютерні програми	[18]
Не вказаний	[6, 8, 10, 24]

Таблиця 9. Типи розмітки даних, які використовуються в оглянутих статтях

Тип розмітки	Статті
Розмічені дані	[12, 13, 14, 16, 17, 18, 21, 27, 28, 31, 33, 34, 35, 39, 40, 42, 43, 44, 46, 48, 47, 25]
Нерозмічені дані	[11, 12, 39, 40, 47]
Не вказано	[6, 7, 8, 9, 10, 15, 19, 20, 22, 24, 23, 26, 29, 30, 32, 36, 37, 38, 41, 45]

Табл. 10 представляє рівень якості даних з оглянутих статтях. Попередньо опрацьовані дані зазвичай проходять очищення, нормалізацію, токенізацію, а іноді й додаткову розмітку. Це дає змогу покращити якість і консистентність даних, а також полегшити навчання. Прикладами попередньо опрацьованих даних можуть бути набори даних із корпусів або баз даних, які вже пройшли певну обробку. Сирі дані беруться безпосередньо з реальних джерел, як-от вебсторінки, соціальні мережі, необроблені тексти тощо. Вони можуть містити шум, некоректне форматування, помилки та інші артефакти. Використання сирих даних може бути корисним для навчання моделей, які мають бути стійкими до реальних умов та здатними обробляти неструктуровані дані. Відсутність інформації про рівень якості даних у значній частині статей може свідчити про те, що автори не приділяють достатньої уваги цьому аспекту або вважають його маловажливим. Водночас якість даних є критичним фактором, що впливає на ефективність і узагальнювальність генерування тексту, тому варто приділяти більше уваги опису та аналізу якості даних у майбутніх дослідженнях.

Таблиця 10. Рівень якості наборів даних в оглянутих статтях

Рівень якості даних	Статті
Попередньо опрацьовані	[13, 14, 16, 17, 18, 44, 47, 48, 39, 34, 31, 42]
Сирі	[7, 11, 28, 33, 35, 37, 39, 34, 31, 42]
Не вказано	[6, 8, 9, 10, 12, 15, 20, 19, 21, 22, 24, 25, 23, 26, 27, 29, 30, 32, 36, 38, 40, 41, 43, 45, 46]

Порівнюючи результати огляду 2022–2024 рр. із попереднім оглядом [2], робимо такі висновки:

- у 2022–2024 рр. з'явилися нові набори даних, як-от XWIKIREF, DailyMed, numericNLG, WITA, DIST-ToTTo, що свідчить про активний розвиток ресурсів для дослідження та застосування методів генерування тексту;
- набори даних E2E, WikiBio, ToTTo, CommonGen, CNN/DailyMail та XSum залишаються популярними і широко використовуються в дослідженнях як у 2015–2021 рр., так і в 2022–2024 рр.;
- спостерігається тенденція до використання більш різноманітних типів даних, як-от таблиці з описом, зображення, музика, переклади, питання-відповідь, відеокліпи та комп'ютерні програми, на додачу до традиційних типів, як-от речення, абзаци та документи;
- розмічені дані залишаються найбільш широко використовуваними, але зростає інтерес до нерозмічених даних та комбінації розмічених і нерозмічених даних;
- хоча якість даних є критичним фактором, що впливає на ефективність генерування текстів, у значній частині досліджень 2022–2024 рр. цей аспект не висвітлюється, що може свідчити про необхідність приділяти більше уваги опису та аналізу якості використаних даних у майбутніх дослідженнях.

Отже, порівняння результатів двох оглядів демонструє, що набори даних для генерування тексту продовжують активно розвиватися, охоплюючи нові домени та типи даних. Водночас деякі популярні набори даних залишаються актуальними та широко використовуваними в дослідженнях. Спостерігається тенденція до використання більш різноманітних типів даних та зростання інтересу до нерозмічених даних і комбінованих підходів. Проте опис якості даних все ще потребує більшої уваги.

Які нові застосування генерування тексту описано в літературі 2022–2024 рр.?

Табл. 11 відображає застосування генерування тексту, які виявлено в аналізованих статтях. Найбільш поширеними застосуваннями є генерування анотацій та машинний переклад, за кожним із яких знайдено 8 статей. Аналіз застосувань демонструє широкий спектр можливостей використання генерування тексту у різних галузях, від обробки структурованих даних до створення емоційно забарвлених текстів та перекладу жестової мови в текст. Розвиток нових методів та архітектур нейронних мереж відкриває нові перспективи для подальшого розширення сфер застосування генерації тексту.

