УДК 004.8+004.85-028.24“653”]:930.2 **Том 39. 2026 р.**

ДОСВІД ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В РОБОТІ З СЕРЕДНЬОВІЧНИМИ МАНУСКРИПТАМИ

Максим ВОЛОЩУК
Товариство з обмеженою відповідальністю ВІНСТАРС ТЕХНОЛОДЖІ,
e-mail: maxvoloshchuk1@gmail.com
DOI: 10.15330/gal.39.171-179, ORCID: 0009-0005-4950-6234,
ISSN 2312-1165 (друкована версія), ISSN 3083-6670 (онлайн версія)

Богдана ЗАРЕМБОВСЬКА
Навчально-науковий центр дистанційного навчання та моніторингу освітньої діяльності
Карпатського національного університету імені Василя Стефаника,
вул. Шевченка, 57, 76018, м. Івано-Франківськ, Україна
e-mail: zarembovskabohdana@gmail.com
DOI: 10.15330/gal.39.171-179, ORCID: 0009-0002-7673-5970
ISSN 2312-1165 (друкована версія), ISSN 3083-6670 (онлайн версія)

У цій роботі наведено принципи та підходи застосування штучного інтелекту (ШІ) у роботі з рукописними історичними документами. З розвитком методів машинного навчання та комп’ютерного зору використання автоматизованих систем для аналізу, структурування та розпізнавання текстів зі сканованих документів набуває все більшого поширення, що підтверджується значною кількістю сучасних наукових досліджень у цій галузі. Використання подібних механізмів стає стандартом для великих архівних проєктів. Однак, попри значну кількість наявних інструментів, більшість з вже готових рішень орієнтовані переважно на обробку документів доби модерну, тоді як проблематика автоматизованої обробки середньовічних манускриптів залишається недостатньо дослідженою через варіативність письма та фізичні пошкодження матеріалів.

У статті ми проаналізували сучасні підходи застосування машинного навчання у роботі з розпізнаванням рукописних історичних текстів, зокрема методи детекції та сегментації структурних елементів документів, а також запропонували власне рішення комп’ютерної системи, здатне обробляти латино-мовні документи епохи Каролінгів та Оттонів IX–XI ст. Особлива складність роботи з документами цього періоду зумовлена специфікою каролінзького мінускула: наявність численних лігатур, виносних елементів та специфічних середньовічних скорочень – створює серйозні перешкоди для стандартних алгоритмів OCR. Водночас документи каролінзької доби мають високу джерельну цінність для історичних і палеографічних досліджень, оскільки фіксують ранні форми адміністративної, правової та писемної практики Західної Європи, а також відображають етапи становлення середньовічної документальної традиції.

Запропонована у цій роботі нами система є модульною та складається з чотирьох взаємопов’язаних моделей машинного навчання, кожна з яких виконує власну роль в загальному процесі обробки документа. Система забезпечує поетапне виявлення текстових рядків і слів, а також їх розпізнавання із застосуванням поєднання різних моделей для пошуку візуально й синтаксично подібних слів. Такий підхід дозволяє підвищити стійкість розпізнавання в умовах обмеженого навчального набору даних і забезпечує кращу адаптацію до особливостей середньовічного письма.

***Ключові слова**: палеографія, штучний інтелект, комп’ютерний зір, розпізнавання рукописного тексту, машинне навчання.*

З раптовою популяризацією великих мовних моделей (*large language models*) у 2022 р., використання ШІ стає все більш поширеним явищем. Крім генерації тексту або зображень, моделі машинного навчання здатні обробляти не лише структуровані дані, а й інші типи інформації – зокрема дані зі сканованих документів, виконуючи їх розпізнавання.

ШІ – досить широке поняття, що має низку визначень, залежно від джерела:

- «Штучний інтелект стосується комп'ютерних систем, які можуть виконувати складні завдання, які зазвичай виконуються людським мисленням, прийняттям рішень, створенням тощо», як вважають при NASA[1].

- «Штучний інтелект (ШІ) – це технологія, яка дозволяє комп'ютерам та машинам імітувати людське навчання, розуміння, вирішення проблем, прийняття рішень, креативність та автономію», на думку фахівців із IBM[2].

- «Штучний інтелект – здатність цифрового комп'ютера або робота, керованого комп'ютером, виконувати завдання, які зазвичай асоціюються з розумними істотами», за словами упорядників «*Енциклопедії Британіки*»[3].

