

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Інститут інформаційних технологій в економіці**

**Кафедра штучного інтелекту, моделювання та статистики**

**Освітньо-професійна програма «Економічна кібернетика і Дата Сайнс»**

Галузь знань 05 «Соціальні та поведінкові науки»

Спеціальність 051 «Економіка»

Форма здобуття освіти: очна (денна)

**КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА**

на тему **«NLP-аналіз впливу штучного інтелекту на структуру ринку праці»**

*(назва теми)*

здобувача Замлинної Яни Володимирівни

*(ПІБ, підпис)*

Науковий керівник: д. ф.-м. н., професор Ольга ПРИТОМАНОВА

*(підпис)*

**Робота допущена до захисту перед екзаменаційною комісією з атестації здобувачів вищої освіти (ЕК)**

Завідувач кафедри: к. ф.-м. н., професор

Галина ВЕЛИКОІВАНЕНКО

*(підпис)*

**Київ 2025**

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВАДИМА ГЕТЬМАНА

Інститут інформаційних технологій в економіці

Кафедра штучного інтелекту, моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика і Дата Сайнс»

Галузь знань 05 Соціальні та поведінкові науки

Спеціальність 051 Економіка

**ПОГОДЖЕНО**

Керівник проектної групи (гарант)  
освітньо-професійної програми  
Олена ПСКУНОВА

**ЗАТВЕРДЖУЮ**

Завідувач кафедри штучного інтелекту,  
моделювання та статистики  
Галина ВЕЛИКОІВАНЕНКО

\_\_\_\_\_ (підпис)

\_\_\_\_\_ 2025 р.

\_\_\_\_\_ (підпис)

\_\_\_\_\_ 2025 р.

**ІНДИВІДУАЛЬНЕ ЗАВДАННЯ**

здобувача вищої освіти

Замлинної Яни Володимирівни

(прізвище, ім'я, по батькові)

\_\_\_\_\_ очної (денної)

\_\_\_\_\_ форми здобуття освіти

на підготовку кваліфікаційної магістерської роботи

*на тему*

«NLP-аналіз впливу штучного інтелекту на структуру ринку праці»

Тему затверджено наказом ректора Університету від «\_\_» \_\_\_\_\_ 20\_\_ р. № \_\_\_\_\_

Кваліфікаційна магістерська робота виконується на матеріалах офіційних сайтів

Міністерство економіки, довілля та сільського господарства України <https://me.gov.ua/>, Реєстру кваліфікацій <https://register.nqa.gov.ua/> та інших відкритих джерел.

## План кваліфікаційної магістерської роботи

<b>Розділ 1</b>	Теоретичні та методичні основи аналізу трансформацій ринку праці під впливом цифровізації та автоматизації
<b>Розділ 2</b>	Моделювання процесів трансформації на ринку праці із застосуванням методів NLP
<b>Розділ 3</b>	Оцінювання гендерних диспропорцій у оплаті праці із застосуванням Logib моделі
<b>Об'єкт дослідження:</b>	процеси трансформації структури ринку праці під впливом штучного інтелекту
<b>Предмет дослідження:</b>	методи і моделі штучного інтелекту, пов'язані з процесами трансформації структури ринку праці
<b>Мета виконання кваліфікаційної магістерської роботи:</b>	застосування методів та інструментів NLP та Big Data, що дозволяють моделювати процеси трансформації структури ринку праці під впливом штучного інтелекту та оцінювати можливі ризики втрати робочого місця через цифровізацію та автоматизацію та гендерних диспропорцій в платі праці

**Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:**

**У розділі 1:**

- 1) огляд джерел для визначення сутності та функцій ринку праці;
- 2) огляд джерел щодо впливу штучного інтелекту на структуру сучасного ринку праці;
- 3) огляд методів і моделей штучного інтелекту дослідження процесів трансформації структури ринку праці.

**У розділі 2:**

- 1) характеристика інформаційної бази для моделювання;
- 2) застосування методів NLP для моделювання процесів трансформації структури ринку праці (розглянути конкретні методи).

**У розділі 3:**

- 1) сформулювати алгоритм оцінювання гендерної нерівності в оплаті праці на основі моделі Logib;
- 2) застосувати розроблений алгоритм для оцінювання гендерних диспропорцій у оплаті праці для організацій з великою та малою кількістю працівників.

Завдання підготував  
науковий керівник

\_\_\_\_\_

(підпис)

Ольга ПРИТОМАНОВА

(ініціали, прізвище)

«\_\_» \_\_\_\_\_ 2025 р.

Завдання одержав  
здобувач

\_\_\_\_\_

(підпис)

Яна ЗАМЛИННА

(ініціали, прізвище)

«\_\_» \_\_\_\_\_ 2025 р.

## Реферат

Кваліфікаційна магістерська робота містить 75 сторінок, 10 таблиць, 23 рисунки, список використаних джерел з 68 найменувань, додатки.

### «NLP-аналіз впливу штучного інтелекту на структуру ринку праці»

*Об'єктом дослідження* кваліфікаційної магістерської роботи є процеси трансформації структури ринку праці під впливом штучного інтелекту.

*Предметом дослідження* є методи і моделі штучного інтелекту, пов'язані з процесами трансформації структури ринку праці.

*Мета і завдання дослідження.* Метою виконання кваліфікаційної магістерської роботи є застосування методів та інструментів NLP та Big Data, що дозволяють моделювати процеси трансформації структури ринку праці під впливом штучного інтелекту та оцінювати можливі ризики втрати робочого місця через цифровізацію та автоматизацію та гендерних диспропорцій в платі праці.

Для досягнення поставленої мети визначені такі *завдання*:

- Проаналізувати сутність та ключові функції ринку праці, дослідити вплив технологій штучного інтелекту на трансформацію його сучасної структури.
- Проаналізувати методи й моделі штучного інтелекту, що застосовуються для дослідження процесів структурних змін на ринку праці.
- Здійснити підготовку даних та побудувати семантичне представлення текстової інформації на основі сучасних векторних моделей (SBERT).
- Сформувати модель оцінювання ризику автоматизації професій із використанням SBERT-ембедінгів та релевантних алгоритмів машинного навчання.
- Реалізувати метод крос-лінгвального вирівнювання професій та виконати структурний аналіз відповідності між національними та міжнародними класифікаціями.
- Розробити та застосувати алгоритм оцінювання гендерної нерівності в оплаті праці на основі моделі Logit.
- Здійснити оцінювання гендерних диспропорцій в оплаті праці для закладу освіти, розташованого у сільській місцевості та місті.

*Теоретична, методична та практична значущість отриманих результатів.* У процесі дослідження систематизовано теоретичні підходи до аналізу ринку праці в умовах цифровізації та впровадження штучного інтелекту, узагальнено наукові погляди щодо впливу автоматизації на структуру зайнятості та трансформацію професій. Поглиблено розуміння взаємозв'язку між розвитком інтелектуальних технологій, ризиком автоматизації професій і зміною вимог до компетенцій працівників. Методична значущість роботи полягає у поєднанні економічного аналізу з методами обробки природної мови та машинного навчання.

Практична значущість результатів полягає в ідентифікації професій з різним рівнем вразливості автоматизації на ринку праці України, оцінюванні впливу штучного інтелекту на структуру зайнятості та виявленні гендерних диспропорцій в оплаті праці. Отримані результати можуть бути використані органами державної влади, роботодавцями та освітніми установами для формування політики зайнятості, розроблення програм перекваліфікації та підвищення цифрових навичок, а також для моніторингу та зменшення гендерної нерівності на ринку праці.

Рік виконання кваліфікаційної магістерської роботи – 2025.

Рік захисту роботи – 2025.

*Ключові слова:* ринок праці, цифровізація, штучний інтелект, класифікація ринку праці, ризик автоматизації професій, структура зайнятості, професії, гендерні диспропорції в оплаті праці, модель Logit.

В і д г у к  
про кваліфікаційну магістерську роботу здобувача  
освітньо-професійної програми «Економічна кібернетика і Дата Сайнс»  
навчально-наукового інституту «Інститут інформаційних технологій  
в економіці КНЕУ імені Вадима Гетьмана»  
**Замлинної Яни Володимирівни**  
на тему «**NLP-аналіз впливу штучного інтелекту на структуру ринку праці**»

1. Актуальність теми обумовлена тим, що розвиток штучного інтелекту, а особливо великих мовних моделей і інтелектуальних агентів, став одним із найбільш впливових факторів структурних змін сучасного світу. Швидка популяризація алгоритмів штучного інтелекту та їх здатність виконувати певні інтелектуальні завдання, що раніше вважалися винятково людською компетенцією, породжує дискусії про можливу втрату робочих місць. Насправді ризик витіснення людини машиною існує, але ступінь його реальності, масштаби та швидкість поширення потребують науково-обґрунтованого аналізу.

2. Позитивні риси кваліфікаційної магістерської роботи: у роботі розроблена NLP-модель, що дозволяє кількісно оцінювати технологічну вразливість професій і слугує основою для прогнозування майбутніх структурних змін у зайнятості. Водночас і соціальні аспекти ринку праці залишаються критично важливими. У роботі, із застосуванням методів машинного навчання та на основі міжнародного досвіду, проведено аналіз гендерних диспропорцій в оплаті праці в Україні.

Кваліфікаційна робота відповідає затвердженому індивідуальному завданню та оформлена відповідно до вимог до кваліфікаційних робіт.

3. Наявність самостійних розробок автора: Поєднання технологічного та соціального аналізу із застосуванням методів ML та NLP дозволило авторці побудувати модель прогнозування ризику автоматизації професій та зробити висновок, що трансформації, спричинені цифровізацією, не усувають соціальних нерівностей, а лише змінюють їхній характер.

4. Цінність теоретичних висновків та практичних рекомендацій, отриманих у роботі. Розроблені математичні моделі дозволяють проаналізувати трансформації структури ринку праці під впливом штучного інтелекту та оцінювати можливі ризики втрати робочого місця через цифровізацію та гендерних диспропорцій в платі праці, що знайшло практичне втілення у діяльності ГО «Інститут професійних кваліфікацій» при розробленні рекомендацій для органів влади, роботодавців та освітніх закладів щодо заходів перекваліфікації та моніторингу гендерної рівності.

Інститут професійних кваліфікацій надав довідку про застосування у своїй діяльності результатів кваліфікаційної магістерської роботи.

Робота Замлинної Я.В. пройшла апробацію на двох конференціях: III Міжнародній науково-практичній конференції «Цифрова економіка» (5-6 червня 2025 року, Київ) та XIII Всеукраїнській науково-практичній конференції форумі молодих економістів-кібернетиків «Моделювання економіки: проблеми, тенденції, досвід» (22-23 листопада 2025 року, Львів). На обох конференціях Яна виступала з доповідями та за результатами роботи опубліковані тези у збірниках конференцій.

5. Загальна оцінка кваліфікаційної магістерської роботи та її допущення до захисту перед ЕК: кваліфікаційна магістерська робота Замлинної Я.В. виконана на *високому рівні* та може бути допущена до захисту перед ЕК.

Науковий керівник: професор кафедри  
штучного інтелекту, моделювання та статистики,  
доктор фізико-математичних наук, професор \_\_\_\_\_ Ольга ПРИТОМАНОВА

«12» грудня 2025 р.