Порівнюючи результати огляду 2022–2024 рр. із попереднім оглядом [2], можна виділити такі спостереження:

- машинний переклад (Machine Translation) та генерування анотацій (Text Summarization) набули більшої популярності у 2022–2024 рр., порівняно з попереднім періодом. У 2022–2024 рр. з'явилися публікації з генерування текстів із таблиць та структурованих даних, що може вказувати на зростання інтересу до такої обробки структурованої інформації;

- контрольоване генерування тексту (Controllable Text Generation) також стало більш поширеним, що свідчить про зростання інтересу до управління генеруванням тексту для отримання більш релевантних та якісних результатів;
- генерування медичних текстів (Medical Text Generation) з'явилося як новий окремий напрям, що може бути пов'язано з активним розвитком методів обробки медичних даних та потребою в автоматизації створення медичної документації;
- з'явилися нові застосування, як-от генерування емоційно забарвлених текстів (Emotional Text Generation), генерування енциклопедичних статей (Encyclopedic Text Generation), генерування технічної документації (Technical Documentation Generation) та переклад жестової мови в текст (Sign Language to Text Translation), що свідчить про розширення сфер використання генерування тексту.
- парафразування та доповнення даних є актуальним застосуванням генерування тексту як у 2015–2021 рр., так і в 2022–2024 рр.;
- деякі застосування, які були популярними у попередньому огляді, як-от генерування поезії, діалогові системи, класифікація текстів і тематичне моделювання, у новому огляді не фігурують серед найбільш згадуваних, що може бути пов'язано зі зміною фокусу досліджень та появою нових перспективних напрямів;
- загалом спостерігається зростання різноманітності застосувань у 2022–2024 рр., порівняно з попереднім періодом, що свідчить про активний розвиток досліджень з генерування тексту та розширення можливостей використання генеративних моделей для вирішення прикладних завдань у різних предметних областях.

Таблиця 11. Застосування генерування тексту

Застосування	**Статті**
Генерування анотацій (Text Summarization)	[9, 18, 23, 41, 37, 45, 47, 48]
Машинний переклад (Machine Translation)	[9, 16, 17, 18, 27, 36, 37, 47]
Генерування тексту зі структурованих даних (Data-to-Text Generation)	[8, 31, 34, 44, 46]
Генерування тексту з таблиць (Table-to-Text Generation)	[13, 30, 34, 36, 43]
Парафразування (Paraphrasing)	[9, 27, 39, 47]
Доповнення даних (Data Augmentation)	[8, 21, 40, 42]
Контрольоване генерування тексту (Controllable Text Generation)	[8, 19, 30]
Генерування тексту за зображеням (Image-based Text Generation)	[14, 29, 37]
Генерування тексту з графів знань (Text Generation from Knowledge Graphs)	[7, 9]
Генерування медичних текстів (Medical Text Generation)	[12, 35]
Генерування емоційно забарвленого тексту (Emotional Text Generation)	[11, 25]
Генерування відповідей на запитання (Question Answering)	[9, 15]
Генерування музичних текстів (Music Text Generation)	[28, 29]
Генерування сценаріїв (Script Generation)	[9, 29]
Генерування новинних заголовків (News Headline Generation)	[33]
Генерування технічної документації (Technical Documentation Generation)	[10]
Кібербезпека (Cybersecurity)	[45]
Генерування енциклопедичних статей (Encyclopedic Text Generation)	[48]
Переклад жестової мови в текст (Sign Language to Text Translation)	[16]

Отже, порівняння результатів двох оглядів демонструє, що сфера застосування генерування тексту продовжує активно розширюватися, охоплюючи нові галузі та напрями. Популярність застосувань генерування тексту з таблиць та графів знань, контрольованого генерування тексту та генерування медичних текстів свідчить про зростання інтересу до методів, які дають змогу ефективно обробляти структуровані дані й отримувати більш релевантні та якісні результати. Водночас традиційні застосування, як-от парафразування, генерування анотацій та машинний переклад, залишаються актуальними й широко використовуваними в дослідженнях.

Які природні мови використовуються для генерування тексту відповідно до літератури 2022–2024 рр.?

Табл. 12 представляє щорічну статистику за мовами генерування тексту протягом досліджуваного періоду. Англійська мова є найбільш популярною протягом усіх трьох років. Для генерування англійських текстів використовуються різноманітні архітектури нейронних мереж, включно з Transformer, BERT, GPT-2, GPT-3, RNN, LSTM, CNN, GAN та Seq2Seq.

