

# ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ ТА МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

УДК 821.161.2:004.8

О. П. ІВАНОВ<sup>1\*</sup>, В. І. ШИНКАРЕНКО<sup>2\*</sup>, В. В. СКАЛОЗУБ<sup>3\*</sup>, А. А. КОСОЛАПОВ<sup>4\*</sup>

<sup>1\*</sup>Каф. «Комп'ютерні інформаційні технології», Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010, тел. +38 (056) 373 15 35, ел. пошта o.p.ivanov@ust.edu.ua, ORCID 0000-0003-1259-6377

<sup>2\*</sup>Каф. «Комп'ютерні інформаційні технології», Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010, тел. +38 (056) 373 15 35, ел. пошта v.i.shynkarenko@ust.edu.ua, ORCID 0000-0001-8738-7225

<sup>3\*</sup>Каф. «Комп'ютерні інформаційні технології», Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010, тел. +38 (056) 373 15 35, ел. пошта v.v.skalozub@ust.edu.ua, ORCID 0000-0002-1941-4751

<sup>4\*</sup>Каф. «Електронні обчислювальні машини», Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010, тел. +38 (050) 575 05 32, ел. пошта a.a.kosolapov@ust.edu.ua, ORCID 0000-0001-8878-568X

## Визначення авторства художнього українськомовного тексту засобами штучного інтелекту за надкороткими уривками

**Мета.** Інтелектуальну пошукову систему Bing можна використати як метод і засіб визначення автора художнього українськомовного тексту. Bing допомагає знайти інформацію про текстовий фрагмент та його автора, але результати пошуку можуть бути неточними або неповними. Основною метою роботи є дослідження ефективності встановлення авторства художніх текстів надсучасними засобами штучного інтелекту за надкороткими уривками. **Методика.** Відібрано десять українських авторів, із багатим наробком художніх творів, які відображають різні аспекти української культури та історії, а також випадкові фрагменти довжиною по 3–7 слів із різних творів цих авторів. Проведено експеримент із визначення авторства 2 000 фрагментів. **Результати.** За допомогою мови програмування Python та пакета skru розроблено програмне забезпечення, що передає запитання та отримує відповіді від вбудованого в Microsoft Skype боту Bing. У текстах відповідей перевірено наявність імені автора фрази та відповідної назви твору. За результатами Іван Франко має найвищий відсоток відповідей, де було згадано ім'я автора (65 %), а Олександр Довженко має найнижчий результат (23 %). Проаналізовано відповіді за довжиною фрагментів. Звісно, чим більша довжина фрагмента тексту, тим більша вірогідність точного встановлення його авторства. Особливості стилю автора проявляються у 20–40 % коротких фрагментах. Останні 60–80 % можуть складати загальноживані мовні конструкції, які ретрансльовано автором із зовнішнього середовища. **Наукова новизна.** У цій роботі вперше представлено методику перевірки авторства фрагментів українськомовного тексту за допомогою боту Bing зі штучним інтелектом. Виконано порівняльний аналіз та наведено експерименти з визначення авторства коротких фрагментів по 3–7 слів. Установлено що навіть досить малі фрагменти тексту мають ознаки, властиві оригінальному стилю автора художніх творів. **Практична значимість.** З'ясовано, якою мірою експерти з визначення авторства природномовних текстів можуть спиратись на наявні надсучасні засоби штучного інтелекту у поєднанні з широкою базою текстів в інтернет-просторі.

**Ключові слова:** визначення авторства; природномовний текст; штучний інтелект; генеративні мовні моделі; ChatGPT; бот Bing; Skype; Microsoft; Bard; Google

### Вступ

Визначення авторства художніх творів є складною проблемою, яка спричиняє багато дискусій у науковому світі. Це питання стосується не тільки літератури, але й інших видів

мистецтва, таких як живопис, музика та кіно. Визначення авторства може мати важливе значення для історії мистецтва, культури або права.

Встановлення авторства є складним завданням, оскільки існує багато факторів, які можуть вплинути на цей процес. Один із підходів полягає у використанні стилometriї – методу, який аналізує стиль написання тексту та порівнює його з відомими творами автора. Цей метод може бути ефективним у випадках, коли автор має унікальний стиль написання, але може бути менш ефективним, якщо автор намагається імітувати стиль іншого письменника.

Іншим підходом є аналіз граматики та лексики тексту. Наприклад, дослідники можуть шукати специфічні слова або фрази, які часто використовує автор, або аналізувати граматичні конструкції та інші особливості мови. Цей метод також може бути ефективним у деяких випадках, але має свої обмеження.

Окрім цих методів, існують інші підходи, такі як аналіз контексту та історичних документів, порівняння з іншими творами того ж жанру тощо. Важливо зазначити, що жоден із цих методів не є абсолютно точним, і кожен із них має свої переваги та недоліки. Тому визначення авторства художнього тексту часто є результатом комплексного аналізу за допомогою різних методів.