## ЗМІСТ

<b>ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ</b> .....	<b>8</b>
<b>ВСТУП</b> .....	<b>9</b>
<b>РОЗДІЛ 1. ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ АНАЛІЗУ ТРАНСФОРМАЦІЙ РИНКУ ПРАЦІ ПІД ВПЛИВОМ ЦИФРОВІЗАЦІЇ ТА АВТОМАТИЗАЦІЇ</b> .....	<b>12</b>
1.1 Сутність та функції ринку праці .....	12
1.2 Вплив штучного інтелекту на структуру сучасного ринку праці.....	21
1.3 Методи і моделі штучного інтелекту дослідження процесів трансформації структури ринку праці .....	25
1.3.1 Огляд методів NLP .....	25
1.3.2 Огляд інструментів Big Data .....	33
<b>РОЗДІЛ 2. МОДЕЛЮВАННЯ ПРОЦЕСІВ ТРАНСФОРМАЦІЇ НА РИНКУ ПРАЦІ ІЗ ЗАСТОСУВАННЯМ МЕТОДІВ NLP</b> .....	<b>36</b>
2.1 Підготовка даних та побудова семантичних представлень .....	36
2.2 Моделювання ризику автоматизації на основі SBERT-ембедінгів .....	41
2.3 Крос-лінгвальне вирівнювання професій та структурний аналіз .....	45
<b>РОЗДІЛ 3. ОЦІНЮВАННЯ ГЕНДЕРНИХ ДИСПРОПОРЦІЙ В ОПЛАТІ ПРАЦІ ІЗ ЗАСТОСУВАННЯМ LOGІВ МОДЕЛІ</b> .....	<b>61</b>
3.1 Алгоритм оцінювання гендерної нерівності в оплаті праці на основі моделі Logib.....	61
3.2 Оцінювання гендерних диспропорцій для навчального закладу в сільській місцевості .....	67
3.3 Оцінювання гендерних диспропорцій для навчального закладу в місті.....	72
<b>ВИСНОВКИ</b> .....	<b>77</b>
<b>ВИКОРИСТАНІ ДЖЕРЕЛА</b> .....	<b>80</b>
<b>ДОДАТКИ</b> .....	<b>87</b>
<b>ДОДАТОК А</b> .....	<b>87</b>

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

NLP (Natural Language Processing) – Обробка природної мови

ChatGPT (Generative Pre-trained Transformer) – генеративний попередньо тренований трансформер; чат-бот та віртуальний помічник з генеративним штучним інтелектом

ILO (International Labour Organization) – Міжнародна організація праці

ШІ – штучний інтелект

ЄС – Європейський Союз

МВФ – Міжнародний валютний фонд

OECD – Організація економічного співробітництва та розвитку

BERT – Bidirectional Encoder Representations from Transformers

TF-IDF – term frequency–inverse document frequency

PCA (Principal Component Analysis) – Метод головних компонент

ERP (Enterprise Resource Planning) – Планування ресурсів підприємства

HRIS (Human Resources Information System) – інформаційна система управління персоналом

ETL (Extract, Transform, Load) – (вилучення, перетворення, завантаження) підхід при якому дані спочатку обробляються на проміжному етапі, а потім завантажуються у цільове сховище.

ELT (Extract, Load, Transform) – (вилучення, завантаження, перетворення) підхід при якому не оброблені дані завантажуються в цільову систему, а їх трансформація відбувається безпосередньо всередині сховища.

ISCO (International Standard Classification of Occupations) – Міжнародна стандартна класифікація професій

HR – Human Resources

## ВСТУП

*Актуальність теми.* Ми живемо у часі глобальних зрушень та цифрової трансформації, коли технології розвиваються швидше, ніж суспільні та економічні системи встигають адаптуватися. Розвиток штучного інтелекту, а особливо великих мовних моделей і інтелектуальних агентів, став одним із найбільш впливових факторів структурних змін сучасного світу. Нові технологічні рішення не лише полегшують виконання рутинних процесів, а й трансформують способи комунікації, управління, виробництва та ухвалення рішень. У результаті цифрові інновації дедалі сильніше впливають на ринок праці – сферу, яка традиційно чутливо реагує на технологічні зрушення.

Стрімка інтеграція технологій штучного інтелекту в економічні процеси формує відчутну невизначеність щодо перспектив зайнятості, оскільки алгоритмічні системи дедалі частіше виконують функції, що раніше належали до виключної компетенції людини. Попри перебільшені оцінки потенціалу автоматизації, фактичний рівень загроз є варіативним і потребує всебічного наукового аналізу. Вплив ШІ має асиметричний характер: низка професій демонструє підвищену вразливість до технологічного заміщення, тоді як інші отримують імпульс до інноваційного розвитку та зростання продуктивності. Це зумовлює необхідність глибокого дослідження структурних трансформацій ринку праці та ідентифікації сфер, що належать до зон підвищеного ризику або потенційної експансії.

*Аналіз останніх досліджень і публікацій.* Розвиток ринку праці в умовах цифровізації привертає значну увагу вітчизняних і зарубіжних науковців. Теоретичні засади його аналізу були сформовані А. Смітом і Д. Рікардо та поглиблені неокласиками А. Маршаллом і А. Пігу, тоді як К. Маркс розкрив природу праці як специфічного товару у капіталістичній системі. Важливий поворот у розумінні механізмів зайнятості здійснив Дж. М. Кейнс, який довів

необхідність державного регулювання для досягнення стійкого рівня зайнятості. У подальших дослідженнях значне місце посіли соціальні та гендерні аспекти трудових відносин, що отримали ґрунтовне наукове осмислення у працях Нобелівської лауреатки К. Голдін. В українському економічному дискурсі питання трансформацій ринку праці, впливу цифровізації та штучного інтелекту розкривають О. Грішнова, А. Колот, А. Калина та інші науковці КНЕУ, зосереджуючи увагу на структурних зрушеннях, нових вимогах до навичок і соціально-трудовах наслідках технологічних змін.

У зарубіжних дослідженнях домінує аналіз впливу автоматизації та ШІ на зайнятість. Д. Гінтон акцентує на ризиках технологічного безробіття та посилення економічної нерівності; Е. Бріньолфсон і А. Макафі підкреслюють подвійний ефект цифрових технологій – витіснення рутинних завдань і появу нових професійних напрямів, що вимагають складніших когнітивних і цифрових компетентностей. Д. Ачемоглу та П. Рестрепо доводять, що автоматизація змінює структуру попиту на працю, зменшуючи потребу в одних видах діяльності та стимулюючи формування нових професійних ролей, що дозволяє комплексно оцінити, як цифровізація та ШІ трансформують ринок праці й формують контури його майбутнього розвитку.

*Об'єктом дослідження* кваліфікаційної магістерської роботи є процеси трансформації структури ринку праці під впливом штучного інтелекту.

*Предметом дослідження* є методи і моделі штучного інтелекту для аналізу процесів трансформації структури ринку праці.

*Мета і завдання дослідження.* Метою виконання кваліфікаційної магістерської роботи є застосування методів та інструментів NLP та Big Data, що дозволяють моделювати процеси трансформації структури ринку праці під впливом штучного інтелекту та оцінювати можливі ризики втрати робочого місця через цифровізацію та автоматизацію та гендерних диспропорцій в платі праці.

Для досягнення поставленої мети визначені такі завдання:

- Проаналізувати сутність та ключові функції ринку праці, дослідити вплив технологій штучного інтелекту на трансформацію його сучасної структури.

- Проаналізувати методи й моделі штучного інтелекту, що застосовуються для дослідження процесів структурних змін на ринку праці.
- Здійснити підготовку даних та побудувати семантичне представлення текстової інформації на основі сучасних векторних моделей (SBERT).
- Сформувати модель оцінювання ризику автоматизації професій із використанням SBERT-ембедінгів та релевантних алгоритмів машинного навчання.
- Реалізувати метод крос-лінгвального вирівнювання професій та виконати структурний аналіз відповідності між національними та міжнародними класифікаціями.
- Розробити та застосувати алгоритм оцінювання гендерної нерівності в оплаті праці на основі моделі Logib.
- Здійснити оцінювання гендерних диспропорцій в оплаті праці для закладу освіти, розташованого у сільській місцевості та місті.

*Теоретична, методична та практична значущість отриманих результатів:* поєднання економічного аналізу ринку праці з методами машинного навчання та NLP, що дозволило побудувати модель прогнозування ризику автоматизації професій і провести кластеризацію професій за рівнем ризику. Застосувати методологію Logib для оцінки гендерних диспропорцій на українському ринку праці. Отримані результати дозволяють виділити групи професій, потенційно вразливі до заміни ШІ, оцінити вплив автоматизації на структуру зайнятості та розробити рекомендації для органів влади, роботодавців і освітніх закладів щодо заходів перекваліфікації та моніторингу гендерної рівності.

*Інформаційна база дослідження відображає:* праці зарубіжних та вітчизняних науковців, нормативні та законодавчі акти, український класифікатор професій, статистичні дані щодо ринку праці та зайнятості, зібрані із державних сайтів статистики.

## РОЗДІЛ 1

# ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ АНАЛІЗУ ТРАНСФОРМАЦІЙ РИНКУ ПРАЦІ ПІД ВПЛИВОМ ЦИФРОВІЗАЦІЇ ТА АВТОМАТИЗАЦІЇ

### 1.1 Сутність та функції ринку праці

Ринок праці відіграє важливу роль в системі економічних відносин. На даному ринку зіштовхуються інтереси роботодавців (державних, приватних) і працездатних людей [1]. Трудові відносини, що формуються на ринку, мають глибокий соціально-економічний зміст, оскільки вони безпосередньо впливають на задоволення життєво важливих потреб більшості громадян держави.

Класики політичної економії А. Сміт і Д. Рікардо доводили, що на даному ринку продається праця, яка має свою природну і свою ринкову ціну. На противагу цьому з марксистської позиції специфічним товаром на ринку праці вважається робоча сила виробника. У сучасній світовій економічній думці розвивається перший підхід з уточненням об'єкта ринку праці як трудових послуг (послуг праці), що надаються найманим працівником роботодавцеві згідно з умовами трудової угоди [2].

Категорії «робоча сила» і «праця», звичайно, не тотожні. Робоча сила – це сукупність інтелектуальних, духовних та фізичних здатностей людини до праці, тоді як праця є процесом використання або споживання робочої сили під час створення духовних чи матеріальних благ [3]. При цьому, в умовах ринкової економіки, робоча сила набуває статусу товару, який працівник продає, а роботодавець купує. Праця ж, як безпосередня виробнича діяльність, є джерелом новоствореної вартості та реалізацією цього проданого потенціалу.

Об'єктом купівлі-продажу на ринку праці є послуги праці. Саме вони реалізуються у процесі виробництва, а після його завершення дістають ринкової

оцінки. Принципово важливо, що ринок оцінює саме послуги праці, тобто реальний внесок працівника у створення продукції, а не його потенційні здібності до праці [2]. При оформленні трудових відносин роботодавці беруть до уваги не лише фахові компетенції, а й особистісні та етичні характеристики кандидатів, надаючи пріоритет добросовісним, відповідальним, товаришким та відданим кадрам. Своєю чергою, персонал охочіше обирає роботу в колективах зі сприятливою соціально-психологічною атмосферою.

Суб'єктами ринку праці виступають учасники трудових відносин: працівники, які є власниками робочої сили та продають свої трудові послуги; підприємці, що володіють капіталом і купують ці послуги; об'єднання найманих працівників і роботодавців, а також держава, яка здійснює регулювання взаємовідносин між ними [2].

Ринок праці є системою взаємодії між його суб'єктами та об'єктом, у межах якої формується попит і пропозиція на працю. Взаємовідносини між учасниками цього ринку забезпечують не лише процес обміну трудових послуг на заробітну плату, а й виконання низки важливих соціально-економічних функцій. Саме через реалізацію цих функцій ринок праці впливає на ефективність зайнятості, рівень доходів населення та загальний розвиток економіки.