Таблиця 12. Статистика за мовами генерування

Мова	2022 р.	2023 р.	2024 р.	Разом	Архітектури
Англійська	17 статей [6, 14, 17, 18, 23, 28, 22, 25, 16, 42, 47, 31, 36, 37, 39, 30, 43]	19 статей [7, 8, 11, 12, 15, 19, 41, 20, 21, 26, 27, 32, 33, 34, 35, 40, 45, 46, 48]	2 статті [13, 44]	38 статей	Transformer, BERT, GPT-2, GPT-3, RNN, LSTM, CNN, GAN, Seq2Seq
Німецька	3 статті [39, 37, 18]	2 статті [45, 27]	–	5 статей	Conditional GAN, StyleGAN, DCGAN
Китайська	1 стаття [37]	3 статті [27, 29, 10]	–	4 статті	Graph Neural Networks, B2T
Французька	1 стаття [39]	–	–	1 стаття	Conditional GAN, StyleGAN, DCGAN
Бенгальська	1 стаття [16]	1 стаття [48]	–	2 статті	CNN, YOLO, mBART
Урду	–	1 стаття [33]	–	1 стаття	GPT-2
Хінді, малаялам, маратхі, орія, панджабі, тамільська	–	1 стаття [48]	–	1 стаття	HipoRank, mBART, mT5
Шекспірівська англійська	1 стаття [29]	–	–	1 стаття	Modified DCGAN
Румунська	1 стаття [18]	1 стаття [27]	–	2 статті	DCGAN, BART
Корейська	–	1 стаття [19]	–	1 стаття	Modified DCGAN

Німецька мова представлена у 5 статтях із використанням архітектури GAN (Conditional GAN, StyleGAN та DCGAN). Китайська мова представлена у 4 статтях із використанням Graph Neural Networks та архітектури B2T. Бенгальська мова представлена у 2 статтях з використанням CNN та YOLO. Румунська мова представлена у 2 статтях із використанням архітектур DCGAN та BART. Французька, урду, шекспірівська англійська та

корейська мови згадуються по одній статті кожна, з використанням різних архітектур, як-от Conditional GAN, StyleGAN, DCGAN та GPT-2. У 2023 р. у [48] досліджується генерування текстів одразу кількома індійськими мовами – хінді, малаялам, маратхі, орія, панджабі та тамільська – з використанням архітектур HipoRank, mBART та mT5.

Порівнюючи результати огляду 2022–2024 рр. із попереднім оглядом [2], можна виділити такі спостереження:

- англійська залишається найбільш широко використовуваною мовою для генерування тексту як у 2015–2021 рр., так і в 2022–2024 рр., проте збільшується кількість досліджень щодо генерування текстів іншими мовами, особливо мовами з обмеженими ресурсами;
- у 2022–2024 рр. з'явилися перші дослідження про генерування текстів на урду, хінді, малаялам, маратхі, орія, панджабі та тамільські мові, що свідчить про зростаючий інтерес до мовної диверсифікації моделей генерування тексту;
- стаття [48] демонструє можливість генерування текстів одразу кількома індійськими мовами з використанням сучасних архітектур, як-от HipoRank, mBART та mT5, що не було представлено в попередньому огляді;
- для генерування текстів різними мовами використовуються як традиційні архітектури (RNN, LSTM, CNN), так і більш сучасні підходи, як-от Transformer, BERT, GPT-2, GPT-3, GAN та Graph Neural Networks;
- спостерігається тенденція до розширення спектра мов моделей генерування тексту та використання більш різноманітних архітектур нейронних мереж.

Отже, порівняння результатів двох оглядів демонструє, що хоча англійська мова залишається домінуючою в дослідженнях із генерування тексту, але зростає інтерес до розробки моделей для інших мов, особливо мов з обмеженими ресурсами. Поява досліджень, присвячених генеруванню текстів мовами урду, хінді, малаялам, маратхі, орія, панджабі та тамільською свідчить про розширення мовних можливостей генераторів тексту. Використання сучасних архітектур нейронних мереж, як-от Transformer, BERT, GPT-2, GPT-3, GAN та Graph Neural Networks, дає змогу покращити якість та ефективність генерування тексту на різних мовах.