Узагальнюючи, ШІ – це комп'ютерна система, що імітує людські когнітивні здібності у виконанні завдань. Системами ШІ можуть вважатися як комплексні нейронні мережі, здатні узагальнювати великі обсяги тексту, так і прості програми, що, наприклад, грають у гру «*Хрестики-нулики*» згідно з наперед визначеною стратегією. Одним із підходів розробки ШІ є створення моделей машинного навчання. Принцип цих програм полягає у тому, що система здатна змінювати свою поведінку завдяки обробці наборів навчальних даних. Одним із найпоширеніших способів застосування машинного навчання є штучні нейронні мережі – системи, дизайн яких був натхненний нейронними мережами людського мозку. Саме ці системи є найбільш ефективними в багатьох сферах обробки даних – наприклад, комп'ютерний зір, обробка природної мови, навчання з підкріпленням та генеративний ШІ.

Навчання штучних нейронних мереж відбувається завдяки обробці великого набору анотованих даних, що називається навчальною вибіркою. Алгоритм навчання намагається встановити такі параметри системи, щоб похибка її роботи була якомога менша.

Щоб натренувати штучну нейронну мережу для розпізнавання історичних документів, дослідники та розробники можуть використовувати публічні набори анотованих даних або провести анотацію власноруч. Одним із прикладів таких наборів даних є DIVA-HisDB[4]. Цей датасет налічує 150 анотованих сторінок середньовічних манускриптів. Анотації стосуються задач аналізу структури документів та виявлення текстових елементів – наприклад, рядків. Вартим уваги також є CATMuS[5]. Цей набір даних налічує зображення рядків та анотації понад 200 манускриптів періоду VIII–XVI ст. Датасет налічує екземпляри 10-ма мовами, зокрема латинською. Цей набір даних слугує чудовою базою для навчання моделі розпізнавання тексту на зображеннях текстових рядків.

Однією з причин, чому більшість досліджень присвячені обробці манускриптів модерної доби, є наявність більшої кількості даних. Достатня кількість тренувальних та тестувальних прикладів є фундаментальною потребою при навчанні систем машинного навчання. На щастя, існують онлайн-громади, які поповнюють набори середньовічних даних новими анотаціями, що дозволяє проводити результативні дослідження.

Однією з особливостей середньовічних джерел є значно гірший стан їх збереження порівняно з документами епохи модерну. Найчастішою проблемою є вицвітання чорнила, що

---

[1] «Artificial intelligence refers to computer systems that can perform complex tasks normally done by human-reasoning, decision making, creating, etc.», NASA. What is Artificial Intelligence? URL: https://www.nasa.gov/what-is-artificial-intelligence.

[2] «Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy», Stryker C., Kavlakoglu E. What is AI? URL: https://www.ibm.com/think/topics/artificial-intelligence.

[3] «Artificial intelligence – the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings», Copeland B. J. *Artificial intelligence* URL: https://www.britannica.com/technology/artificial-intelligence.

[4] Simistira F., Seuret M., Eichenberger N., Garz A., Liwicki M., and Ingold R. DIVA-HisDB: A Precisely Annotated Large Dataset of Challenging Medieval Manuscripts. *International Conference on Frontiers in Handwriting Recognition*, 2016. P. 471–476.

[5] Clérice Th., Pinche A. Vlachou-Efstathiou M., Chagué A., Camps J.–B., et al. *CATMuS Medieval: A multilingual large-scale cross-century dataset in Latin script for handwritten text recognition and beyond*. 2024. DOI: hal-04453952.

дуже істотно ускладнює процес обробки. Також часто можна зустріти втрачені фрагменти пергаментів, плями та інші типи пошкоджень.

Як правило, процес обробки середньовічних документів складається з двох кроків (див. рис. 1):

1. Виявлення необхідних елементів документа (наприклад, рядки, текстові блоки тощо).
2. Розпізнавання тексту у виявлених об'єктах (найчастіше зображеннях рядків).

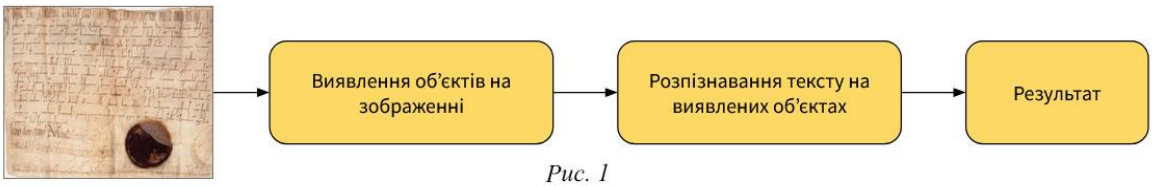

Рис. 1

Для виявлення об'єктів можуть використовуватися моделі детекції або сегментації. Моделі детекції здатні визначати координати бажаних об'єктів у форматі обмежувальної рамки – прямокутника, що описує межі елемента документа. Ці моделі здатні класифікувати різні об'єкти, наприклад, відрізняти виявлені зображення від заголовків.