Інфраструктура ринку праці – це, насамперед, системне явище, тому для ефективного функціонування необхідний зв'язок між усіма її елементами, навіть між конкуруючими в одній сфері (як то у сфері працевлаштування та рекрутингу). Співпраця між ними є запорукою збільшення ефективності кожного учасника. Досвід економічно розвинених країн показує, що співпраця між державними та недержавними посередниками на ринку праці приносить позитивний ефект обом сторонам [4].

Сегментація ринку праці – це процес поділу ринку робочої сили на окремі групи відповідно до певних критеріїв, що залежать від цілей дослідження, аналізу або управління. Ринок праці можна сегментувати за різними ознаками: географічним розташуванням (територіально), рівнем професійної підготовки та кваліфікації, соціальними характеристиками, видами діяльності та сферами

виробництва, співвідношенням числа працівників і роботодавців, а також залежно від умов конкуренції [5].

Функціонування ринку праці зможе забезпечити виконання функцій, притаманних моделі соціально орієнтованої економіки:

- *суспільного поділу праці* [6]. Ринок праці виконує функцію розмежування між найманими працівниками та роботодавцями, забезпечуючи водночас розподіл трудових ресурсів за професійно-кваліфікаційними ознаками, галузями економічної діяльності та регіональною структурою;
- *інформаційну* [6]. У ринкових умовах ціни на товари, рівень попиту й пропозиції, а також відсоткові ставки за кредитами постійно змінюються, що безпосередньо впливає на стан ринку праці. Сам ринок надає учасникам процесу купівлі-продажу товару «робоча сила» необхідну інформацію про умови найму, рівень заробітної плати, наявність робочих місць і працівників, а також про якість робочої сили. Сучасний ринок праці, хоча й реагує на зміни з певним запізненням, усе ж відображає результати конкурентної боротьби, показує середні витрати на робочу силу, визначає дисбаланс між попитом і пропозицією праці та інформує зацікавлених учасників ринку про ці процеси;
- *посередницьку* [6]. Ринок праці виступає посередником між роботодавцями та найманими працівниками, які взаємодіють для задоволення спільних інтересів і потреб щодо вигідних умов обміну товару «робоча сила». У розвиненій економіці працівник має можливість обирати найкраще для себе місце роботи, враховуючи умови праці, рівень оплати, кваліфікаційні вимоги та професійні інтереси. Ринок праці виконує важливу соціальну роль – стимулює працівників до пошуку більш складної, високооплачуваної та престижної роботи відповідно до їхніх здібностей, ділових якостей і професійної майстерності. У той же час роботодавці отримують змогу підбирати найбільш кваліфіковану й відповідну своїм потребам робочу силу;
- *відтворювальну* [6]. Головна функція ринку праці полягає у забезпеченні рівноваги між попитом і пропозицією робочої сили. Задоволення потреб роботодавців відбувається завдяки збільшенню пропозиції, підвищенню

якості та вартості робочої сили, які постійно змінюються й урівноважують одна одну. Саме на ринку праці діє закон вартості, що забезпечує суспільне визнання витрат праці на відтворення товару «робоча сила» та оцінку його суспільної корисності. Тут формуються об'єктивно необхідні умови відтворення цього товару та визначається його вартість. Основою механізму формування ціни на ринку праці є трудова теорія вартості, а також взаємозалежність і співвідношення попиту та пропозиції;

- *стимулюючу* [6]. Суть цієї функції полягає в тому, що ринок орієнтує економіку на задоволення потреб людей, роблячи всіх учасників конкуренції матеріально зацікавленими у цьому процесі. З одного боку, через дію конкуренції ринок праці стимулює ефективніше використання трудових ресурсів для підвищення продуктивності та прибутковості виробництва. З іншого – він формує певні вимоги до якості робочої сили, спонукаючи працівників удосконалювати свої професійні навички й підвищувати конкурентоспроможність, що забезпечує успішну реалізацію їхньої праці на ринку;
- *оздоровлюючу* [6]. Конкуренція сприяє очищенню суспільного виробництва від економічно слабких і неефективних підприємств. Ринок виступає найоб'єктивнішим суддею та найсправедливішим механізмом економічного добору, який забезпечує виживання ефективних господарств і витіснення нежиттєздатних. Цей процес є суворим і невблаганним, однак причиною поразки зазвичай є недостатня компетентність. Виконуючи цю функцію, товарний ринок підтримує прибутковість економіки загалом. Подібним чином діє й ринок праці, надаючи перевагу працівникам із високими професійними якостями та кваліфікацією. Натомість ті, хто має нижчі показники робочої сили, можуть розраховувати лише на малокваліфіковану, низькооплачувану роботу або ризикують залишитися безробітними;
- *регулюючу* [6]. Ринок загалом впливає на всі сфери економіки, передусім на сферу виробництва. Саме він визначає, що, для кого і як виробляти. У певному сенсі ринок бере участь у формуванні пропорцій суспільного виробництва та

розвитку регіонів, сприяючи перерозподілу робочої сили між галузями й територіями – від менш ефективних до пріоритетних і прибуткових. Таким чином, він регулює надлишок трудових ресурсів, забезпечує їх раціональне розміщення та ефективне використання. Зокрема, ринок праці спрямовує рух трудових ресурсів у народному господарстві, стимулюючи підприємців утримувати лише ту кількість і кваліфікацію працівників, які забезпечують стабільний прибуток підприємства.

Виконуючи свої основні функції, ринок праці відіграє ключову роль у формуванні економічної активності населення, забезпеченні зайнятості та ефективному використанні трудових ресурсів. Саме через нього відбувається узгодження інтересів працівників і роботодавців, що сприяє стабільності та розвитку економіки.

Розрізняють зовнішній, або професійний, ринок праці і внутрішній ринок. Зовнішній ринок охоплює відносини між продавцями і покупцями робочої сили в масштабах галузі, регіону, країни. Це відносини, що виникають з метою наймання працівників відповідної спеціальності, професії, а отже, потребують точної класифікації робіт і чіткого визначення їх змісту. Зовнішній ринок характеризується значною плинністю кадрів, тобто він припускає можливість вільного переходу з одного місця роботи на інше [7].

Внутрішній ринок передбачає рух кадрів всередині підприємства, переміщення з однієї посади на іншу. Це переміщення може відбуватися як по вертикалі, так і по горизонталі. По вертикалі - переведенням на інше робоче місце з підвищенням у посаді або на роботу, що потребує вищої кваліфікації. По горизонталі - переведенням на інше робоче місце без змін у кваліфікації, без підвищення в посаді. Розвиток внутрішнього ринку сприяє зниженню плинності кадрів, оскільки підприємство зацікавлене в збереженні працівників, які знають специфіку його виробництва [7].

В Україні зовнішній ринок праці є більш розвиненим, ніж внутрішній. Це зумовлено високою мобільністю робочої сили між підприємствами, активною трудовою міграцією, а також недостатнім розвитком системи внутрішнього

просування кадрів. Водночас великі компанії та державні установи поступово формують елементи внутрішнього ринку праці, створюючи програми кар'єрного росту, підвищення кваліфікації та внутрішнього навчання персоналу.

На зміну структури ринку праці значний вплив мають економічні зміни та структурні трансформації в економіці. Перехід від індустріальної до постіндустріальної економіки супроводжується розвитком секторів послуг та інформаційних технологій. Це призводить до зміни попиту на певні професії та кваліфікаційні групи, спричиняючи структурні зміни на ринку праці. Одночасно технологічний прогрес та автоматизація змінюють характер зайнятості: деякі професії зникають, водночас виникають нові, що потребують іншого набору навичок, що безпосередньо впливає на структуру ринку праці та рівень кваліфікаційних вимог [8].

Демографічні фактори також відіграють важливу роль у формуванні структури ринку праці. Зміни у складі населення, такі як старіння, міграційні процеси та коливання рівня народжуваності, безпосередньо впливають на попит і пропозицію робочої сили в різних регіонах та секторах економіки. Не менш важливим є рівень освіти та професійної підготовки працівників. Зміни в системі освіти та програмах перепідготовки кадрів формують якість робочої сили, що, в свою чергу, визначає доступ до певних сегментів ринку та конкурентоспроможність працівників [8].

На структуру ринку праці впливають також соціально-політичні чинники. Політика зайнятості, трудове законодавство та державні програми підтримки працівників визначають умови праці та рівень зайнятості в різних секторах економіки. Процеси глобалізації та міжнародної конкуренції сприяють інтеграції ринків праці, зміні попиту на робочу силу та появі нових форм зайнятості, таких як дистанційна робота та фріланс, що також змінює структуру ринку [8].

Структура ринку праці формується під впливом комплексної взаємодії економічних, технологічних, демографічних, освітніх, соціально-політичних та глобалізаційних факторів. Ці фактори визначають розподіл робочої сили між

секторами, рівень кваліфікації та професійні навички, а також сприяють адаптації ринку праці до змін у суспільстві та економіці загалом [8].

Враховуючи широкий спектр залежності ринку праці від економічних, соціальних і технологічних чинників, існує необхідність у впорядкуванні професійної діяльності відповідно до рівня кваліфікації, сфери зайнятості та функціональних обов'язків працівників. Тому варто зосередити увагу на можливій класифікації професій.

За основу розроблення Національного класифікатору України "Класифікатор професій" було прийнято Міжнародну стандартну класифікацію професій (ISCO 88: International Standard Classification of Occupations/ILO, Geneva), яку Міжнародна конференція статистики праці Міжнародного бюро праці рекомендувала для переведення національних даних у систему, що полегшує міжнародний обмін професійною інформацією [9].

Національний класифікатор України «Класифікатор професій» використовується центральними органами виконавчої влади, органами місцевого самоврядування, Конфедерацією роботодавців України [10] та всіма суб'єктами господарювання для правильного внесення записів про роботу до трудових книжок працівників [11].

Професійні назви робіт, подані у класифікаторі, рекомендується застосовувати при формуванні нових назв професій і посад, що виникають у зв'язку з розвитком сучасних видів економічної діяльності та технологій [11].

Класифікатор поділяє професії на дев'ять основних розділів, кожен із яких відображає специфіку діяльності та вимоги до освіти працівника [10].

1. *Законодавці, вищі державні службовці, керівники, менеджери (управителі)* – професії, визначенням та формуванням державної політики, законодавчим регулюванням; вищим державним управлінням; правосуддям та прокурорським наглядом; керівництвом об'єднаннями підприємств, підприємствами, установами, організаціями та їхніми підрозділами незалежно від форм власності та видів економічної діяльності характеризуються високим рівнем аналітичного мислення та відповідальністю. Розділ охоплює широке коло професій,

пов'язаних із здійсненням різноманітних функцій управління та керівництва, які в цілому суттєво відрізняються за своєю складністю та відповідальністю [10].