Генеративні мовні моделі є одним із напрямів, що найбільш активно розвивається в галузі штучного інтелекту. Ці моделі використовують для генерації тексту, перекладу, суміщення тексту та інших завдань, пов'язаних з обробкою природної мови. Перший із найбільш популярних типів генеративних мовних моделей – це послідовні моделі, зокрема рекурентні нейронні мережі (RNN) та їх варіації, зокрема LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit). Ці моделі використовують для генерації тексту, передбачення наступного слова в послідовності тощо. Останнім часом усе більшої популярності набувають трансформерні моделі, такі як BERT (Bidirectional Encoder Representations from Transformers) та GPT (Generative Pretrained Transformer). Вони застосовують механізм уваги для кращого розуміння контексту та здатні до генерації більш складних текстових послідовностей.

Мовні моделі призначені для генерації тексту та продовження речень, але вони не завжди здатні точно визначити авторство тексту. Однією з причин цього є те, що статистичні методи не завжди достатньо точні для визначення ав-

торства. Крім того, мовні моделі можуть генерувати текст, який схожий на стиль певного автора, але не є його твором.

Великі мовні моделі, такі як ChatGPT, використовують статистичні методи для передбачення наступного слова в реченні. Вони аналізують контекст речення та використовують його для визначення ймовірностей різних слів, які можуть бути наступними. Модель обирає слово з найвищою ймовірністю та пропонує його як наступне слово в реченні [8]. Статистична модель мови – це розподіл ймовірностей над послідовностями слів. За такою послідовністю, наприклад, довжини  $m$ , модель призначає ймовірність усього рядка. Модель мови забезпечує контекст для розрізнення слів та фраз, які звучать схоже [2].

Використання мовних моделей для завдань визначення авторства тексту може бути складним і потребує додаткових досліджень та розробки нових методик. Мовні моделі – це лише інструменти, які можуть допомогти у вирішенні певних завдань, але вони не завжди здатні точно визначити авторство тексту. Тому для визначення авторства художнього тексту часто потрібно використовувати комплексний підхід, який поєднує різні методи аналізу.

*Основні можливості боту Bing.* Bing – це пошукова система від компанії Microsoft, яка поєднує можливості мовної моделі з інтернет-пошуком. Це дозволяє користувачам шукати інформацію за допомогою природної мови та отримувати більш точні результати. Microsoft вбудовує Bing у багато своїх продуктів, таких як Windows, Office, Skype [9] та інші, щоб забезпечити користувачам зручний доступ до інформації. Це дозволяє користувачам швидко знаходити потрібну інформацію та підвищує продуктивність пошуку.

Пошукова система Bing працює на наступному поколінні мовної моделі OpenAI, яка є більш потужною, ніж ChatGPT, та спеціально налаштованою для пошуку [5]. Вона використовує ключові знання та досягнення з ChatGPT та GPT-4 і є ще швидшою, точнішою та більш функціональною [7]. Bing також розширює використання моделей представлення природної мови для розширення своїх функцій щодо глобального семантичного виділення та генерації відповідей на запитання.

Bing використовує пошук для відповіді на запити користувачів, щоб забезпечити найбільш точну та актуальну інформацію. Коли користувач вводить запит, система шукає відповідну інформацію у своїй базі даних та на інших ресурсах, щоб знайти найкращі результати. Bing також використовує розширені алгоритми ранжування та фільтрації, щоб забезпечити найбільш релевантні результати для кожного запиту. Ранжування належить до процесу оцінки та впорядкування результатів пошуку за їх релевантністю до запиту. Фільтрація – це процес вибору певних результатів на основі заданих критеріїв, наприклад, типу контенту, мови або дати публікації [3].

Bard – це новий AI-чат-бот від Google, який був запущений для конкуренції з ChatGPT та функцією AI-чату Bing [6]. З 10 травня 2023 року Google Bard доступний у понад 180 країнах світу, а не тільки у США та Великобританії. Bard може виконувати різні завдання, пов'язані з текстом, такі як надання відповідей та підсумків, створення різних форм контенту та інше.

Bard використовує інформацію з інтернету, щоб надавати відповіді на запити користувачів. А щоб отримати доступ до свіжої та актуальної інформації з інтернету, він використовує мовну модель PaLM 2 від Google, [4]. Це дозволяє Bard надавати більш точні та повні відповіді на запити користувачів. На відміну від інших AI-чат-ботів, таких як ChatGPT, Bard має кнопку «Google it», яка дає можливість користувачам перевірити факти та дізнатися, які джерела були використані, наприклад, Вікіпедія. Це дозволяє користувачам отримати більше інформації про те, як Bard формулює свої відповіді та звідки система отримує інформацію.