Порівнюючи розподіл мов у старому та новому оглядах із розподілом мов за кількістю моделей на Hugging Face [52], звернемо увагу на такі факти:

- англійська мова домінує в усіх трьох розподілах. У старому та новому оглядах вона найчастіше використовується для генерування тексту, а на Hugging Face для неї доступно найбільше моделей – 51 738. Це свідчить про значну увагу дослідників і розробників до англійської мови та про велику кількість ресурсів для неї;
- китайська мова посідає друге місце за кількістю моделей на Hugging Face (4 546) та згадується в кількох статтях у новому огляді. Це вказує на зростаючий інтерес до генерації тексту китайською мовою та розвиток відповідних ресурсів;
- французька, іспанська, російська та німецька мови мають значну кількість моделей на Hugging Face – від 2 326 до 4 049, але рідше згадуються в оглядах. Це може свідчити

про те, що, незважаючи на наявність ресурсів, дослідження з генерування тексту для цих мов не так широко представлені в літературі;

- мови з обмеженими ресурсами, як-от бенгальська, урду, арабська та хінді, згадуються в новому огляді, що свідчить про зростаючий інтерес до розробки моделей генерування тексту для цих мов. Однак кількість доступних моделей на Hugging Face для цих мов (від 670 до 1 674) значно менша, ніж для англійської;
- на Hugging Face представлено понад 200 мов, що значно більше, ніж згадується в оглядах. Це вказує на те, що дослідження з генерування тексту охоплюють лише частину мов, для яких доступні моделі та ресурси;
- японська, корейська, індонезійська, арабська та деякі інші мови мають значну кількість моделей на Hugging Face (від 1 674 до 2 920), але майже не згадуються в оглядах. Це може свідчити про потенціал для подальших досліджень із генерування тексту цими мовами.

Результати аналізу розподілів мов показують, що, незважаючи на домінування англійської мови в дослідженнях та наявних ресурсах, спостерігається зростаючий інтерес до генерування тексту іншими мовами, особливо мовами з обмеженими ресурсами. Однак кількість доступних моделей та ресурсів для цих мов все ще значно менша, порівняно з англійською. Доступність великої кількості моделей на Hugging Face для деяких мов, які рідко згадуються в оглядах, вказує на потенціал подальших досліджень та розробок у цій галузі.

Висновки

За результатами проведеного систематичного огляду застосування штучних нейронних мереж для генерування текстового контенту у 2022–2024 рр. виявлено 6 тенденцій:

1. Збільшується кількість статей у наукових журналах, порівняно з матеріалами конференцій, що може свідчити про більш ґрунтовне висвітлення питань генерування тексту.
2. З-поміж передових методів глибокого навчання для генерування тексту домінують моделі на основі архітектури Transformer – GPT-2, GPT-3, BERT та їх модифікації. Також набувають популярності підходи з використанням механізмів уваги та контрольованого генерування тексту. Загалом спостерігається тенденція до переходу від традиційних до більш інноваційних та ефективних моделей.
3. Найбільш широко використовуваними метриками для оцінювання ефективності моделей генерування тексту залишаються BLEU та ROUGE, а також оцінювання якості людьми (Human Evaluation). Водночас у 2022–2024 рр. з'явилися нові метрики, як-от BERTScore, Fluency, Coherence, Diversity, N-gram Overlap та Embedding Similarity, що свідчить про активний розвиток методів оцінювання якості згенерованого тексту.
4. Набори даних для генерування тексту продовжують активно розвиватися, охоплюючи нові домени та типи даних. Використовується більше різноманітних типів даних – таблиці з описом, зображення, музика, переклади тощо – та зростає інтерес до нерозмічених даних і комбінованих підходів.

5. Сфера застосування генерування тексту розширюється, охоплюючи нові галузі та напрями. Популярним є застосування генерування тексту з таблиць та графів знань, контрольоване генерування тексту та генерування медичних текстів, що свідчить про зростання інтересу до методів, які дають змогу ефективно обробляти структуровані дані та отримувати більш релевантні і якісні результати.
6. Хоча англійська мова продовжує домінувати в дослідженнях із генерування тексту, спостерігається зростаючий інтерес до розробки моделей на інших мовах, особливо на мовах з обмеженими ресурсами. Використання сучасних архітектур нейронних мереж дає змогу покращити якість та ефективність генерування тексту для різних мов.

Представлений систематичний огляд доповнює та розширює попередні дослідження, зокрема огляд [2] за 2015–2021 рр., шляхом аналізу новітніх досягнень у сфері нейронного генерування тексту продовж 2022–2024 рр. На відміну від попередніх оглядів, приділено особливу увагу інноваційним архітектурам моделей, як-от Transformer-based (GPT-2, GPT-3, BERT), механізмам уваги та контрольованому генеруванню тексту. До того ж огляд висвітлює появу нових метрик, зокрема BERTScore та Diversity Score, що доповнюють традиційні показники якості, як-от BLEU та ROUGE.

Ще однією відмінністю огляду є акцент на зростанні інтересу до застосування генеративних моделей для низькоресурсних мов та розширенні сфер застосування, включно із створенням анотацій, машинним перекладом, генеруванням тексту на основі таблиць та графів знань, а також генеруванням медичних текстів. Ці результати підкреслюють потенціал нейронного генерування тексту для вирішення широкого спектра прикладних завдань у різних предметних областях.