2. *Професіонали* – фахівці з глибокими знаннями галузі фізичних, математичних, технічних, біологічних, агрономічних, медичних чи гуманітарних наук; професійні завдання полягають у накопичення наукових знань, застосування певних теорій, методів та концепцій для вирішення наявних проблем. Діяльність потребує освітньої кваліфікації за першим (бакалаврським), другим (магістерським), третім (освітньо-науковим/освітньо-творчим) рівнем вищої освіти [10].
3. *Фахівці* – працівники, що застосовують наукові положення й методи у виконанні в одній чи більше галузях природознавчих, технічних і гуманітарних наук; зазвичай вимагають передвищої або базової вищої освіти [10].
4. *Технічні службовці* – виконують підготовку, обробку й збереження інформації. Професійні завдання пов'язані з роботою на друкарських машинках чи інших конторських машинах, виконанням секретарських обов'язків, обслуговуванням клієнтів (записами та опрацюванням цифрових даних, поштове обслуговування, операції грошового обігу, надання довідок, реєстрація чи передавання інформації). Для даного розділу професій може бути рівень кваліфікації "молодший спеціаліст", а також професії, що вимагають повної загальної середньої та професійно-технічної освіти чи повної загальної середньої освіти та професійної підготовки на виробництві [10].
5. *Працівники сфери торгівлі та* – забезпечують побутові, торговельні, транспортні та охоронні послуги; багато професій цієї групи мають рутинний характер. Більшість професій, вміщених до цього розділу, вимагає повної загальної середньої та професійно-технічної освіти чи повної загальної середньої освіти та професійної підготовки на виробництві. Ряд професій можуть мати освітньо-кваліфікаційний рівень молодшого спеціаліста [10].
6. *Кваліфіковані робітники сільського та лісового господарств, риборозведення та рибальства* – професії, що передбачають знання, необхідні для лісового господарства, сільськогосподарського виробництва, риборозведення та рибного

промислу. Професійна діяльність у цій сфері охоплює вирощування сільськогосподарських культур, розведення тварин, рибальство та риборозведення, а також догляд, збереження й раціональне використання лісових ресурсів. Основною метою такої діяльності є отримання продукції для подальшої реалізації на ринку – торговельним підприємствам, збутовим організаціям або безпосереднім споживачам. До цього розділу належать професії, що вимагають повної загальної середньої та професійно-технічної освіти чи повної загальної середньої освіти та професійної підготовки на виробництві [10].

7. *Кваліфіковані робітники з інструментом* – спеціалісти будівництва, виробництва та видобутку, що використовують інструменти та обладнання, визначення стадій робочого процесу, характеристик та призначення кінцевої продукції. Ці професії вимагають повної загальної середньої та професійно-технічної освіти чи повної загальної середньої освіти та професійної підготовки на виробництві. Для деяких професій, пов'язаних з виконанням робіт високої кваліфікації, потрібна кваліфікація молодшого спеціаліста [10].
8. *Робітники з обслуговування, експлуатації та контролювання за* – експлуатація й нагляд за машинами та технологічними лініями, у тому числі автоматизованими системами; розроблення корисних копалин, нагляд за їх видобутком, ведення робочого процесу та виробництво продукції на устаткуванні чи машинах, керування транспортними засобами, пересувними установками, складання виробів із деталей та вузлів. Кваліфікаційні вимоги до цих професій включають повну загальну середню освіту та здобуття професійно-технічної або виробничої підготовки. Однак, для виконання висококваліфікованих завдань у окремих випадках може вимагатися наявність диплома молодшого спеціаліста [10].
9. *Найпростіші професії* – завдання пов'язані з продажем товарів на вулиці, прибиранням, збереженням та охороною майна, пранням, чищенням, прасуванням та виконанням низькокваліфікованих робіт у видобувній, риболовній, сільськогосподарській, будівельній та промисловій галузях. Для

виконання професійних завдань досить базової загальної середньої освіти або початкової загальної освіти та мінімальної професійної підготовки на виробництві чи інструктажу [10].

Опис кожного розділу професій демонструє значну різноманітність обов'язків і водночас підкреслює, що окремі трудові функції можуть бути потенційно автоматизовані. Сьогодні частину рутинних або технічно визначених дій здатні виконувати як роботизовані системи, так і великі мовні моделі, що все активніше інтегруються у робочі процеси.

У цьому контексті стає очевидним, що розгляд структури професій уже неможливо відокремити від ширших технологічних змін. Зростання можливостей штучного інтелекту та його швидке поширення змінюють підходи до організації праці, трансформують вимоги до компетенцій і формують нові виклики для роботодавців та працівників. Саме тому логічним є подальший аналіз впливу ШІ на ринок праці.

## **1.2 Вплив штучного інтелекту на структуру сучасного ринку праці**

Приблизно два роки тому світ отримав відкритий доступ до ChatGPT – і це стало переломним моментом для ринку праці. Модель дуже швидко перетворилася на інструмент, що полегшує й прискорює виконання багатьох типових завдань. Наслідки такого стрімкого впровадження ШІ бізнес відчув уже на початку 2025 року [12].

Швидке поширення систем штучного інтелекту формує нову траєкторію розвитку сучасних ринків праці та радикально змінює логіку взаємодії між технологіями, людськими ресурсами та продуктивністю. Вплив ШІ не є одностороннім він одночасно може посилювати ризики структурного безробіття та створювати передумови для формування нових сегментів зайнятості. Ця подвійність і становить ключову дилему сучасних трудових ринків – чи веде ШІ до

витіснення робочої сили, чи навпаки стає каталізатором якісної трансформації робочих місць і появи нових професійних ролей [13].

Якщо є розглядати вплив ІІ на різні сектори економіки – від медицини та логістики до сільського господарства і фінансової сфери – можна визначити, що технологічний прогрес створює нерівномірний вплив на зайнятість. У низці галузей автоматизація стимулює заміну людини високопродуктивними алгоритмами (наприклад, у логістиці чи банківському секторі). Водночас в інших напрямках, як-от медицина, робототехніка чи інженерія, ІІ виступає ключовим інструментом підвищення ефективності фахівця, а не його заміни. Таким чином, підтверджується ідея про те, що розвиток ІІ не супроводжується універсальною втратою робочих місць, а призводить до зернистої структурної перебудови трудового середовища[14].

Найбільш уразливими до автоматизації є рутинні та стандартизовані види робіт, які легко формалізуються у вигляді алгоритмічних процедур. У таких секторах ІІ може дійсно посилювати безробіття, оскільки машинні системи демонструють вищу швидкість, точність та меншу вартість у виконанні однотипних операцій. Водночас у професіях, що спираються на аналітичні, соціальні, креативні компетенції, ІІ частіше виконує роль інструмента підсилення, а не заміни працівника. Таким чином, технологія не одноманітно руйнує ринок праці, а різними шляхами перебудовує його структуру.

З іншого боку, ІІ варто розглядати як довгостроковий потенціал щодо створення нових сфер зайнятості. У процесі технологічного переходу з'являються спеціалізації, які не існували раніше: розробка, підтримка, етичний моніторинг моделей ІІ, аналіз даних, тренування моделей, управління автоматизованими системами. Поряд із цим, у вже існуючих галузях виникає потреба в працівниках, здатних взаємодіяти з інтелектуальними системами, інтерпретувати їхні результати та поєднувати машинні висновки з контекстним людським судженням. Ця логіка формує підхід, за яким зайнятість не зникає, а змінює свої характеристики [13].

Масштаби та глибина проникнення ІІ в економічні процеси значною мірою зумовлені рівнем технологічної готовності країн. Європейський контекст,

широко представлений через звіти ЄС, МВФ, Банку Італії та ОЕСД, показує, що виклики автоматизації нерозривно пов'язані зі спроможністю держав інвестувати в цифрову інфраструктуру, технологічний капітал та розвиток компетенцій. Особлива увага приділяється дефіциту навичок, який стає критичним фактором у формуванні нових трудових моделей. Технології самі по собі не створюють безробіття – його провокують прогалини у системі освіти, відсутність програм перепідготовки та повільна адаптація ринку до нових технічних стандартів [14].

Особливо важливою для розуміння майбутнього ринку праці є частина, присвячена прогнозам щодо зникнення професій та появи нових. Згідно з проаналізованими Piro міжнародними дослідженнями, найбільших трансформацій зазнають професії середнього рівня кваліфікації, що традиційно виконують повторювані або формалізовані завдання. Натомість зростає попит на спеціалістів у галузях аналізу даних, кібербезпеки, управління алгоритмічними системами, розробки ШІ та підтримки високотехнологічного виробництва. Кількість робочих місць у цілому не зменшується – відбувається зміщення структури зайнятості на користь компетенцій, які дозволяють працівникам взаємодіяти з інтелектуальними системами [14].

Проте на разі існує феномен так званої «бульбашки ШІ» який полягає не у тому, що штучний інтелект є марною або переоціненою технологією, а в тому, що суспільні очікування щодо його можливостей значно випереджають реальні темпи розвитку та масштаби практичного застосування. Цей ефект є типовим для періодів швидких технологічних змін – ринок, медіа та політичний дискурс схильні створювати наратив, у якому інновація видається більш універсальною, розумною та трансформаційною, ніж вона є насправді у поточний момент [15].

В основі «бульбашки ШІ» лежить гіперболізована віра в те, що алгоритми можуть автоматизувати майже всі види інтелектуальної праці, приймати об'єктивні рішення, забезпечити повну економічну ефективність та стати ключовим рушієм економічного прориву. Проте на практиці більшість сучасних моделей ШІ мають суттєві обмеження: вони залежать від великих масивів даних, важко адаптуються до нетипових ситуацій, часто генерують помилки або упередження, і потребують

масштабної людської участі на етапах навчання, налаштування та контролю. Це означає, що реальне поглинання ринку праці ШІ відбувається значно повільніше, ніж передбачають популярні прогнози [15].

Гіперочікування формуються також через інформаційний шум: бізнес, інвестори й держави перебувають у великій конкуренції за лідерство у сфері ШІ, що стимулює створення гучних заяв, амбітних стратегій і перебільшених обіцянок. У результаті технологія стає символом інноваційності, навіть якщо її фактичне впровадження у конкретних секторах обмежене або перебуває на експериментальній стадії. Так виникає ринковий та суспільний дисбаланс: капітал, інтерес і політика рухаються швидше, ніж технологічні можливості [15].

Важливо, що даний феномен не заперечує реального потенціалу штучного інтелекту – вона лише підкреслює різницю між очікуваннями та реальністю. Історія інновацій демонструє, що після фази надмірного ентузіазму зазвичай настає період корекції, коли ринок відсікає надмірно завищені очікування і переходить до стадії більш тверезої, масштабованої та економічно обґрунтованої інтеграції технології. Аналогічно і з ШІ: попри нинішнє «накачування» ринку гучними прогнозами, довгостроковий ефект буде формуватися не через моду на ШІ, а через реальні кейси успіху, інституційні зміни та здатність робочої сили адаптуватися до нових вимог[15].

Попри те, що ринок поступово переходить від гучних обіцянок до практичних, перевірених сценаріїв застосування штучного інтелекту, саме реальні технологічні механізми визначають, наскільки ефективно ШІ може бути інтегрований у конкретні професійні чи дослідницькі процеси. Це означає, що для коректного аналізу впливу ШІ важливо не лише розглядати економічні чи соціальні наслідки, а й розуміти технічну природу його роботи. Зокрема, необхідно усвідомлювати, яким чином моделі інтерпретують інформацію, обробляють текст і «розуміють» значення слів.

## 1.3 Методи і моделі штучного інтелекту дослідження процесів трансформації структури ринку праці

### 1.3.1 Огляд методів NLP

При роботі із словами для нас важливо передати системі їх сенс. Адже будь яке слово у нашій свідомості є зрозумілим для нас, проте для системи ні. Тому існує потреба пояснити наші слова для комп'ютера його мовою. Тому виникає потреба в утворенні embeddings.

Ембедінги (embeddings) у природній мовній обробці – це низькорозмірні векторні подання, які відображають мовні одиниці (слова, значення слів, фрази, речення чи цілі документи) у багатовимірному просторі таким чином, щоб їхні семантичні властивості та подібності були збережені. В основі цієї концепції лежить ідея репрезентації значення через геометричні відношення між векторами: мовні одиниці зі схожим змістом мають розташовуватися близько одна до одної у векторному просторі. Первинно ембедінги розроблялися для слів (word embeddings), однак з розвитком методів представлення знань вони були поширені на інші рівні, включно з сенсовими ембедінгами, графовими поданнями та представленнями речень і документів. Сучасні підходи також включають контекстуалізовані ембедінги (наприклад, ELMo, BERT), у яких одне й те саме слово може мати різні векторні представлення залежно від контексту, що дозволяє моделі точніше відображати значення та забезпечує глибше семантичне розуміння тексту [16].