### Мета

Інтелектуальну пошукову систему Bing можна використати як метод і засіб визначення автора художнього україномовного тексту. Bing допомагає знайти інформацію про текстовий фрагмент та його автора, але результати пошуку можуть бути неточними або неповними.

Більшість авторів використовують загальні мовні конструкції, шаблони та мають свою авторську особливість. Це означає, що навіть надто малі фрагменти тексту можуть містити

певні авторські особливості, які дозволяють визначити автора тексту. Основною метою роботи є визначити, наскільки можливо та ефективно можна з'ясувати авторство художніх творів за короткими фрагментами тексту.

### Методика

Відібрано десять українських авторів, які створили багато творів, що відображають різні аспекти української культури та історії: Микола Хвильовий; Валер'ян Підмогильний; Всеволод Нестайко; Панас Мирний; Михайло Коцюбинський; Іван Франко; Олександр Довженко; Марко Вовчок; Остап Вишня; Іван Багряний. Було вибрано випадкові фрагменти довжиною по 3–7 слів із різних творів цих авторів. Загалом для експерименту відібрано близько 2 000 фрагментів.

База авторів і творів збігається з представленою у роботах [10, 11]. Фрагменти обрано випадковим чином із випадкових творів, без урахування пунктуації та структурних елементів. Тобто фрагмент може включати слова закінчення одного речення та початок наступного. Приклади фрагментів:

– «довгасті чорні очі» (автор М. Коцюбинський, твір «На камені»);

– «Мої очі і мої» (автор О. Вишня, твір «Вальдшнеп»);

– «піднімаючи довгі ноги потім зірвався й побіг» (автор В. Нестайко, твір «Чарівний талісман»);

– «вона схожа не на ріку а на» (автор О. Довженко, твір «Аероград»);

– «й діти малі й усі слухали усі» (автор М. Вовчок, твір «Невільничка»).

За допомогою мови програмування Python та пакета skru [1] було створено програму, що передає запитання та отримує відповіді від вбудованого у Skure боту Bing. Skru – це бібліотека для Python, яка дозволяє взаємодіяти з API Skure. Бібліотека автоматизує роботу зі Skure, дозволяє надсилати повідомлення та отримувати відповіді з чатів або розмов.

В експерименті боту Bing було запропоновано відповіді на запитання такого формату «Дай коротку відповідь, хто автор рядка: %?». Приклад схеми діалогу:

Програма надсилає запит до Bing «Дай коротку відповідь, хто автор рядка: ‘Як умру то поховайте?’».

Bing обробляє запит та шукає інформацію про автора цього рядка.

Bing відповідає: «Автором рядка ‘Як умру то поховайте’ є Тарас Шевченко, великий український поет і художник. Він написав цей рядок у своєму вірші “Заповіт”, який став символом національної свободи і гідності».

Зазначимо, що Bing зазвичай не тільки визначає авторство, тобто відповідає на пряме запитання, але й надає більш розширену інформацію щодо автора, твору та іншу додаткову інформацію (рис. 1).

Звісно, у разі зміни контексту діалогу можна отримати дещо інші результати, наприклад, якщо звукити пошук, запропонувавши перелік можливих авторів, або вказати регіон походження, роки написання. Проте в наших експериментах проведено діалог із ботом за наведеною вище схемою.

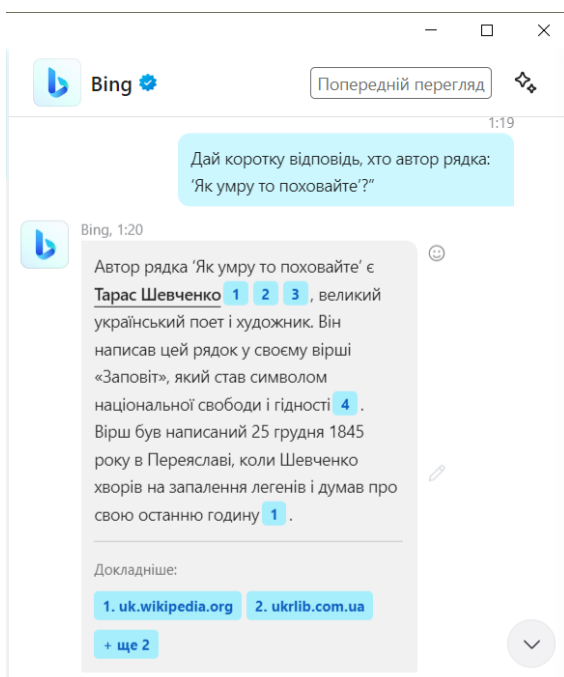


Рис. 1. Приклад діалогу з ботом Bing у програмі Skype

Fig. 1. An example of a dialog with the Bing bot in the Skype program

## Результати

Усі відповіді було зібрано в загальну таблицю для попередньої обробки. Частина відповідей віднесено до категорії помилкових: текст збережено з помилками, або текст відповіді стосується не поточного питання. Помилкові відповіді було видалено. Для подальшого аналізу використано 1 600 відповідей боту Bing. У текстах відповідей було перевірено присутність імені автора фрази та відповідної назви твору.