Перед побудовою моделі ми виконали детальний аналіз зібраного набору даних, щоб визначити оптимальний підхід – як до детекції, так і до класифікації. Зокрема, було проаналізовано кількість зразків кожного класу, що дало змогу оцінити ступінь дисбалансу та визначити, як доцільніше навчати моделі. Додатково ми дослідили близькість слів за відстанню Геммінга, щоб зрозуміти, наскільки візуально подібні між собою класи присутні в датасеті. Також було виконано аналіз самих зображень документів: наявність шумів, артефактів сканування, нерівномірне освітлення та перекоси тексту, що безпосередньо впливає на точність детекції. Окремо розглянуто розподіл частоти вживання слів, кількість унікальних слів у кожному файлі та частку рідковживаних слів, що допомогло оцінити загальну складність задачі та визначити необхідні техніки обробки зображень і балансування. Результати цього аналізу сформували підґрунтя для вибору архітектури моделей і стратегії підготовки даних[6].

Серхіо Торрес Агіляр (*Sergio Torres Aguilar*) порівняв різні моделі детекції об'єктів на базі архітектур YOLO та трансформерів у задачі структурного аналізу документів та виявлення об'єктів[7], використовуючи набір із трьох публічних датасетів. Дослідник виявив, що використання орієнтованих обмежувальних рамок є простим, однак фундаментальним покращенням. Ми дійшли того ж висновку, навчаючи модель виявлення рядків.

Натомість, моделі семантичної сегментації утворюють «маску» зображення, що позначає які пікселі належать до виявленого об'єкта, а які ні. Ці моделі виділяють силует виявлених елементів, що часто є точнішим підходом, ніж обведення прямокутником. Так само як і моделі детекції, моделі сегментації здатні класифікувати різні виявлені об'єкти.

Мелоді Буайє (*M'elodie Boillet*), Крістофер Керморван (*Christopher Kermorvant*) і Тьєррі Паке (*Thierry Paquet*) провели дослідження та розробили систему сегментації рядків, здатну розпізнавати велике різноманіття шрифтів та стилів[8]. Їхня робота показує, наскільки добре ШІ здатен узагальнювати інформацію та адаптуватися до нових даних при правильному підході.

Після успішного виявлення елементів документу системі необхідно з'ясувати що написано на виявлених об'єктах. У сучасних застосунках використовуються, як правило, два підходи:

---

[6] Волощук М., Зарембовська Б. Використання штучного інтелекту (машинного навчання) для розчитки середньовічних історичних документів. *Студентські історичні зошити.* 2024. Вип. 16. С. 116.

[7] Aguilar S. From Codicology to Code: A Comparative Study of Transformer and YOLO-based Detectors for Layout Analysis in Historical Documents. 2025. DOI: 10.48550/arXiv.2506.20326.

[8] Boillet M., Kermorvant Ch., Paquet Th. Robust text line detection in historical documents: learning and evaluation methods. *International Journal on Document Analysis and Recognition* (IJDAR). 2022. Vol. 25. P. 1–20. DOI: 10.1007/s10032-022-00395-7.

1. *Optical character recognition* (OCR) – підхід, що визначає значення кожного символу окремо, завдяки чому збирає повний текст написаного на зображенні.

2. *Handwritten text recognition* (HTR) – область обробки рукописного тексту, з усіма його особливостями, такими як почерк, стиль письма тощо. Як правило, для кожного домену використовують унікальні штучні нейронні мережі.

Системи OCR часто погано обробляють рукописний текст, тим паче на історичних матеріалах, особливо середньовічних. Як правило, для цих задач використовуються штучні нейронні мережі, які приймають на вхід зображення текстових рядків та повертають розпізнаний текст. Певною мірою, архітектурно ці нейронні мережі нагадують моделі розпізнавання мовлення з аудіосигналів, однак на вхід приймають зображення тексту, а не звукову інформацію.