Досить популярною є модель BERT (Bidirectional Encoder Representations from Transformers) – це попередньо навчена мовна модель, яка використовує двонаправлений контекст для підвищення продуктивності над завданнями з обробки природної мови. Використовуючи архітектуру Трансформатора, він захоплює нюансовані значення слів, розглядаючи контекст з обох напрямків у реченні [17].

Мережа Sentence-BERT (SBERT) є модифікацією попередньо навченої мережі BERT, яка використовує структури сіамських і триплетних мереж для отримання семантично змістовних векторних подань речень [18]. Сіамська нейронна мережа – це мережа, що складається з двох ідентичних нейронних мереж, які мають спільні ваги. Обидві мережі працюють паралельно, а їхні вихідні результати порівнюються, зазвичай за допомогою метрики відстані, такої як косинусна відстань. Сіамські мережі добре підходять для задач визначення подібності [19]. Властивість спільного використання ваг у сіамських мережах гарантує узгоджені передбачення, оскільки кожна з мереж обчислює ту саму функцію [20].

TF-IDF (term frequency–inverse document frequency) – це підхід у галузі обробки природної мови та інформаційного пошуку, який дозволяє оцінити значущість слів у тексті. Метод визначає, наскільки документ є релевантним певному запиту, призначаючи кожному терміну вагу залежно від того, як часто він трапляється в конкретному документі та наскільки рідкісним є серед всієї сукупності документів. Ключова ідея TF-IDF полягає у тому, що важливість слова зростає, якщо воно часто використовується в документі, але водночас рідко зустрічається в інших документах колекції [21].

У межах цієї роботи важливо розглянути як класичні статистичні методи представлення тексту, так і сучасні нейромережеві підходи, оскільки вони відображають різні парадигми розуміння мови. TF-IDF забезпечує інтерпретоване, структуроване та просте у реалізації подання текстів, яке залишається популярним у задачах пошуку, базової класифікації та аналізу частотності. Натомість SBERT, як представник контекстуалізованих ембедінгів на базі трансформерів, дозволяє моделі вловлювати семантичні зв'язки між реченнями та працювати на рівні змістової подібності, що недоступно для традиційних методів.

Порівняння моделей буде здійснене за допомогою Ridge-регресії (регуляризація L2) – це техніка регуляризації, яка додає до функції втрат штрафний член, пропорційний квадрату коефіцієнтів моделі. Крім класичної ролі в боротьбі з мультиколінеарністю, Ridge використовується для зменшення перенавчання та

стабілізації моделей при роботі з високовимірними даними (наприклад, TF-IDF або ембедінги), оскільки регуляризація сприяє «стисканню» великих ваг, але не виключає ознаки повністю. У контексті машинного навчання це робить Ridge ідеальним вибором як базову, лінійну та стабільну модель – вона дозволяє оцінити, наскільки ембедінги містять корисну інформацію, без великого ризику перенавчання [22].

У контексті дослідження моя мета полягала у порівнянні якості різних методів текстового подання (TF-IDF та SBERT), а не у побудові максимально складної моделі прогнозування. Тому Ridge-регресія була вибрана як контрольована, стабільна та лінійна модель, яка мінімізує вплив самої моделі на результати та дозволяє оцінити саме інформативність ембедінгів.

Основні метрики згідно яких я оцінюю якість моделі наступні:

- $R^2$  – коефіцієнт детермінації, числове значення якого вказує якою мірою варіація залежної змінної  $Y$  пояснюється варіацією незалежних змінних. Чим ближчий коефіцієнт до одиниці, тим більша варіація залежної змінної визначається варіацією незалежних змінних [23].
- MAE – середня абсолютна помилка. Обчислює середнє абсолютне значення різниці між прогнозованими та фактичними значеннями [24].
- MSE (mean squared error) – середньоквадратична помилка. Визначається як середнє квадратичне відхилення прогнозованих значень від фактичних [24].
- Коефіцієнт Спірмена – це непараметричний критерій оцінювання сили та напрямку монотонної залежності між двома змінними, який ґрунтується на порівнянні їхніх рангів замість використання фактичних числових значень. Він дає змогу визначити, наскільки узгоджено змінюються показники у прямому чи зворотному напрямку. Значення коефіцієнта лежить у межах від -1 до 1, а саме крайні значення відображають повний монотонний зв'язок, а нульове значення свідчить про його відсутність [25].
- Кореляція (лат. correlation – відповідність) – статистична залежність між величинами, яка не має строго функціонального характеру. Кореляційна залежність виникає за умови, коли одна з величин залежить не тільки від заданої

другої, а й від деяких випадкових факторів; або, коли серед умов, від яких залежать обидві величини, є загальні для них обох [26, с.22].

Кореляційний зв'язок може проявлятися у різних формах, від повної незалежності до функціональної залежності. Крім того, характер взаємозв'язку між величинами може відрізнятися. Таким чином, важливо визначити форму, напрямок і ступінь кореляції між ними. За формою кореляція може бути прямолінійною або криволінійною, а за напрямком - прямою або оберненою[26,с.23].

У роботі також буду досліджувати найкращу модель шляхом побудови декількох та вибору найкращої із них. Проте перед цим здійсниться підготовка даних, одним із етапом якою буде масштабування. Алгоритми машинного навчання часто погано працюють із немасштабованими ознаками. Коли характеристики даних мають різний діапазон значень, моделі можуть навчатися надто повільно або давати нестабільні результати. Саме тому масштабування один із ключових етапів попередньої обробки. StandardScaler приводить ознаки до нульового середнього та одиничного стандартного відхилення, MinMaxScaler перетворює їх у визначений інтервал, зазвичай [0; 1], а RobustScaler забезпечує стійкість до викидів, використовуючи медіану та міжквартильний розмах. Усі ці методи реалізовані у scikit-learn і можуть бути інтегровані в пайплайни для коректної та відтворюваної обробки даних [27].

Види моделей, що будуть досліджені у роботі:

- Лінійна регресія – статистична модель, яка встановлює залежність між цільовою змінною та набором предикторів шляхом пошуку лінійної комбінації ознак, що найкраще наближує спостережувані дані. Метод мінімізує суму квадратів відхилень між реальними та прогнозованими значеннями, забезпечуючи пояснення коефіцієнтів і можливість аналізу значущості ознак. Лінійна регресія є базовою моделлю, що є основою для більш важчих методів [28].
- RandomForest – ансамблевий метод, який поєднує велику кількість незалежних рішучих дерев для отримання більш точних і стабільних прогнозів. Ґрунтується на ідеї випадкового підбору підмножин ознак та випадкового семплювання спостережень, що дозволяє деревам бути менш корельованими між собою. Така

стратегія значно зменшує ризик перенавчання та робить модель стійкою до шуму у даних, зберігаючи при цьому здатність моделювати складні нелінійні залежності [29].

- XGBoost підвид градієнтного бустингу, а саме метод послідовного побудування слабких моделей (зазвичай дерев рішень), де кожне наступне дерево коригує помилки попередніх. Модель навчається у напрямку градієнта функції втрат, що забезпечує високу точність на складних даних. XGBoost, як розширення цього підходу оптимізує обчислення та вводить регуляризацію, що підвищує швидкість і стійкість моделі. Один із найефективніших для табличних даних завдяки здатності виявляти складні, нелінійні закономірності [30].
- Багатошарова нейронна мережа (MLP) – це тип глибинної моделі, що складається з кількох послідовних шарів нейронів, кожен з яких застосовує лінійне перетворення та нелінійну активацію. Нелінійні активаційні функції дозволяють моделі апроксимувати довільно складні залежності у даних [31].

Ідея роботи полягатиме у використанні двох датасетів із застосуванням transfer learning. Це методика, за якої модель, попередньо натренована на великому та загальному датасеті, використовується для розв'язання іншої, більш вузької задачі. Такий підхід дозволяє повторно використовувати вже сформовані представлення ознак, що суттєво зменшує кількість необхідних локальних даних і прискорює навчання. Як зазначають автори базового підручника «*Deep Learning*» (Ian Goodfellow, Yoshua Bengio та Aaron Courville) [31] та François Chollet у «*Deep Learning with Python*» [32], сила transfer learning полягає в тому, що модель переходить до нової задачі, уже володіючи високорівневими абстракціями та універсальними закономірностями, сформованими під час навчання на великих масивах даних. Це дозволяє алгоритму значно швидше адаптуватися до специфіки нової предметної області та досягати більшої узагальненої здатності порівняно з моделями, що навчаються з нуля.

Навчання має бути інформативним, щоб при перекладі не було втрачено важливої інформації, тому проведу семантичне вирівнювання.

Сучасні підходи до семантичного представлення в природній мові спираються на принцип розподільчої семантики, згідно з яким значення лексеми визначається її контекстом, а не як абстрактна сутність. Цей принцип є фундаментом сучасних векторних ембеддінгів, у яких слова з подібними контекстами розташовуються близько в багатовимірному просторі [33].

Під час масштабного попереднього тренування мультимовні трансформери, такі як XLM-R, формують спільний векторний простір для численних мов, що забезпечує ефективне крос-лінгвальне вирівнювання контекстуальних представлень та полегшує перенесення знань між мовами [34].

Модель Sentence-BERT (SBERT) модифікує класичний BERT, застосовуючи сіамську або тріплет-мережу для отримання ембеддінгів речень, які можна безпосередньо порівнювати за допомогою косинусної подібності. Косинусна подібність вимірює кут між векторами незалежно від їхньої довжини, що дозволяє фокусуватися на напрямі вектора (семантичному змісті), а не на його величині [35].

Застосовуватиму навчену модель на українських даних та буду досліджувати результати. Так як кількість професій досить велика, цікаво простежити візуальне сприйняття, тому у даному випадку візуальне сприйняття можливе за допомогою методу головних компонент.

Метод головних компонент (PCA) – це класичний статистичний підхід до зниження розмірності даних, що дозволяє зберегти максимально можливу частку інформації під час проєкції високорозмірних векторів на простір меншої розмірності. PCA виконує лінійне перетворення початкових даних у нову систему координат, де перша головна компонента відповідає напрямку найбільшої дисперсії, друга – наступному за величиною, і так далі [36].

У межах задач обробки природної мови PCA часто застосовується для візуалізації семантичних просторових структур, зокрема для аналізу ембеддінгів, які зазвичай мають сотні вимірів. Використання PCA дозволяє відобразити подібність між об'єктами (наприклад, професіями) у двовимірному просторі, при цьому

зберігаючи ключові закономірності їхнього розташування у вихідному багатовимірному просторі [37].

Побудова PCA-графіка здійснюється шляхом обчислення ковариаційної матриці ознак, визначення власних векторів та проєкції кожного первинного векторного представлення на перші дві головні компоненти. Отримані координати (PC1 та PC2) дозволяють наочно проаналізувати структуру даних, виявити групування об'єктів та дослідити відповідність між їхніми ознаками та поведінкою моделі [38].

Згодом після проведення тренування на українських даних здійснюватимуть кластерний аналіз. Це набір методів, які використовуються для класифікації багатовимірних спостережень, де кожне спостереження описується набором вихідних змінних  $X_1, X_2, \dots, X_m$ . Основна мета кластерного аналізу полягає в об'єднанні подібних об'єктів у групи, які називаються кластерами. Термін "кластер" походить з англійської мови (cluster) і означає згусток, пучок або групу. У науковій літературі близькими за змістом поняттями є клас, таксон і згущення [39, с. 57].