Наступна таблиця представляє відсоток відповідей, у яких правильно визначено ім'я автора. За результатами, Іван Франко має найвищий відсоток відповідей, де було згадано ім'я автора (65 %), а Довженко має найнижчий відсоток (23 %).

Таблиця 1

### Середній відсоток точного і помилкового визначення автора фрагмента

Table 1

#### The average percentage of accurate and incorrect identification of the author of a fragment

Автор	точно	помилково
М. Хвильовий	24 %	76 %
В. Підмогильний	25 %	75 %
В. Нестайко	40 %	60 %
П. Мирний	56 %	44 %
М. Коцюбинський	27 %	73 %
І. Франко	65 %	35 %
О. Довженко	23 %	77 %
М. Вовчок	47 %	53 %
О. Вишня	51 %	49 %
І. Багряний	36 %	64 %

Зведена діаграма (рис. 2) показує розподіл точного та помилкового визначення автора фрагмента.

Точність визначення авторства тексту підвищується зі збільшенням довжини фрази. Чим довший текст, тим більше інформації він містить про стиль написання, мову, лексику та інші ознаки, які допомагають визначити автора.

## ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ ТА МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

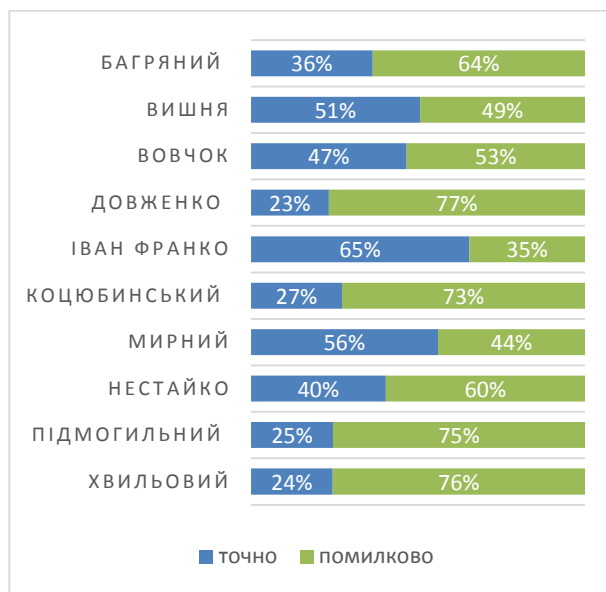


Рис. 2. Результат точного розпізнавання авторства

Fig. 2. The result of accurate authorship recognition

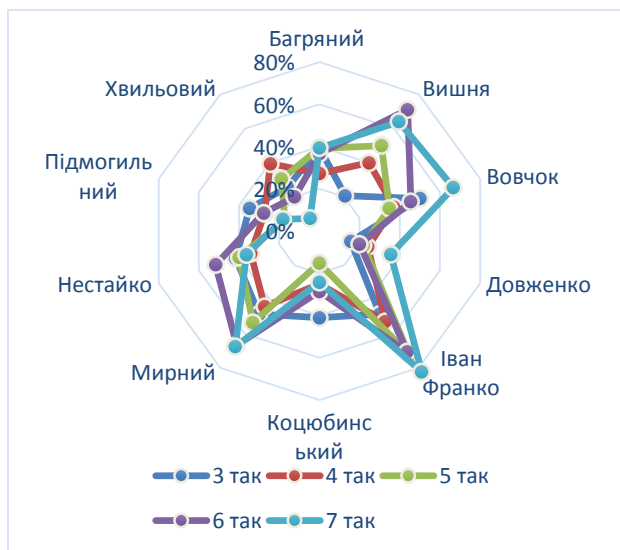


Рис. 3. Відсоток правильного визначення авторства за довжиною фрагмента

Fig. 3. Percentage of correct identification of the author by the length of the fragment

Унікальність автора може позначатися навіть на коротких фрагментах довжиною у 3–7 слів (рис. 2). Збільшення довжини текстового фрагмента сприяє правильному встановленню його автора (рис. 3).