Серхіо Торрес Агіляр та Вінсент Жоліве (*Vincent Jolivet*) у своєму дослідженні змогли створити систему розпізнавання тексту на текстових рядках, враховуючи дані з різних джерел[9].

На цей час нам відомо про кілька онлайн-сервісів обробки історичних документів. Одним із них є *Transcribus* (https://www.transkribus.org/) – комерційний проєкт для обробки рукописних матеріалів. Підтримує низку мов та типів документів. Однак наразі можна сказати, що галузь радше на етапі досліджень, адже готових онлайн-рішень для транскрипції манускриптів не так багато.

Справа в тому, що латинські рукописні тексти раннього середньовіччя становлять особливу складність для автоматичного розпізнавання: нестандартизоване письмо, варіативні орфографічні форми, лігатури, своєрідні з'єднання букв та скорочення. Звісно ж, на ефективність роботи впливають також і механічні пошкодження матеріалу. Більшість готових OCR та HTR систем у таких умовах демонструють дуже низьку точність, оскільки вони розраховані на сучасні шрифти або на добре структуровані документи без шумів і варіативності. Крім того, середньовічна писемність часто мала систему скорочень слів, що значно ускладнює роботу OCR та HTR-системам, які, як правило, розпізнають текст літера за літерою.

На цьому тлі розроблена нами система має принципову відмінність. Наш механізм для розчитки історичних латинських текстів побудований не як стандартна система для розпізнавання текстів, а як поетапна модель, оптимізована під рукописи епохи Каролінгів (див. рис. 2).

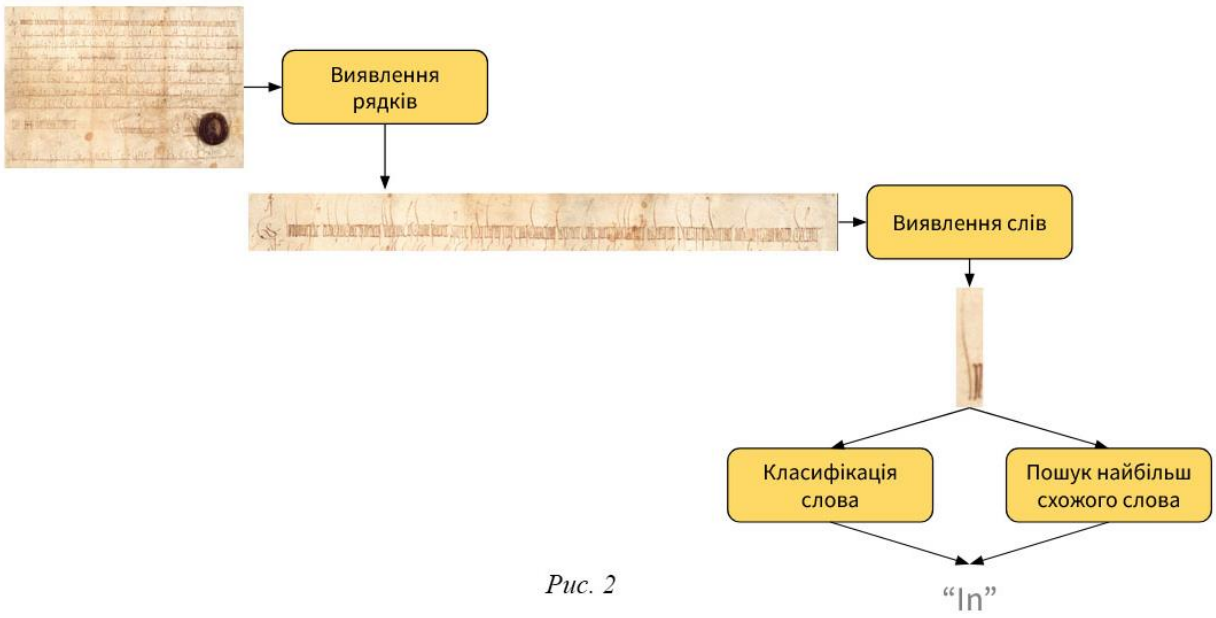

Рис. 2

[9] Aguilar S., Jolivet V. Handwritten Text Recognition for Documentary Medieval Manuscripts. 2022. DOI: hal-03892163v1.

Ми розділили підхід транскрипції документів на кілька етапів:

1. виявлення рядків;
2. виявлення слів всередині рядків;
3. розпізнавання слів з допомогою класифікації;
4. розпізнавання з допомогою пошуку синтаксично схожих слів.