Серед найбільш перспективних технологій та підходів, що з'явилися у 2022–2024 рр., відзначимо моделі на основі архітектури Transformer, зокрема GPT-3 та його модифікації. Ці моделі демонструють вражаючі результати у генеруванні зв'язного та семантично релевантного тексту; вони здатні генерувати тексти на основі невеликої кількості прикладів (few-shot learning) та можуть бути адаптовані для вирішення широкого спектра завдань NLP. Іншим перспективним напрямом є контрольоване генерування тексту, яке дає змогу управляти стилем, тональністю та семантикою згенерованого тексту за допомогою додаткових сигналів керування.

Водночас актуальними залишаються виклики, пов'язані з адаптацією моделей генерування тексту для низькоресурсних мов та розробкою ефективних підходів до генерування тексту в умовах обмеженості навчальних даних. Попри певний прогрес, все ще існує потреба в розробці спеціалізованих архітектур та методів попереднього навчання, які б дали змогу покращити якість генерування тексту в таких умовах.

Відзначимо потенціал дифузійних моделей (Diffusion Models), які використовують ітеративний денойзинг для генерування високоякісних текстів. Ці моделі можуть бути особливо ефективними для генерування довгих послідовностей і мають потенціал для покращення якості та різноманітності згенерованого тексту.

Важливим напрямом подальших досліджень є розробка більш інтерпретовних та пояснюваних моделей генерування тексту, які дають змогу краще зрозуміти особливості процесу генерації для різних мов та доменів і полегшують взаємодію між людиною та

штучним інтелектом. Окремої уваги заслуговують етичні питання та відповідні потенційні ризики. З розвитком потужних мовних моделей зростають ризики автоматичного створення фейкових новин, дезінформації та маніпулятивного контенту. Тому важливо досліджувати методи детекції та протидії таким загрозам, а також розробити етичні принципи та дієві рекомендації щодо відповідального використання технологій генерування тексту.

Отримані результати демонструють активний розвиток та еволюцію методів генерування текстового контенту з використанням штучних нейронних мереж у 2022–2024 рр., порівняно з попереднім періодом. Застосування нових архітектур, підходів та метрик дає змогу покращити якість і релевантність згенерованого тексту, а також розширити сферу застосування цих технологій.

Результати систематичного огляду можуть бути корисними для дослідників, розробників і практиків у галузі обробки природної мови та штучного інтелекту, оскільки вони надають актуальну інформацію про сучасні тенденції, методи й перспективні напрями розвитку нейронного генерування текстового контенту. Отримані висновки можуть бути використані під час вибору методів, архітектур, метрик та наборів даних для розробки нових моделей і систем генерування тексту, а також для визначення пріоритетних напрямів подальших досліджень у цій області з урахуванням актуальних викликів та етичних аспектів.

Література

1. Ganegedara, T. (2018). *Natural Language Processing with TensorFlow: Teach language to machines using Python's deep learning library*. Packt Publishing. https://tinyurl.com/3xps3c5u
2. Fatima, N., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Soomro, A. (2022). A systematic literature review on text generation using deep neural network models. *IEEE Access*, 10, 53490-53503. https://doi.org/10.1109/ACCESS.2022.3174108
3. OpenAI (2022). Introducing ChatGPT. https://openai.com/blog/chatgpt
4. (2023). Large language models – Google Trends. https://trends.google.com/trends/explore?date=2022-01-01%202023-12-21&q=large%20language%20models&hl=en
5. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71
6. Bas, A., Topal, M. O., Duman, Ç., & Van Heerden, I. (2022). A brief history of deep learning-based text generation. In *J. M. Alja'Am, S. AlMaadeed, S. A. Elseoud, & O. Karam (eds.), Proceedings of the International Conference on Computer and Applications* (pp. 1–4). IEEE. https://doi.org/10.1109/ICCA56443.2022.10039545
7. Zhu, J., Ma, X., Lin, Z., & De Meo, P. (2023). A quantum-like approach for text generation from knowledge graphs. *CAAI Transactions on Intelligence Technology*. https://doi.org/10.1049/cit2.12178
8. Zhang, H., Song, H., Li, S., Zhou, M., & Song, D. (2023). A survey of controllable text generation using transformer-based pre-trained language models. *ACM Computing Surveys*, 56, 64. https://doi.org/10.1145/3617680