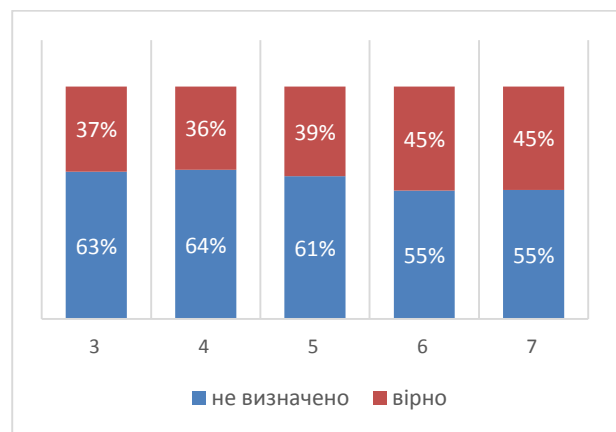


Рис. 4. Загальний розподіл точного визначення автора за довжиною фрагмента тексту

Fig. 4. General distribution of the exact author identification by the length of the text fragment

Розглянемо помилково встановлені назви творів. Отримані відповіді боту Bing, які мали точне визначення автора фрагмента, було перевірено на наявність назви твору, з якого фрагмент було взято. Відсутність назви твору у відповіді може свідчити про невисоку точність пошуку або прояв авторського стилю. Але самого запитання визначити назву твору за фрагментом тексту не було. Наведені результати показують найбільший відсоток правильного визначення авторства Івана Франка (рис. 1), але він має один із найнижчих (28 %) показників точного визначення назви твору (рис. 5).

Розглянемо результати помилкового встановлення авторства тексту.

Деякі фрагменти було визнано народними або взятими з історичних епосів. Це означає, що вони належать до народної традиції або літературного жанру, а не до конкретного автора. Визначення авторства таких фраз може бути складним або неможливим, оскільки вони часто переходили з покоління в покоління в усній формі та могли зазнавати змін і модифікацій.

Іноді авторство фрагмента може бути приписане більш популярному автору, навіть якщо він насправді належить менш відомому автору. Це може статися через ранжування джерел інформації або через неправильне сприйняття публікою. Крім того, фрази можуть зустрічатися в перекладах з іноземних творів, що також ускладнює визначення їх авторства.

## ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ ТА МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

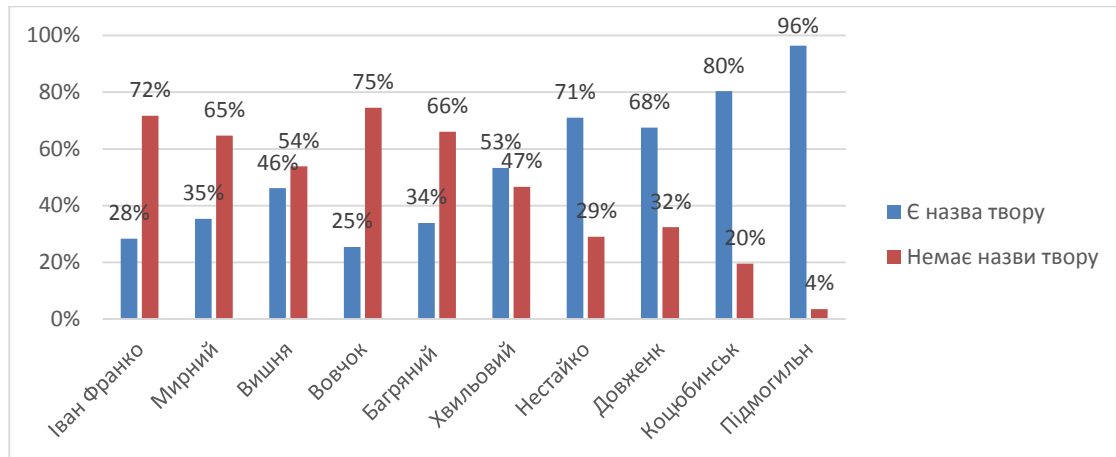


Рис. 5. Відсоток точного визначення назви твору у відповідях

Fig. 5. Percentage of accurate determination of the title of the work in the answers

Таблиця 2

**Загальний відсоток появи помилкового автора (представлено тільки відповіді більше 1 %)**

Table 2

**The total percentage of false authors (only answers over 1% are presented)**

Помилково визначений автор	Відсоток появи
Леся Українка	25 %
Тарас Шевченко	19 %
Іван Франко	17 %
Іван Котляревський	10 %
Ліна Костенко	8 %
Василь Симоненко	6 %
Іван Драч	3 %
Юрій Андрухович	2 %
Олександр Довженко	2 %
Василь Стус	2 %

Для відповідей боту Bing, у яких автор визначений неправильно, було проведено аналіз імені автора, з яким відбулося «сплутування». Таблиця демонструє, що з багатьох помилково встановлених імен можна виділити основних популярних україномовних авторів (рис. 6–7).

Розглянемо показники унікальності авторського стилю. Кожен автор має свій унікальний стиль написання, який можна визначити за допомогою різних показників. Ось деякі з них:

– словник: автори часто використовують унікальний набір слів та фраз, які характеризують їх стиль написання;

– граматики: структура речень та використання граматичних конструкцій також можуть бути показниками унікальності авторського стилю;

– тематика: автори часто пишуть на певні теми або жанри, які характеризують їх творчість;

– тон: тон написання, такий як сарказм, іронія або емоційність, також може бути показником унікальності авторського стилю.