Провідний фахівець Центру медієвістичних студій, доктор філософії з історії та археології Остап Кардаш надав нам зображення та транскрипції 31-го документа епохи Каролінгів та Оттонів. На основі цих даних було проведено наше дослідження. Для їх використання у навчанні моделей ШІ нам необхідно було провести анотацію. Для цього ми використали онлайн-інструменти, що дозволяють обводити координати об'єктів на зображеннях та експортувати ці дані в потрібний нам формат. Крім того, довелося провести окремий процес формування навчальних та тестувальних вибірок строго під вимоги конкретних моделей ШІ.

Отримавши анотації, ми мали можливість проаналізувати їх для кращого розуміння даних. Одним із основних напрямів аналізу стало визначення розподілу частоти слів. Природно для людської мови мати слова, які повторюються частіше, ніж інші, та слова які вживаються вкрай рідко. Ця властивість підтвердилася і для наших даних. Наприклад, сполучники та прийменники, такі як «et», «nostri» тощо, траплялися найчастіше. Усі надані документи мали схожу структуру та починалися зі слів «In nomine Sanctae Trinitatis…», тому логічно, що ці слова також траплялися частіше за інші. Цей аналіз показує незбалансованість даних, що є досить важливим фактором, адже ШІ моделі за своєю сутністю є статистичними і якість їх навчання напряму залежить від властивостей тренувальної вибірки.

Першим етапом обробки документа в нашій системі є розпізнавання рядків. Для цього ми навчили YOLO (*You Only Look Once*) модель, використовуючи наші анотації. Попри відносно невелику кількість даних, модель показала досить хороші метрики на тестових прикладах та точно виділяла координати рядків. Для збільшення точності її роботи, був використаний формат орієнтованих рамок, що, крім координат, містить і кут нахилу об'єкта. Це важливо, адже рядки, написані чітко «під лінійку», трапляються дуже рідко. Виявлення рядків також допомагає зручніше впорядкувати надалі виявлені слова.

Для виявлення слів застосовувалася також модель YOLO архітектури, однак з використанням звичайних обмежувальних рамок, без інформації про нахил. Модель навчалася виявляти слова на зображеннях рядків, що є простішою задачею, ніж виявлення на цілому документі.

Після визначення координат слів система переходить до наступного етапу – розпізнавання.

Основний компонент розпізнавання – це класифікатор, який працює швидко й точно у випадках, коли слово вже зустрічалось у навчальному наборі достатньо часто. Другий компонент – ембедер (англ. *embedder*), механізм, який знаходить схожість між зображеннями.

Система працює таким чином, що класифікатор намагається одразу назвати слово, базуючись на тому, до якого слова (класу) воно найбільш імовірно відноситься, але якщо він не впевнений, тобто ймовірність відношення до всіх знайомих класів надто низька, або слово нетипове (рідкісна форма, новий варіант написання, пошкоджене зображення), спрацьовує другий компонент – ембедер. Він порівнює потрібне для нас слово із зображеннями інших слів і на основі цього підбирає найбільш близького за синтаксисом і візуальними ознаками кандидата.

Класифікатор дозволяє працювати швидко й точно на знайомих словах, за рахунок своєї алгоритмічної простоти, а друга частина вберігає архітектуру від помилок, що виникають при роботі класифікатора. Саме тому наш підхід до створення проєкту чудово підходить для роботи зі специфічними даними: система не лише вміє якісно впізнавати знайоме, але й добре справляється з невідомим, мінімізуючи похибку там, де стандартні OCR-системи часто припускаються помилок.

Після обробки документа системою користувач отримує інформацію про виявлені об'єкти та розпізнані слова (див. рис. 3, 4, 5).