9. Yu, W., Zhu, C., Li, Z., Hu, Z., Wang, Q., Ji, H., & Jiang, M. (2022). A survey of knowledge-enhanced text generation. *ACM Computing Surveys*, 54, 227. https://doi.org/10.1145/3512467
10. Wu, T., Wang, H., Zeng, Z., Wang, W., Zheng, H.-T., & Zhang, J. (2023). Enhancing text generation with cooperative training. *Frontiers in Artificial Intelligence and Applications*, 372, 2704-2711. https://doi.org/10.3233/FAIA230579
11. Du, H., Xing, W., & Pei, B. (2023). Automatic text generation using deep learning: providing large-scale support for online learning communities. *Interactive Learning Environments*, 31, 5021–5036. https://doi.org/10.1080/10494820.2021.1993932
12. Chen, Q., Sun, H., Liu, H., Jiang, Y., Ran, T., Jin, X., … Niu, Z. (2023). An extensive benchmark study on biomedical text generation and mining with ChatGPT. *Bioinformatics*, 39, btad557. https://doi.org/10.1093/bioinformatics/btad557
13. Alonso, I., Agirre, E. (2024). Automatic logical forms improve fidelity in table-to-text generation. *Expert Systems with Applications*, 238. https://doi.org/10.1016/j.eswa.2023.121869
14. Kreiss, E., Fang, F., Goodman, N. D., & Potts, C. (2022). Concadia: Towards image-based text generation with a purpose. In *Y. Goldberg, Z. Kozareva, & Y. Zhang (eds.).* Proceedings *of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 4667–4684). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.308
15. Rao, K. Y., Rao, K. S., & Narayana, S. V. S. (2023). Conditional-aware sequential text generation in knowledge-enhanced conversational recommendation system. *Journal of Theoretical and Applied Information Technology*, 101, 2820–2836. http://www.jatit.org/volumes/Vol101No7/30Vol101No7.pdf
16. Tazalli, T., Aunshu, Z. A., Liya, S. S., Hossain, M., Mehjabeen, Z., Ahmed, M. S., & Hossain, M. I. (2022). Computer vision-based Bengali sign language to text generation. In *5th IEEE International Image Processing, Applications and Systems Conference* (pp. 1–6). IEEE. https://doi.org/10.1109/IPAS55744.2022.10052928
17. Teng, Z., Chen, C., Zhang, Y., & Zhang, Y. (2022). Contrastive latent variable models for neural text generation. In J. Cussens & K. Zhang (Eds.), *Proceedings of Machine Learning Research* (Vol. 180, pp. 1928–1938). ML Research Press. https://proceedings.mlr.press/v180/teng22a.html
18. An, C., Feng, J., Lv, K., Kong, L., Qiu, X., Huang, X. (2022). CONT: contrastive neural text generation. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS'22 (p. 160). Curran Associates Inc., Red Hook, NY, USA. https://dl.acm.org/doi/10.5555/3600270.3600430
19. Seo, H., Jung, S., Jung, J., Hwang, T., Namgoong, H., & Roh, Y.-H. (2023). Controllable text generation using semantic control grammar. *IEEE Access*, 11, 26329-26343. https://doi.org/10.1109/ACCESS.2023.3252017
20. Zhou, W., Jiang, Y. E., Wilcox, E., Cotterell, R., & Sachan, M. (2023). Controlled text generation with natural language instructions. In *A. Krause, E. Brunskill, C. K., B. Engelhardt, S. Sabato, & J. Scarlett (eds.). Proceedings of Machine Learning Research* (Vol. 202, pp. 42602–42613). ML Research Press. https://proceedings.mlr.press/v202/zhou23g/zhou23g.pdf