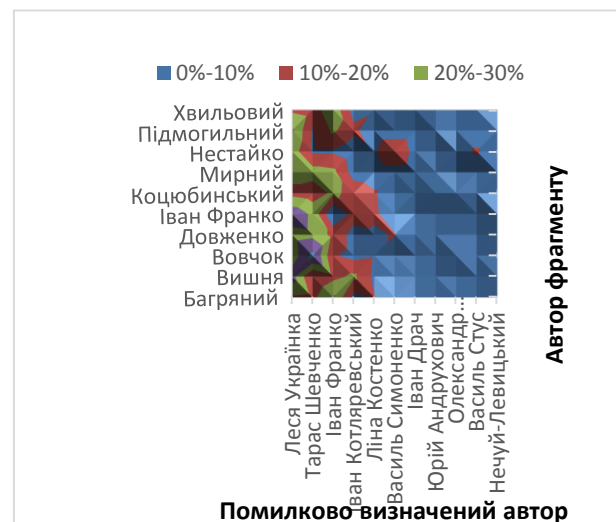


Рис. 7. Розподіл відсотка помилкового встановлення авторства фрагмента до справжнього автора

Fig. 7. Distribution of the percentage of false attribution of a fragment to the real author

## ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ ТА МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

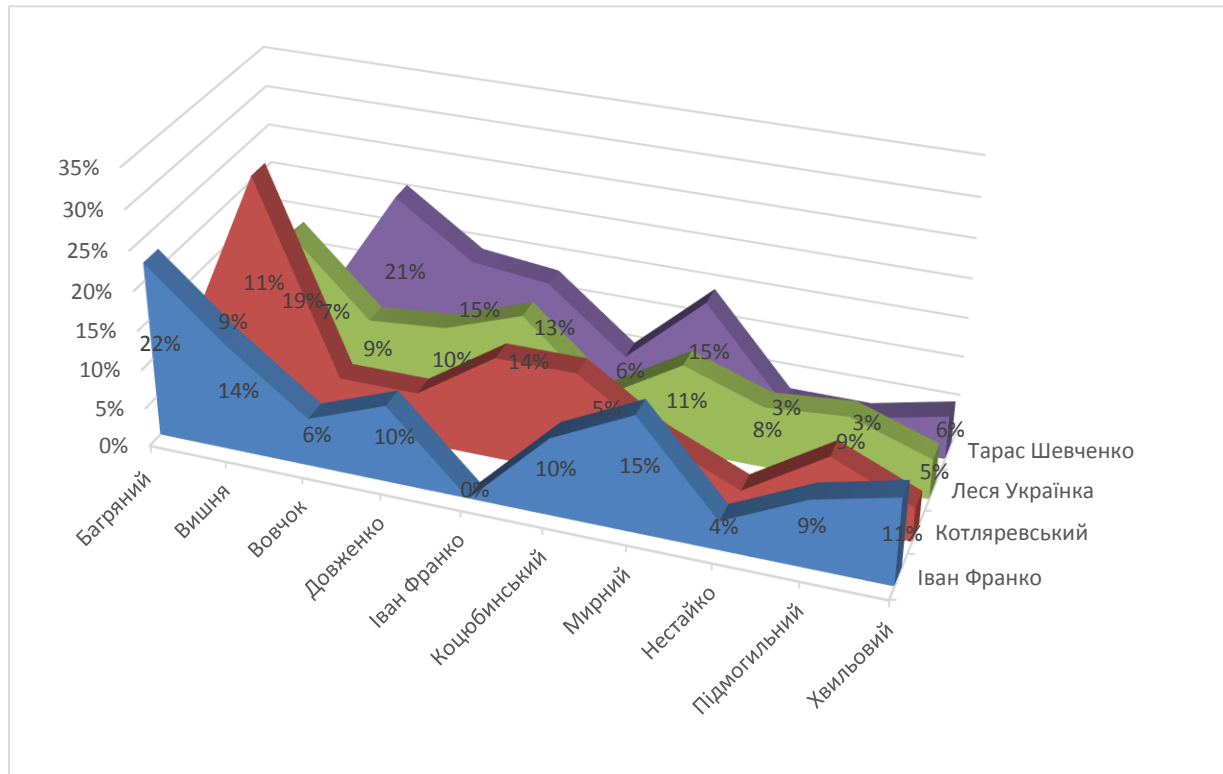


Рис. 8. Перша четвірка помилково визначених авторів, відсоток появи по авторам фрагментів

Fig. 8. The top four misidentified authors, percentage of appearances by author of fragments

Наведено лише деякі з показників, які можна використати для визначення унікальності авторського стилю. Кожен автор має індивідуальну комбінацію цих та інших показників, яка робить його стиль унікальним. Питання щодо методики формалізації запитів до інтелектуальних ботів з урахуванням показників унікальності стилю потребують подальших досліджень.