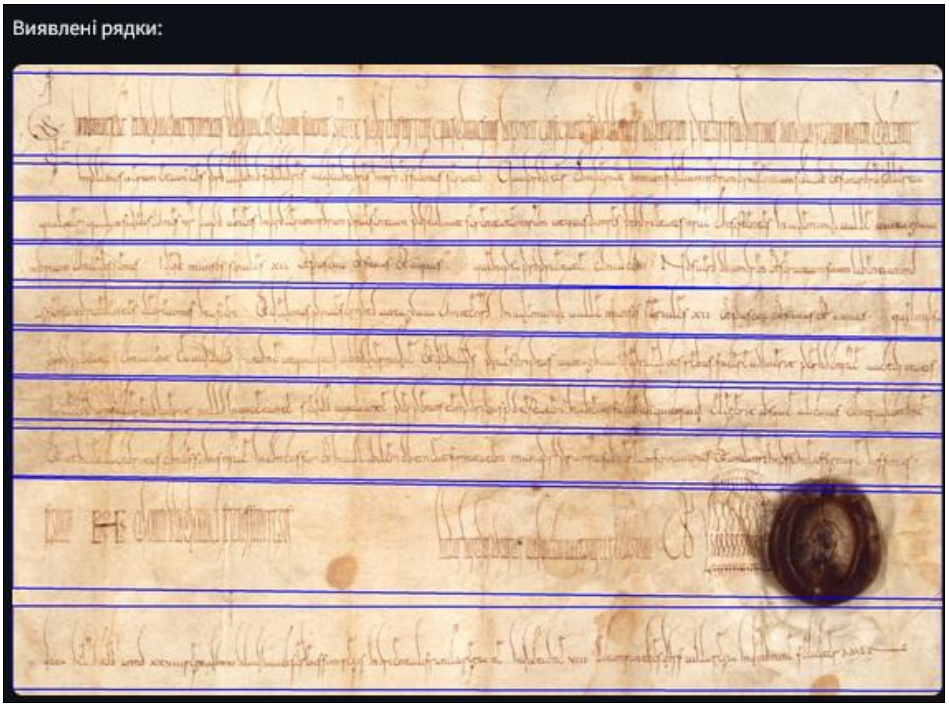

Рис. 3

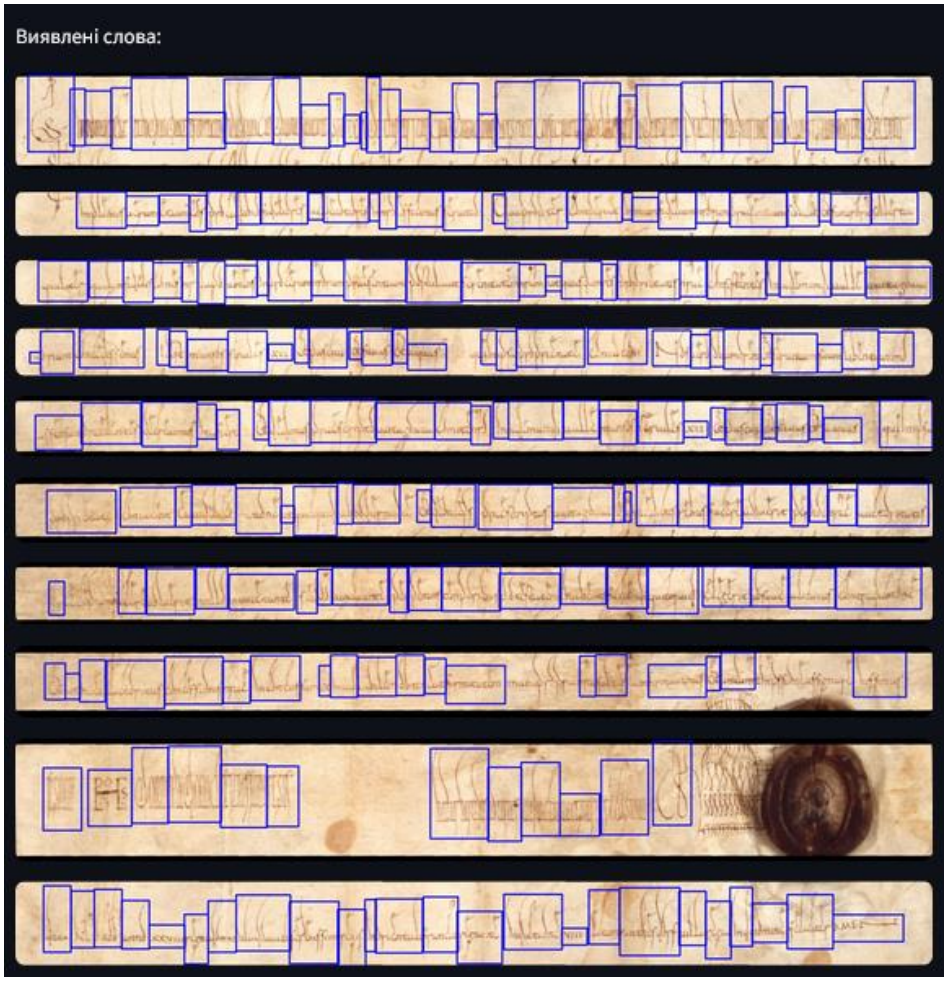

Рис. 4

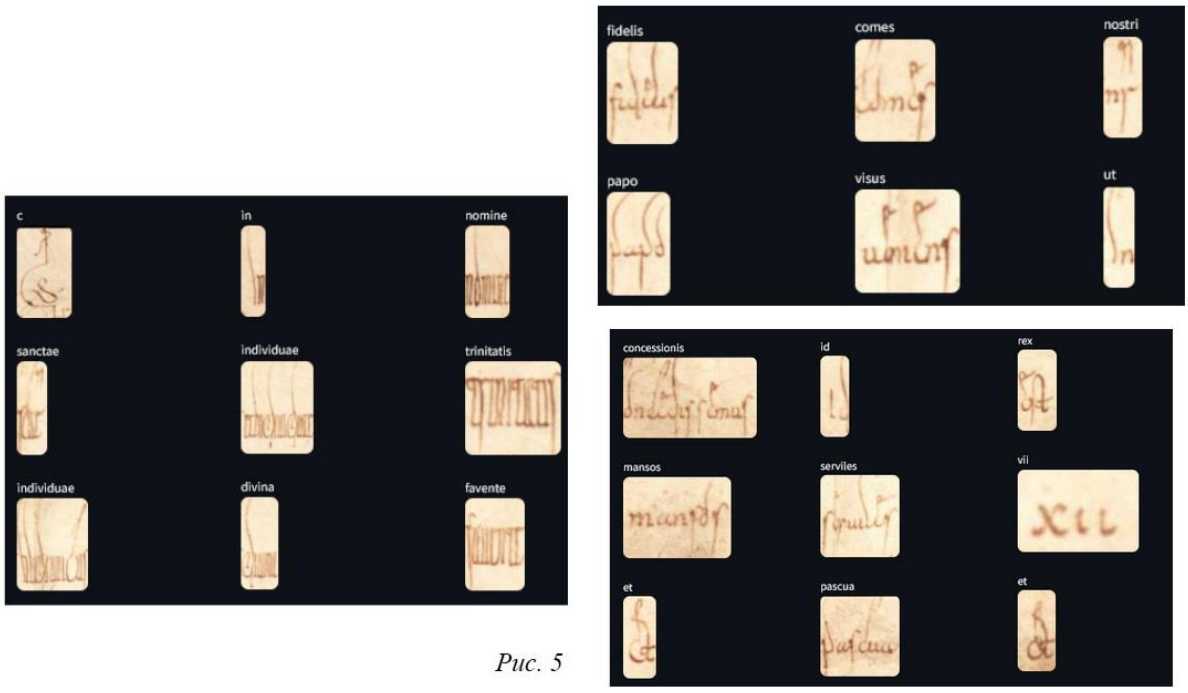

Рис. 5

Така гібридна стратегія дозволяє уникнути типової проблеми звичайних OCR-систем: помилкового «вгадування» незнайомого для моделі слова, що суттєво підвищує стабільність і якість на реальних рукописах. Завдяки поєднанню швидкого класифікатора та більш обережного ембедера, система зберігає точність навіть на рідкісних словах, пошкоджених фрагментах або нестандартних варіантах письма. Такий підхід забезпечує надійний баланс між швидкістю обробки та якістю результату, роблячи технологію ефективним інструментом для транскрипції та аналізу історичних документів.

Наша система краще адаптована до можливих скорочень у тексті, оскільки сприймає зображення слів у цілому, а не намагається зібрати їх по літерах. Також можна з упевненістю сказати, що наша система добре працює з недосконалими матеріалами, адже пошкоджені частини документів не потрапляють на розпізнавання. Крім того, результати роботи системи є легко зрозумілими, адже якщо програма допускається помилки, користувач здатен побачити, де саме та яке саме слово було розпізнано неправильно, та чи було воно виявлене.

Однак, звичайно, система має кілька обмежень. Оскільки розпізнавання тексту відбувається слово за словом, важливо мати набір даних із достатньою вибіркою різних слів. Це вимагає значно більшої кількості роботи для анотації даних. Також обмеженням є те, що система повертає не текст, а радше впорядкований набір слів.

Це обмеження і є основним фокусом нашої майбутньої роботи – перетворення набору слів у повноцінний текст. У наших планах є використання нових архітектур ШІ, наприклад, рекурентних мереж та трансформерів, для побудови грамотно оформленого тексту. Крім того, ми зосереджені на оптимізації наявних підходів для пришвидшення роботи системи та покращення поточних результатів.