21. Bayer, M., Kaufhold, M.-A., Buchhold, B., Keller, M., Dallmeyer, J., Reuter, C. (2023). Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers. *International Journal of Machine Learning and Cybernetics*, 14, 135–150. https://doi.org/10.1007/s13042-022-01553-3
22. Hong, S., Moon, S., Kim, J., Lee, S., Kim, M., Lee, D., & Kim, J.-Y. (2022). DFX: A low-latency multi-FPGA appliance for accelerating transformer-based text generation. In *Proceedings of the Annual International Symposium on Microarchitecture* (pp. 616–630). IEEE Computer Society. https://doi.org/10.1109/MICRO56248.2022.00051
23. Ghazvininejad, M., Karpukhin, V., Gor, V., & Celikyilmaz, A. (2022). Discourse-aware soft prompting for text generation. In *Y. Goldberg, Z. Kozareva, & Y. Zhang (eds.).* Proceedings *of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 4570–4589). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.303
24. Koplin, J. J. (2023). Dual-use implications of AI text generation. *Ethics and Information Technology*, 25, 32. https://doi.org/10.1007/s10676-023-09703-z
25. Pautrat-Lertora, A., Perez-Lozano, R., & Ugarte, W. (2022). EGAN: Generatives adversarial networks for text generation with sentiments. In *F. Coenen, A. Fred, & J. Filipe (eds.). International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (Vol. 1, pp. 249-256). Science and Technology Publications. https://doi.org/10.5220/0011548100003335
26. Wu, J., Guo, Y., Gao, C., & Sun, J. (2023). An automatic text generation algorithm of technical disclosure for catenary construction based on knowledge element model. *Advanced Engineering Informatics*, 56, 101913. https://doi.org/10.1016/j.aei.2023.101913
27. Li, Y., Cui, L., Yan, J., Yin, Y., Bi, W., Shi, S., & Zhang, Y. (2023). Explicit syntactic guidance for neural text generation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 14095–14112). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.788
28. Chu, X. (2022). Feature extraction and intelligent text generation of digital music. *Computational Intelligence and Neuroscience*, 2022. https://doi.org/10.1155/2022/7952259
29. Shahriar, S. (2022). GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network. *Displays*, 73, 102237. https://doi.org/10.1016/j.displa.2022.102237
30. Strobelt, H., Kinley, J., Krueger, R., Beyer, J., Pfister, H., & Rush, A. M. (2022). GenNI: Human-AI collaboration for data-backed text generation. *IEEE Transactions on Visualization and Computer Graphics*, 28, 1106–1116. https://doi.org/10.1109/TVCG.2021.3114845
31. Yin, X., & Wan, X. (2022). How do Seq2Seq models perform on end-to-end data-to-text generation? In *S. Muresan, P. Nakov, & A. Villavicencio (eds.). Proceedings of the Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 7701–7710). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.531
32. Montella, S., Nasr, A., Heinecke, J., Bechet, F., & Rojas-Barahona, L. M. (2023). Investigating the effect of relative positional embeddings on AMR-to-text generation with structural adapters. In *EACL 2023 – 17th Conference of the European Chapter of the*

Association for Computational Linguistics (pp. 727–736). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.eacl-main.51

33. Fatima, N., Daudpota, S. M., Kastrati, Z., Imran, A. S., Hassan, S., Elmitwally, N. S. (2023). Improving news headline text generation quality through frequent POS-Tag patterns analysis. *Engineering Applications of Artificial Intelligence*, 125, 106718. https://doi.org/10.1016/j.engappai.2023.106718
34. Seifossadat, E., & Sameti, H. (2024). Improving semantic coverage of data-to-text generation model using dynamic memory networks. *Natural Language Engineering*, 30, 454-479. https://doi.org/10.1017/S1351324923000207
35. Meyer, C., Adkins, D., Pal, K., Galici, R., Garcia-Agundez, A., & Eickhoff, C. (2023). Neural text generation in regulatory medical writing. *Frontiers in Pharmacology*, 14. https://doi.org/10.3389/fphar.2023.1086913
36. Lu, X., Welleck, S., West, P., Jiang, J., Kasai, D., Khashabi, R., Le Bras, L., Qin, Y., Yu, R., Zellers, N. A., Smith, Y., & Choi, Y. (2022). NEUROLOGIC AFesque decoding: Constrained text generation with lookahead heuristics. In *NAACL 2022 – 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 780–799). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.57
37. Xu, W., Tuan, Y., Lu, Y., Saxon, M., Li, L., & Wang, W. Y. (2022). Not all errors are equal: Learning text generation metrics using stratified error synthesis. In *Y. Goldberg, Z. Kozareva, & Y. Zhang (eds.). Findings of the Association for Computational Linguistics: EMNLP 2022* (pp. 6588–6603). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-emnlp.489
38. Hanafi, A., Bouhorma, M., & Elaachak, L. (2022). Machine learning-based augmented reality for improved text generation through recurrent neural networks. *Journal of Theoretical and Applied Information Technology*, 100, 518–530. http://www.jatit.org/volumes/Vol100No2/18Vol100No2.pdf
39. Le, H., Le, D.-T., Weber, V., Church, C., Rottmann, K., Bradford, M., & Chin, P. (2022). Semi-supervised adversarial text generation based on Seq2Seq models. In *EMNLP 2022 – Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track* (pp. 264–272). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-industry.26
40. Yue, X., Inan, H. A., Li, X., Kumar, G., McAnallen, J., Shajari, H., Sun, H., Levitan, D., & Sim, R. (2023). Synthetic text generation with differential privacy: A simple and practical recipe. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 1321–1342). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.74
41. Lin, Z., Gong, Y., Shen, Y., Wu, T., Fan, Z., Lin, C., Duan, N., & Chen, W. (2023). Text generation with diffusion language models: A pre-training approach with continuous paragraph denoise. In *Proceedings of the 40th International Conference on Machine Learning*. JMLR.org. https://dl.acm.org/doi/abs/10.5555/3618408.3619275