Розрізняють певну кількість методів кластерного аналізу. У даній роботі буде використано метод К-середніх (K-means clustering) який застосовується, коли дослідник має попередні припущення щодо кількості кластерів. У цьому методі дослідник задає кількість кластерів заздалегідь, і алгоритм кластеризації розподіляє дані таким чином, щоб кластери максимально відрізнялися один від одного. Перевагою цього методу є можливість перевірки статистичної значущості відмінностей між утвореними кластерами [40, с. 109].

Робастність моделі в задачах машинного навчання визначається як її здатність зберігати стабільність результатів при зміні навчальної вибірки, наявності шуму в даних або зміні параметрів ініціалізації. На практиці це означає, що модель повинна продукувати близькі значення прогнозу навіть у разі незначних змін входу, структури вибірки або початкових умов навчання [41].

У задачах обробки природної мови робастність має особливе значення через високу варіативність текстових даних, контекстну залежність семантики та багатовимірність векторних просторів. Моделі, що працюють з ембедінгами

(зокрема Sentence-BERT), чутливі до структури корпусу, тому перевірка стабільності їхніх результатів є обов'язковою умовою наукової достовірності експерименту.

Стандартними методами перевірки робастності в задачах ML є:

- bootstrap-аналіз полягає в багаторазовому формуванні вибірок із поверненням на основі початкового набору даних і повторному навчанні моделі на кожній із них. У результаті формується розподіл значень метрики якості або прогнозу, за яким можна оцінити дисперсію та стандартне відхилення результатів [42].
- ablation study використовується для визначення внеску окремих груп ознак у загальну продуктивність моделі шляхом їх послідовного вилучення [43].
- перевірка стабільності на різних значеннях random seed визначає початкові умови стохастичних процесів, зокрема ініціалізацію параметрів та формування підвбірок. Зміна цього параметра може впливати на результати навчання та фінальні показники якості моделі. Перевірка стабільності полягає у багаторазовому повторенні навчання моделі XGBoost з незмінними гіперпараметрами, але різними значеннями random seed. Якщо отримані метрики демонструють незначну варіацію, це свідчить про низьку чутливість моделі до випадкових чинників і підтверджує її робастність [44].

У межах даного дослідження перевірка робастності критично важлива через використання transfer learning (Oxford – Ukraine), складну структуру моделі (ембедінги у поєднанні із інженерними ознаками та побудова моделі XGBoost) та відсутність реальних цільових міток для українських професій. Тому робастність у даній роботі виступає не допоміжним, а валідаційним елементом, який підтверджує надійність отриманих результатів.

Використання NLP-методів, ембедінгів і алгоритмів машинного навчання у поєднанні з перевіркою робастності створює надійну основу для аналізу впливу ШІ на структуру професій та ринок праці загалом. Однак дослідження трансформацій у сфері праці неможливе без урахування ще одного критично важливого виміру –

гендерної рівності в оплаті праці. Адже технологічні зміни можуть не лише створювати нові можливості, а й посилювати або відтворювати існуючі нерівності.

### 1.3.2 Огляд інструментів Big Data

Рівна оплата праці є ключовим фактором досягнення рівності на робочому місці та важливим елементом належного корпоративного управління. Прозора система оплати праці та аналіз рівної оплати праці допомагають роботодавцям сприяти та впроваджувати рівну оплату праці у сфері їхнього впливу.

Рівна оплата за рівну працю є ключовим компонентом гендерної рівності на робочому місці. Крім того, рівна оплата праці сприяє довірі працівників до своєї компанії та підвищує привабливість організації на ринку праці. Свідомо розроблена, прозора система оплати праці формує основу для справедливої компенсації для всіх працівників, незалежно від статі. Федеральні ресурси підтримують компанії та організації в аналізі рівності оплати праці та розробці чіткої та прозорої системи оплати праці.

Незрозуміла різниця в заробітній платі може являти собою гендерно-специфічну дискримінацію в оплаті праці. Дискримінація в оплаті праці виникає, коли жінки та чоловіки отримують різну оплату за працю рівної цінності. Праця вважається еквівалентною, коли вимоги та навантаження, які вона накладає на неї, порівнянні.

Щоб розпочати аналіз за допомогою моделі, варто підготувати дані. І якщо для малих підприємств можна зробити це вручну, то для великих компаній з багатотисячним персоналом це є надзвичайно складною задачею, адже інформація про співробітників зберігається у численних системах – ERP, HRIS, системах обліку часу, системах нарахування заробітної плати. Традиційна ручна обробка таких даних просто не підходить: вона займає багато ресурсів, супроводжується високим ризиком помилок і не гарантує достатньої якісної стандартизації. У цьому

контексті використання технологій Big Data у поєднанні зі штучним інтелектом стає ключовим для автоматизованої та масштабованої інтеграції даних.

Великі організації оперують мільйонами записів працівників з сотнями змінних, що вимагає складної архітектури для централізації даних. Щоб зібрати цю інформацію, рекомендується використовувати потокові технології (наприклад, Apache Kafka або Spark Streaming) або ETL/ELT-інструменти (як-от Amazon Web Services), які здійснюють регулярне автоматичне завантаження даних з різних систем HR до єдиного сховища даних – Data Lake [45].

У Data Lake дані зберігаються на різних рівнях (raw, curated, cleansed), що дозволяє зберегти первісні записи та водночас мати очищені та стандартизовані дані для подальшого аналізу. Такий підхід відповідає принципам Big Data-обробки, оскільки дає змогу працювати як зі структурованими, так і з напівструктурованими або неструктурованими даними, наприклад з текстовими описами посад чи резюме[45].

Застосування алгоритмів машинного навчання та ШІ дозволяє автоматизувати цей процес і значно підвищити якість підготовлених даних. Ключовим завданням є стандартизація таких характеристик, як посадова функція, рівень кваліфікації, освіта. У великих компаніях назви посад можуть бути дуже різноманітні, і без автоматизації NLP їх важко привести до єдиної системи. Big Data дає змогу структурувати неструктурований текст, наприклад опис вакансій, за допомогою NLP-моделей, автоматично класифікуючи посадові назви за міжнародними класифікаціями (ESCO, ISCO) та готуючи дані, сумісні з вимогами Logib [45].

Після того як усі дані очищені, нормалізовані і стандартизовані, ШІ може виконати фінальну валідацію перед експортом: перевірити логіку значень, повноту полів, відповідність структурі Logib і підготувати фінальний файл у форматі, необхідному для аналізу (наприклад, Excel-файл для Logib). Цей конвеєр повністю автоматизує підготовку даних і суттєво знижує навантаження на HR-аналітиків, перетворюючи процес із багатогодинної ручної праці на ефективний і надійний pipeline.

Крім того, сучасні тренди підтверджують доцільність такого підходу. Зокрема, аналітики відзначають, що штучний інтелект у HR може прискорити інтеграцію нових працівників, автоматизувати обробку кадрової інформації та звільнити ресурси, які раніше витрачалися на рутинні задачі, що особливо актуально для компаній з великою кількістю персоналу [46]. Українські та міжнародні практики також демонструють, що Big Data разом з AI все активніше використовуються для управління великими кадровими масивами: автоматизованої інтеграції систем, очищення даних, побудови аналітичних моделей та оцінки ризиків, пов'язаних з рівністю та продуктивністю [47]. Такий підхід дозволяє підготувати дані для детального аналізу.

## РОЗДІЛ 2

# МОДЕЛЮВАННЯ ПРОЦЕСІВ ТРАНСФОРМАЦІЇ НА РИНКУ ПРАЦІ ІЗ ЗАСТОСУВАННЯМ МЕТОДІВ NLP

### 2.1 Підготовка даних та побудова семантичних представлень

Безпосередньо аналіз професій розпочинаю із збору датасету. Першочергово основою для даних стане Національний класифікатор України "Класифікатор професій" [11]. Представлені професії містять чітку назву, яка затверджена в українському реєстрі із відповідним кодом професії. Це робить дослідження унікально орієнтованим на український ринок праці.

Досліджуваний датасет складається із чотирьох колонок, а саме «Код», «Назва», «Група» та «Опис». Для формування повного тексту професії у подальшому я об'єднаю «Назва» та «Опис», щоб NLP аналіз був більш ґрунтовний. Тобто фокус був здійснений не лише на назву професії і намагання моделі опрацювати та зрозуміти професію виключно по назві, а намагання робити висновки по опису кваліфікацій та обов'язків згідно службових обов'язків працівників.

Проте, щоб якісно зробити прогноз професій які можуть бути автоматизовані має бути певна база на яку можна опиратись. У іншому випадку немає можливості здійснити чітку перевірку і тоді дослідження стає менш точним. Тому у даному випадку за основу використаю дослідження «*The Future of Employment: How Susceptible Are Jobs to Computerisation?*» Oxford Martin Programme on Technology and Employment [48], яке вперше запропонувало метод оцінювання автоматизованості професій. Разом із тим, у моїй роботі застосовано оновлений датасет, що ґрунтується на сучасних оцінках ризику автоматизації. Ці оцінки сформовані на основі офіційних даних Bureau of Labor Statistics, зокрема

класифікації професій SOC[49]; task-based моделей, які аналізують набір трудових завдань усередині професій, відповідно до методології McKinsey [50] та Brookings Institution [51]; міжнародних досліджень OECD (Організація економічного співробітництва та розвитку), що демонструють відмінності у ризиках автоматизації між професіями залежно від структури завдань [52].

Ідея роботи полягає у дослідженні професій на можливість автоматизації і в подальшому зміненні штучним інтелектом. Дану роботу виконуватиму за допомогою мови Python, код роботи наведено у *Додатку А*.

Почала роботу із підготовки власного файлу, а саме очищення, нормалізація. Це важливо, адже у даному випадку маю роботу не просто із даними, а із словами. Тому у даному випадку це ще більш важливо.

Переходжу до побудови ембедінгів Sentence-BERT та TF-IDF. Це важливий перший крок роботи, адже необхідно підготувати наявну інформацію до того, щоб система її розуміла. Порівняння цих двох підходів є принципово важливим, оскільки дає змогу оцінити, як різні способи векторизації – частотні та контекстні впливають на результати моделювання, та визначити, який з них є більш ефективним у конкретному завданні аналізу текстових даних. У табл. 2.1 наведені результати порівняльного аналізу.

Таблиця 2.1 – порівняння моделей ембедінгів Sentence-BERT та TF-IDF

	R <sup>2</sup>	MAE	MSE
SBERT	0.503362	0.216937	0.071138
TF-IDF	-0.003119	0.346576	0.143687

*Джерело: розроблено автором*

Провівши аналіз результатів можна зробити висновок, що SBERT значно переважає над класичною моделлю TF-IDF. Дану перевагу вказав коефіцієнт детермінації, який значно вищий і має додатне значення, що вказує на здатність SBERT відтворювати хоча б половину варіації ризику автоматизації, тоді як TF-IDF взагалі не вловлює жодної структури в даних. Також згідно середньої абсолютної та середньоквадратичної помилок SBERT показує набагато нижче значення. Це підтверджує, що контекстні семантичні ембедінги значно краще

відображають зміст професій, ніж поверхнева частотна модель TF-IDF, і тому саме SBERT є оптимальним вибором для задачі оцінки ризику автоматизації.

Для кращого порівняння використала критерій Спірмена. На рис. 2.1 наведено порівняння якості моделей, побудованих на TF-IDF та SBERT-ембедінгах. Результати демонструють суттєву перевагу SBERT над класичним TF-IDF за всіма ключовими метриками. Найбільш показовим є коефіцієнт рангової кореляції Спірмена, який для SBERT досягає приблизно 0.68, тоді як для TF-IDF він практично відсутній. Це означає, що модель на SBERT не лише точніше прогнозує значення, але й правильно зберігає відносний порядок професій за ступенем автоматизованості, що є критично важливим у контексті аналізу ринку праці.