EXPERIENCE OF USING ARTIFICIAL INTELLIGENCE IN WORKING WITH MEDIEVAL MANUSCRIPTS

Maksym VOLOSHCHUK
Limited Liability Company Winstars AI
e-mail: maxvoloshchuk1@gmail.com
ORCID: 0009-0005-4950-6234

Bohdana ZAREMBOVSKA
Educational and Scientific Center for Distance Learning and Monitoring of Educational Activities of the Vasyl Stefanyk Carpathian National University,
57 Shevchenko st., 76018, Ivano-Frankivsk, Ukraine
e-mail: zarembovskabohdana@gmail.com
ORCID: 0009-0002-7673-5970

The article presents the principles and approaches to the application of artificial intelligence (AI) in the work with handwritten historical documents. With the development of machine learning methods and computer vision, the use of automated systems for the analysis, structuring, and recognition of texts from scanned documents has become increasingly widespread, as evidenced by a substantial body of contemporary scholarly research in this field. The use of such mechanisms is becoming a standard practice in large-scale archival projects. However, despite the significant number of available tools, most existing solutions are primarily oriented toward the processing of documents from the modern period, while the problem of automated processing of medieval manuscripts remains insufficiently studied due to the variability of handwriting and the physical deterioration of the materials.

In this article, we analyze current approaches to the application of machine learning in the recognition of handwritten historical texts, in particular methods for the detection and segmentation of structural elements of documents. We also propose our own computer system capable of processing Latin-language documents from the Carolingian and Ottonian periods of the 9th–11th centuries. The particular complexity of working with documents from this period is due to the specifics of Carolingian minuscule, including the presence of numerous ligatures, ascenders and descenders, and characteristic medieval abbreviations, which pose serious challenges for standard OCR algorithms. At the same time, documents of the Carolingian period possess high source value for historical and palaeographic research, as they record early forms of administrative, legal, and written practices in Western Europe and reflect key stages in the formation of the medieval documentary tradition.

The system proposed in this paper is modular and consists of four interconnected machine learning models, each performing a specific role in the overall document processing pipeline. The system provides step-by-step detection of text lines and words, as well as their recognition through the combination of different models designed to identify visually and syntactically similar words. This approach improves the robustness of recognition under conditions of limited training data and ensures better adaptation to the specific characteristics of medieval handwriting.

***Key words**: paleography, artificial intelligence, computer vision, handwritten text recognition, machine learning.*

REFERENCES

Aguilar S. (2025). From Codicology to Code: A Comparative Study of Transformer and YOLO-based Detectors for Layout Analysis in Historical Documents. DOI: 10.48550/arXiv.2506.20326. (in English).

Copeland B. J. Artificial intelligence. URL: https://www.britannica.com/technology/artificial-intelligence (in English).

Boillet M., Kermorvant Ch., Paquet Th. (2022). Robust text line detection in historical documents: learning and evaluation methods. *International Journal on Document Analysis and Recognition* (IJDAR). Vol. 25. P. 1–20. DOI: 10.1007/s10032-022-00395-7 (in English).

Stryker C., Kavlakoglu E. What is AI? DOI: https://www.ibm.com/think/topics/artificial-intelligence. (in English).

Simistira F., Seuret M., Eichenberger N., Garz A., Liwicki M., and Ingold R. (2016). DIVA-HisDB: A Precisely Annotated Large Dataset of Challenging Medieval Manuscripts. *International Conference on Frontiers in Handwriting Recognition*. P. 471–476 (in English).

NASA. What is Artificial Intelligence? DOI: https://www.nasa.gov/what-is-artificial-intelligence (in English).

Aguilar S. T., Jolivet V. Handwritten Text Recognition for Documentary Medieval Manuscripts. 2022. DOI: hal-03892163v1. (in English).

Clérice Th., Pinche A., Vlachou-Efstathiou M., Chagué A., Camps J.-B. et al. (2024). CATMuS Medieval: A multilingual large-scale cross-century dataset in Latin script for handwritten text recognition and beyond. DOI: hal-04453952 (in English).

Voloshchuk, M., Zarembovska, B. (2024). Vykorystannia shtuchnoho intelektu (mashynnoho navchannia) dlia rozchytky serednovichnykh istorychnykh dokumentiv. *Students'ki istorychni zoshyty*. Vol. 16. S. 116–125 (in Ukrainian).

Надійшла до видавництва 10 листопада 2025 р.
Прийнята до друку 28 січня 2026 р.
Опублікована 6 травня 2026 р.