42. Amin, M. S., Mazzei, A., Anselma, L. (2022). Towards Data Augmentation for DRS-to-Text Generation. *CEUR Workshop Proceedings*, 3287, 141–152. https://ceur-ws.org/Vol-3287/paper14.pdf
43. Chen, M., Lu, X., Xu, T., Li, Y., Zhou, J., Dou, D., Xiong, H. (2022). Towards Table-to-Text Generation with Pretrained Language Model: A Table Structure Understanding and Text Deliberating Approach. In *Y. Goldberg, Z. Kozareva, Y. Zhang (eds.). Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2022 (pp. 8199–8210). Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/2022.emnlp-main.562
44. Agarwal, V., Ghosh, S., BSS, H., Arora, H., Raja, B. R. K. (2024). TrICy: Trigger-Guided Data-to-Text Generation With Intent Aware Attention-Copy. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 1173–1184. https://doi.org/10.1109/TASLP.2024.3353574
45. Si, W. M., Backes, M., Zhang, Y., & Salem, A. (2023). Two-in-one: A model hijacking attack against text generation models. In *32nd USENIX Security Symposium* (Vol. 3, pp. 2223–2240). USENIX Association. https://www.usenix.org/system/files/usenixsecurity23-si.pdf
46. Gong, H., Feng, X., & Qin, B. (2023). Quality control for distantly-supervised data-to-text generation via meta learning. *Applied Sciences*, 13, 5573. https://doi.org/10.3390/app13095573
47. Mou, L. (2022). Search and learning for unsupervised text generation. *AI Magazine*, 43, 344-352. https://doi.org/10.1002/aaai.12068
48. Taunk, D., Sagare, S., Patil, A., Subramanian, S., Gupta, M., & Varma, V. (2023). XWikiGen: Cross-lingual summarization for encyclopedic text generation in low resource languages. In *ACM Web Conference 2023 – Proceedings of the World Wide Web Conference* (pp. 1703–1713). Association for Computing Machinery. https://doi.org/10.1145/3543507.3583405
49. Introducing the next generation of Claude (2024). https://www.anthropic.com/news/claude-3-family
50. Awesomegpts.ai (2024). Scholar GPT. https://chatgpt.com/g/g-kZ0eYXlJe-scholar-gpt?oai-dm=1
51. Slobodianiuk, A. V. (2024). Papers' review. https://docs.google.com/spreadsheets/d/e/2PACX-1vR6ZUaeeBjVgVl-do6QXm-Pua-HdztOxjC4DUqunrSDZ_-YSRz-Ng9xktYH9b0LDT502SiVy3YePx9F/pubhtml
52. Hugging Face (2024). Languages. https://huggingface.co/languages

Рукопис отримано – 04/03/2025 р.; прийнято до публікації – 26/03/2025 р.

Evolution of neural text generation models: A systematic review of research from 2022–2024

Artem Slobodianiuk, Serhiy Semerikov

Abstract

Recent years have witnessed significant advancements in neural text generation driven by the emergence of large language models and growing interest in this field. This systematic review identifies and summarizes current trends, approaches, and methods in neural text generation from 2022 to 2024, complementing a previous review covering 2015–2021. Following the PRISMA methodology, 89 articles were initially selected from the Scopus database, of which 43 articles remained after applying inclusion and exclusion criteria. The review reveals a shift towards innovative model architectures such as Transformer-based models (GPT-2, GPT-3, BERT), attention mechanisms, and controllable text generation. While BLEU, ROUGE, and human evaluation remain the most popular evaluation metrics, new metrics like BERTScore have emerged. Datasets span diverse domains and data types, with growing interest in unlabeled data. Applications have expanded to areas such as text summarization, machine translation, table-to-text generation, knowledge graph-based generation, and medical text generation. Although English dominates, there is increasing research on low-resource languages such as German and Chinese. The review also highlights current challenges in the field, including adapting models for low-resource languages, generating text with limited training data, and ethical considerations related to the use of powerful language models. The authors emphasize the importance of developing more efficient and interpretable architectures, improving controllable text generation methods, and creating new evaluation metrics. Taking into account current challenges and ethical considerations, the review also points to future research.

Keywords: neural architectures, neural text generation, controlled text generation, deep learning, systematic review, natural language processing, evaluation metrics, datasets, applications, low-resource languages, transformers, attention mechanisms.