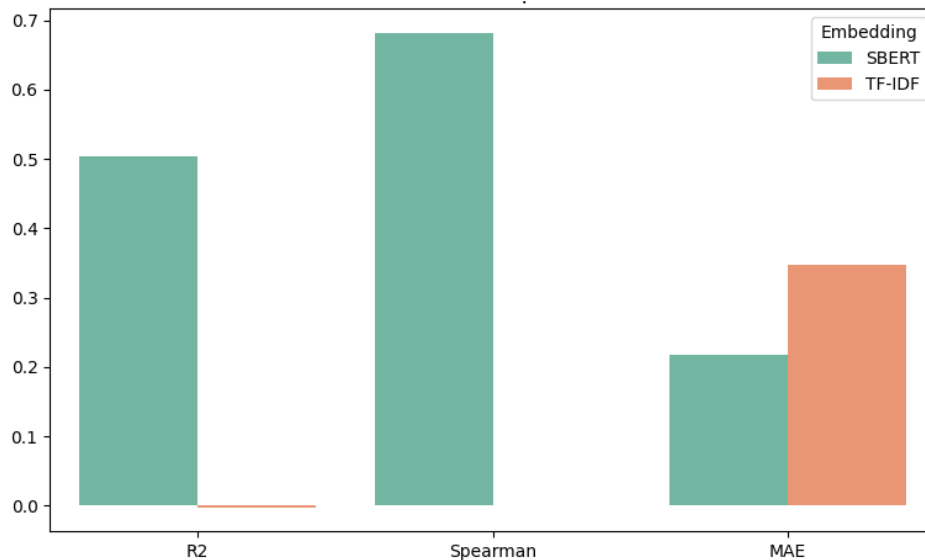


Рисунок 2.1 – Порівняння якості SBERT та TF-IDF

*Джерело: розроблено автором*

Використання коефіцієнта Спірмена є обґрунтованим на даному етапі дослідження, оскільки для аналізу автоматизації першочергове значення має не абсолютне значення прогнозу, а його здатність коректно ранжувати професії за рівнем ризику. У різних країнах, зокрема при переході від американського датасету до українського, абсолютні значення можуть зміщуватися, однак відносний порядок професій (які з них є більш чи менш вразливими до автоматизації) залишається основним предметом дослідження. Саме тому кореляція Спірмена є найбільш доречною метрикою – вона оцінює, наскільки модель правильно зберігає

структуру ранжування, а не точну величину. Це дозволяє забезпечити коректність аналізу та робить висновки дослідження стабільними і методологічно узгодженими.

Однак, навіть маючи якісні векторні представлення тексту, необхідно врахувати, що ембедінги, попри свою контекстну глибину, не завжди здатні явно відобразити функціональні характеристики професій. Тому наступним етапом було введення додаткових доменних ознак, які описують професії через наявність технічних, соціальних, ручних та творчих компонент. Це дозволяє моделі не лише розуміти текст, але й кількісно оцінювати ключові аспекти характеру праці, що є критично важливим для задачі прогнозування ризику автоматизації.

Після того як я сформувала ембедінги, стало зрозуміло, що самих тільки векторів все одно мало. Ембедінги добре передають загальний зміст, але вони не завжди чітко виділяють конкретні характеристики професій. Наприклад, для моделі важливо знати, чи професія пов'язана з технікою, чи з людьми, чи це фізична робота, чи творчість. Тому я додатково зробила чотири невеликі словники ключових слів, кожен під свою категорію.

Перший словник називається «tech\_words» та відповідає за технічний напрям. Я просто збрала слова, які найчастіше зустрічаються у професіях технічного напрямку: «машина», «робот», «інженер», «система», «обладнання», «алгоритм», «програмування» і так далі. До цього списку також увійшли слова, пов'язані з автоматизацією та комп'ютерними процесами.

Другий «social\_words» містить слова про взаємодію між людьми: «комунікація», «керування», «навчання», «клієнт», «учитель», «підтримка», «персонал», «мова», «громада». Ці терміни характерні для професій, де важлива саме робота з людьми, а такі професії, як правило, автоматизуються найменше.

Третій список «manual\_words» – це ручна праця. У ньому зібрані слова на кшталт «ремонт», «зварювання», «будівництво», «укладання», «токарь», «водій», «оператор», «інструмент». Тут я орієнтувалась на типові описання робітничих професій.

І окремо винесла творчий «creative\_words» набір слів: «дизайн», «журналіст», «фотограф», «контент», «актори», «реклама», «копірайтинг», «архітектор», «продюсер». Такі слова зустрічаються в професіях, де багато творчих або медійних задач.

Ці словники допомагають краще розпізнати професію за її сенсовим навантаженням, а не роблять розподіл ґрунтуючись на власний опір. Якщо розглядати професії без таких словників, то до однієї групи можуть відноситись спеціальності, які різні за галуззю, проте в описі мають схожий мотив.

Проте важливою перевіркою є можливість кореляції між утвореними словниками у моделі, що продемонстровано на рис. 2.2. Адже при неналежній побудові словника, може виникати накладання слів і відповідно словники між собою корелюватимуть.

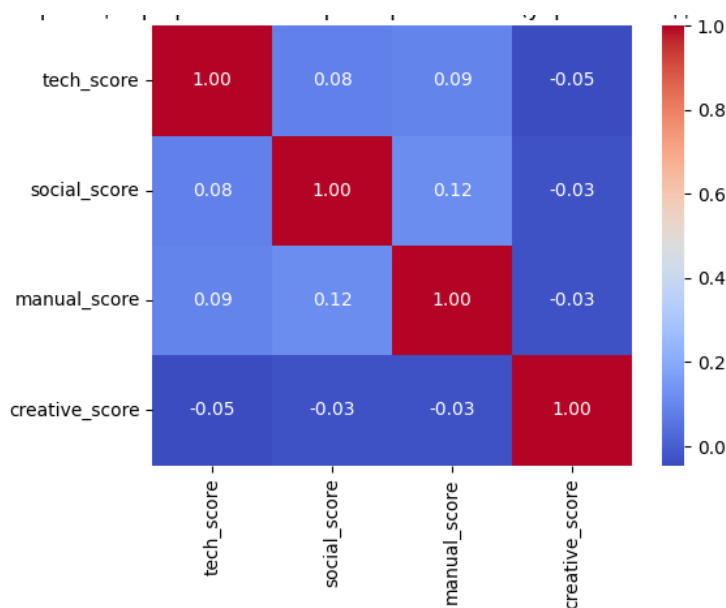


Рисунок 2.2 – Кореляція між професійними характеристиками

*Джерело: розроблено автором*

Матриця кореляцій між словниками різних груп вказує на відсутність значимих лінійних зв'язків. Усі коефіцієнти кореляції знаходяться в межах від – 0.05 до 0.12, що вказує на майже повну незалежність спеціалізованих навичок. Це означає, що професії, які вимагають високого рівня технічних компетенцій, не обов'язково характеризуються високими або низькими показниками соціальних, креативних чи ручних навичок і навпаки.

Навички, які оцінюються через текстові описи професій, формують окремі, майже не пересічні виміри. Кожна група навичок відображає автономний аспект професійної діяльності, що є важливим для подальшого моделювання ризику автоматизації: професія може бути технічно складною, але не креативною; або висококреативною, але з мінімальними соціальними чи ручними компонентами. Відсутність мультиколінеарності також свідчить, що використані індикатори навичок не дублюють одне одного та можуть безпечно застосовуватися в моделі без ризику перекручування оцінок.

## **2.2 Моделювання ризику автоматизації на основі SBERT-ембедінгів**

У дослідженні використано п'ять моделей машинного навчання: лінійну регресію, Ridge-регресію, RandomForest, XGBoost та багат шарову нейронну мережу (MLP), оскільки вони представляють різні підходи до моделювання залежності між текстовими та інженерними ознаками професій і ризиком їх автоматизації. Лінійна регресія обрана як базовий інтерпретований метод, що дозволяє оцінити лінійні взаємозв'язки між ознаками. Ridge-регресія доповнює її, оскільки включає L2-регуляризацію, що зменшує вплив мультиколінеарності у високовимірних даних (зокрема, у SBERT-ембедінгах) і забезпечує більш стабільні оцінки коефіцієнтів. RandomForest застосовано як ансамблевую модель, здатну відтворювати складні нелінійні залежності та демонструвати стійкість до шуму. XGBoost, як сучасна високопродуктивна реалізація градієнтного бустингу, дозволяє досягати високої точності завдяки послідовному коригуванню помилок попередніх дерев і вбудованій регуляризації. Модель MLP включена для оцінки ефективності глибоких нейронних архітектур, здатних автоматично виявляти складні нелінійні патерни у векторних представленнях тексту. Використання цих моделей забезпечує методологічну повноту експерименту, дозволяючи порівняти різні підходи, результати яких наведені у таблиці 2.2 та на рис. 2.3, – від простих

лінійних до ансамблевих та нейронних і вибрати найбільш точний та стабільний алгоритм прогнозування.

Таблиця 2.2 – Порівняння результатів навчання моделей

	CV_R <sup>2</sup> _mean	CV_R <sup>2</sup> _std	R <sup>2</sup>	MAE	MSE	Spearman
XGBoost	0.337735	0.064538	0.451899	0.234572	0.078510	0.674665
RandomForest	0.339897	0.033094	0.404573	0.254438	0.085289	0.660541
Ridge	-0.333935	0.075922	0.128102	0.278468	0.124891	0.554280
MLP	-0.316500	0.169245	-0.200378	0.324839	0.171942	0.442401
LinearRegression	-3.218422	1.070813	-0.433880	0.366819	0.205389	0.415100

Джерело: розроблено автором

Порівняльна оцінка моделей демонструє суттєву різницю в здатності прогнозувати рівень автоматизації на основі SBERT-ембедінгів. Найкращий результат отримано для XGBoost, який забезпечив найвище значення коефіцієнту детермінації, найменшу абсолютну та середньоквадратичні похибки та найвищу рангову кореляцію Спірмена. Подібну, хоча дещо нижчу якість продемонстрував Random Forest, що вказує на ефективність ансамблевих деревоподібних методів у роботі з високорозмірним семантичним простором. Натомість лінійні моделі показали низькі або негативні значення коефіцієнту детермінації, що свідчить про їхню неспроможність коректно моделювати нелінійні залежності у векторних представленнях SBERT. Модель MLP також не сформувала стійкого узагальнення, отримавши негативне значення R<sup>2</sup>, що вказує на перенавчання.