#### Наукова новизна та практична значимість

У цій статті вперше поставлено завдання розробити методику перевірки авторства фрагментів україномовного тексту за допомогою боту Bing зі штучним інтелектом. Виконано порівняльний аналіз та наведено експерименти визначення авторства коротких фрагментів із 3–7 слів.

#### Висновки

Установлено, що короткі та надкороткі фрагменти довжиною 3–7 слів уже дозволяють визначити автора художнього україномовного тексту у випадку творів майстерних авторів.

Особливості стилю автора проявляються у 20–40 % коротких фрагментах. Останні 60–80 % можуть складати загальноживані мовні конструкції, які ретрансльовано автором із зовнішнього середовища.

Рейтингування в системі пошуку та генеративні мовні моделі можуть бути використані для завдання визначення авторства тексту, але це не є точним методом. Системи пошуку дозволяють знаходити інформацію про текст та його автора, але результати пошуку можуть бути неточними або неповними. Генеративні мовні моделі, з іншого боку, можуть генерувати текст, який схожий на стиль певного автора, але не є його твором.



## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. An unofficial Python library for interacting with the Skype HTTP API. *SkPy 0.10.6*. 2023. URL: <https://pypi.org/project/SkPy/>
2. Bengio Y. «Neural net language models». *Scholarpedia*. 2008. Vol. 3. Iss. 1. P. 3881. DOI: <https://doi.org/10.4249/scholarpedia.3881>
3. Bonifacic I. *Microsoft's next-gen Bing uses a 'much more powerful' language model than ChatGPT*. 2023. URL: <https://www.engadget.com/microsofts-next-gen-bing-more-powerful-language-model-than-chatgpt-182647588.html>
4. Chowdhery A., Narang Sh., Devlin J., Bosma M., Mishra G., Roberts A., ... Fiedel N. PaLM : Scaling Language Modeling with Pathways. *arXiv*. 2022. P. 1–87. DOI: <https://doi.org/10.48550/arXiv.2204.02311>
5. *Confirmed: the new Bing runs on OpenAI's GPT-4*. 2023. URL: [https://blogs.bing.com/search/march\\_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4/](https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4/)
6. Hoffmangoogole Ch. Bard: How to Use Google's AI Chatbot. *How-To Geek*. URL: <https://www.howtogeek.com/880668/google-bard-how-to-use-googles-ai-chatbot/>
7. Montti R. How Bing AI Search Uses Website Content. *Search Engine Journal*. 2023. <https://www.searchenginejournal.com/how-bing-ai-search-uses-web-content/480643/>
8. Perez E., Kiela D., Cho K. True few-shot learning with language models. *Advances in Neural Information Processing Systems*. 2021. P. 34.
9. Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web... URL: <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>
10. Shynkarenko V. I., Demidovich I. M. Authorship Determination of Natural Language Texts by Several Classes of Indicators with Customizable Weights. *COLINS-2021 : 5th International Conference on Computational Linguistics and Intelligent Systems* (Kharkiv, 22-23 April 2021). Kharkiv, Ukraine, 2021. Vol. 1. P. 832–844.
11. Shynkarenko V. I., Demidovich I. M. Natural Language Texts Authorship Establishing Based on the Sentences Structure. *COLINS-2022 : 6th International Conference on Computational Linguistics and Intelligent Systems* (Gliwice, 12-13 May 2022). Gliwice, Poland, 2022. P. 328–337.

O. P. IVANOV<sup>1\*</sup>, V. I. SHYNKARENKO<sup>2\*</sup>, V. V. SKALOZUB<sup>3\*</sup>, A. A. KOSOLAPOV<sup>4\*</sup>

<sup>1\*</sup>Dep. «Computer Information Technologies», Ukrainian State University of Science and Technologies, Lazaryana St., 2, Dnipro, Ukraine, 49010, tel. +38 (056) 373 15 35, e-mail o.p.ivanov@ust.edu.ua, ORCID 0000-0003-1259-6377

<sup>2\*</sup>Dep. «Computer Information Technologies», Ukrainian State University of Science and Technologies, Lazaryana St., 2, Dnipro, Ukraine, 49010, tel. +38 (056) 373 15 35, e-mail v.i.shynkarenko@ust.edu.ua, ORCID 0000-0001-8738-7225

<sup>3\*</sup>Dep. «Computer Information Technologies», Ukrainian State University of Science and Technologies, Lazaryana St., 2, Dnipro, Ukraine, 49010, tel. +38 (056) 373 15 35, e-mail v.v.skalozub@ust.edu.ua, ORCID 0000-0002-1941-4751

<sup>4\*</sup>Dep. «Electronic Computing Machines», Ukrainian State University of Science and Technologies, Lazaryana St., 2, Dnipro, Ukraine, 49010, tel. +38 (050) 575 05 32, e-mail a.a.kosolapov@ust.edu.ua, ORCID 0000-0001-8878-568X

## Determining the Authorship of a Ukrainian-Language Literary Text by Means of Artificial Intelligence from Ultra-Short Excerpts

**Purpose.** The intelligent search engine Bing can be used as a method and a means of determining the author of a Ukrainian-language text. Bing helps to find information about a text fragment and its author, but the search results may be inaccurate or incomplete. The main purpose of the paper is to study the effectiveness of establishing the authorship of literary texts by state-of-the-art artificial intelligence tools based on ultra-short excerpts. **Methodology.** Ten Ukrainian authors with a rich body of fiction reflecting various aspects of Ukrainian culture and history were selected, as well as random fragments of 3–7 words each from different works of these authors. An experiment was conducted to determine the authorship of 2,000 fragments. **Findings.** Using the Python programming language and the skpy package, we developed software that sends questions and receives answers from the Bing bot built into Microsoft Skype. The answers were checked for the name of the author of the phrase and the corresponding title of the work. According to the results, Ivan Franko has the highest percentage of answers where the author's name was mentioned (65%), and Oleksandr Dovzhenko has the lowest result (23%). The answers were analyzed by the length of the fragments. Of course, the longer the length of a text fragment, the greater the likeli-



## ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ ТА МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

hood of accurately identifying its authorship. Features of the author's style are manifested in 20–40 % of short fragments. The remaining 60–80% may be commonly used language constructions that the author relayed from the external environment. **Originality.** In this work, for the first time, the method of checking the authorship of fragments of Ukrainian-language text using the Bing bot with artificial intelligence is presented. A comparative analysis was performed and experiments were given to determine the authorship of short fragments of 3–7 words. It has been established that even quite small fragments of the text have signs characteristic of the original style of the author of artistic works. **Practical value.** It has been determined to what extent experts in determining the authorship of natural language texts can rely on existing state-of-the-art artificial intelligence tools in combination with an extensive database of texts in the Internet space.

**Keywords:** authorship detection; natural language text; artificial intelligence; generative language models; ChatGPT; Bing bot; Skype; Microsoft; Bard; Google

## REFERENCES

1. An unofficial Python library for interacting with the Skype HTTP API. (2023). *SkPy 0.10.6*. Retrieved from <https://pypi.org/project/SkPy/> (in English)
2. Bengio, Y. (2008). Neural net language models. *Scholarpedia*, 3(1), 3881. DOI: <https://doi.org/10.4249/scholarpedia.3881> (in English)
3. Bonifacic, I. (2023). *Microsoft's next-gen Bing uses a 'much more powerful' language model than ChatGPT*. Retrieved from <https://www.engadget.com/microsofts-next-gen-bing-more-powerful-language-model-than-chatgpt-182647588.html> (in English)
4. Chowdhery, A., Narang, Sh., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Fiedel, N. (2022). PaLM: Scaling Language Modeling with Pathways. *arXiv*, 1-87, DOI: <https://doi.org/10.48550/arXiv.2204.02311> (in English)
5. *Confirmed: the new Bing runs on OpenAI's GPT-4*. (2023). Retrieved from [https://blogs.bing.com/search/march\\_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4/](https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4/) (in English)
6. Hoffmangoogole, Ch. (2023). Google Bard: How to Use Google's AI Chatbot. *How-To Geek*. Retrieved from <https://www.howtogeek.com/880668/google-bard-how-to-use-googles-ai-chatbot/> (in English)
7. Montti, R. (2023). How Bing AI Search Uses Website Content. *Search Engine Journal*. Retrieved from <https://www.searchenginejournal.com/how-bing-ai-search-uses-web-content/480643/> (in English)
8. Perez, E., Kiela, D., & Cho, K. (2021). True few-shot learning with language models. *Advances in Neural Information Processing Systems*, 34. (in English)
9. *Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web...* Retrieved from <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/> (in English)
10. Shynkarenko, V. I., & Demidovich, I. M. (2022). Natural Language Texts Authorship Establishing Based on the Sentences Structure. In *COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems* (pp. 328-337). Gliwice, Poland. (in English)
11. Shynkarenko, V. I., & Demidovich, I. M. (April, 2021). Authorship Determination of Natural Language Texts by Several Classes of Indicators with Customizable Weights. In *COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems* (Vol. 1, pp. 832-844). Kharkiv, Ukraine. (in English)

Надійшла до редколегії: 01.03.2023

Прийнята до друку: 29.06.2023