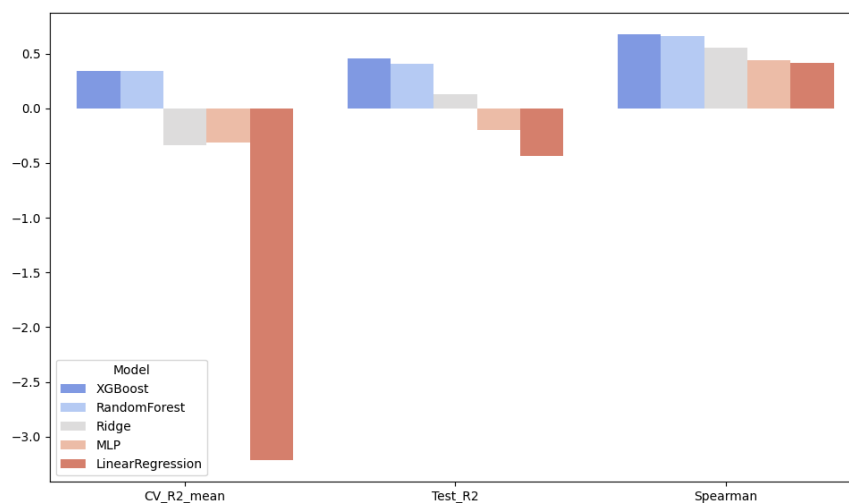


Рисунок 2.3 – Порівняння моделей за якістю прогнозу ризику автоматизації

Джерело: розроблено автором

Графічне представлення результатів підтверджує поведінку моделей та дозволяє візуально оцінити їхню стабільність. На відміну від лінійних та нейромережових алгоритмів, ансамблеві дерева демонструють не лише високі значення  $R^2$ , але й узгодженість між якістю в крос-валідації та на тестовій вибірці, що свідчить про їхню здатність уникати перенавчання. Особливо показовою є метрика Спірмена: XGBoost і Random Forest зберігають високу рангову кореляцію, що вказує на їхню здатність відтворювати структурний порядок професій за рівнем автоматизованості, незалежно від абсолютної похибки прогнозу. Натомість MLP та лінійні моделі демонструють розрив між крос-валідаційними та тестовими оцінками, що видно на графіку як сильне падіння результатів при переході до реальних даних.

Проте якщо аналізувати всі показники, то можна вважати, що коефіцієнт детермінації низький для всіх моделей. Варто врахувати той факт, що акцент робиться на дослідженні професій згідно їхньої назви та опису. Дослідження «The Acceptable R-Square in Empirical Modelling for Social Science Research» Dr Peterson K Ozili (2023) [53] доводить, що в соціальних та поведінкових науках прийнятні значення  $R^2$  значно нижчі, ніж у технічних дисциплінах, і діапазон 0,20 – 0,50 вважається нормальним для моделей, що працюють з багатовимірними та слабкоструктурованими даними. У задачах, де результат формується під впливом великої кількості прихованих факторів – як-от текстові описи професій та складні прояви автоматизації – низький або середній рівень коефіцієнту детермінації є очікуваним. Тому в контексті цього дослідження основну увагу приділено не абсолютному рівню  $R^2$ , а стабільності моделі та здатності коректно відтворювати рангову структуру ризику автоматизації.

Навчання моделі проводилось безпосередньо на кращій моделі XGBoost. Візуальне представлення результатів продемонстровано на рис 2.4

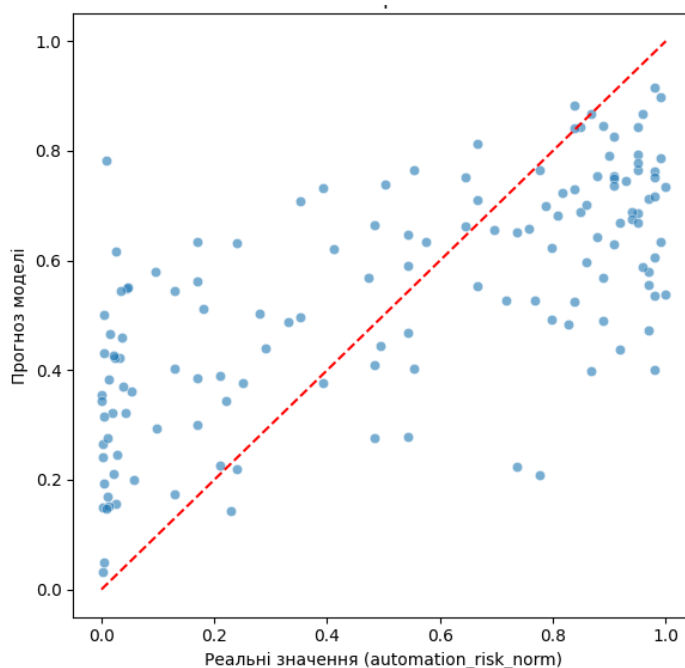


Рисунок 2.4 – Порівняння фактичних значень ризику автоматизації з прогнозами моделі XGBoost

*Джерело: розроблено автором*

Зображення демонструє, що модель формує загалом коректну тенденцію: професії з високим реальним ризиком здебільшого отримують і високі прогнозовані значення, що підтверджує здатність XGBoost відтворювати глобальну структуру залежності. Водночас помітним є певне розсіювання точок навколо діагоналі, особливо у середньому діапазоні значень, що вказує на існування локальних відхилень і відображає складність моделювання контекстних ембедінгів SBERT.

Найважливішим є те, що модель правильно розрізняє низькі та високі рівні автоматизованості: для професій із мінімальним ризиком передбачені значення також концентруються в нижній частині графіка, тоді як професії з ризиком, близьким до 1.0, формують щільні кластери у верхньому правому секторі. Це узгоджується з високою ранговою кореляцією Спірмена, яку показує XGBoost, та підтверджує його здатність адекватно відтворювати відносний порядок професій за ризиком автоматизації навіть за наявності певних похибок у точкових прогнозах.

## 2.3 Крос-лінгвальне вирівнювання професій та структурний аналіз

У цій роботі я використовую підхід transfer learning, оскільки для України просто не існує офіційних даних про те, які професії мають високий чи низький ризик автоматизації. Такі оцінки є лише для США, зокрема в рамках досліджень Oxford та BLS. Через це я не можу навчити модель безпосередньо на українських даних – у них немає цільової змінної, яку модель мала б передбачати.

Тому спочатку я беру американський датасет, де для кожної професії є і текстовий опис, і значення ризику автоматизації. Ці тексти я перетворюю на вектори за допомогою SBERT, щоб модель могла читати їх не як звичайний текст, а як набір смислових ознак. На цьому етапі XGBoost вчиться помічати взаємозв'язок між змістом опису професії та тим, наскільки вона схильна до автоматизації.

Потім, уже на другому етапі, я застосовую цю модель до українського списку професій. Українські описи також пропускаю через SBERT, і це важливий момент: SBERT є багатомовною моделлю, тому українські й англійські описи опиняються в одному й тому самому семантичному просторі. Для обґрунтування перенесення ризику автоматизації з англійського Oxford Dataset на українські професії було виконано семантичне вирівнювання текстових описів. На основі моделі Sentence-BERT (paraphrase-multilingual-MiniLM-L12-v2) згенеровано ембедінги для українського та англійського переліку професій, після чого обчислено косинусну подібність між усіма парами записів. Отримана матриця подібностей на рис. 2.5 дозволила визначити найближчі англійські відповідники для кожної української професії (top-5), а також оцінити загальну якість крос-лінгвального вирівнювання.

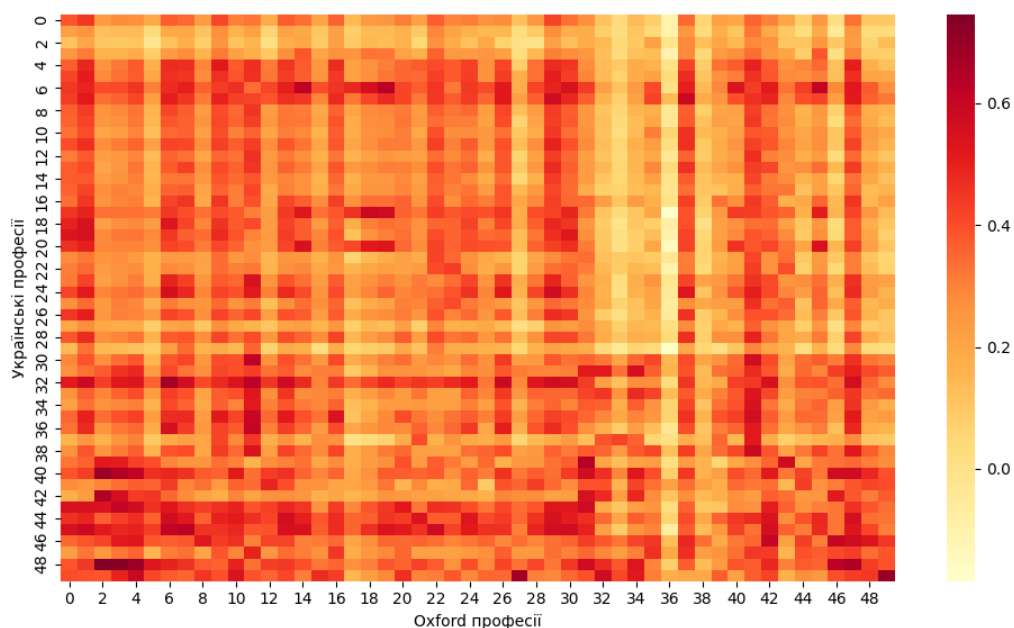


Рисунок 2.5 – Семантична подібність між українськими та іноземними професіями

*Джерело: розроблено автором*

Теплова карта фрагмента матриці (50×50) демонструє чіткі області підвищеної подібності, що підтверджує коректне групування споріднених професій різними мовами. Додатково були розраховані два узагальнені показники наведені у таблиці 2.3: середня максимальна схожість по всіх українських професіях та частка записів, для яких значення подібності мають високі значення. Обидва показники свідчать про достатню якість семантичної відповідності між UKR і OXFORD.

Таблиця 2.3 – Семантична подібність професій

Код	Українська назва професії	Англійська назва професії	Частка подібності	Ризик автоматизації
6636	Помічник ветеринара	veterinary assistants and laboratory animal ca...	0.915779	0.868314
1214	Асистент фізичного терапевта	physical therapist assistants	0.905323	0.015397
8160	Технік лісового господарства	forest and conservation technicians	0.877640	0.422609
3840	Лаборант (медицина)	medical and clinical laboratory technicians	0.875100	0.473258
8236	Технік-технолог (електротехніка)	electrical and electronics engineering technic...	0.866023	0.848055
4688	Машиніст установки з оброблення транспортних з...	cleaners of vehicles and equipment	0.863567	0.371961

## Продовження таблиці 2.3

1	2	3	4	5
1757	Вихователь	education administrators preschool and childca...	0.856468	0.012358
99	Інженер з охорони та захисту лісу	forest fire inspectors and prevention specialists	0.855634	0.045786
4073	Лікар-хірург	physicians and surgeons	0.855326	0.001418
1173	Асистент вчителя спеціальної освіти	teacher assistants	0.855003	0.564425

*Джерело: розроблено автором*

Модель формує спільний семантичний простір для двох мов, що дозволяє коректно зіставляти професії та переносити оцінки ризику автоматизації. У більшості випадків вирівнювання є достатньо надійним для подальшого моделювання.

Для кращого розуміння якості моделі провела перевірку чи дійсно назва професії у поєднанні із описом краще за оцінку лише назви професії. Було сформовано два варіанти вхідних даних: (1) лише назва професії та (2) назва разом із коротким описом. Для обох варіантів згенеровано контекстуальні ембедінги за допомогою Sentence-BERT (SBERT), що забезпечує семантично осмислені представлення речень, придатні для порівняння за косинусною подібністю. Результати аналізу наведені у таблиці 2.4

Таблиця 2.4 – Співставлення моделі із назвою та назвою+описом

Модель	R <sup>2</sup>	MAE	MSE	Spearman
Назва	0.563954	0.062280	0.006278	0.711033
Назва + Опис	0.732329	0.049103	0.003854	0.828256

*Джерело: розроблено автором*

Використання розширеного формату тексту (назва + опис) стабільно покращує всі показники моделі, особливо ранговий коефіцієнт Спірмена, що свідчить про більш точне узгодження прогнозів з фактичним порядком професій. Це підтверджує, що додатковий контекст значно збагачує семантичне представлення професії, особливо у випадках багатозначних або узагальнених назв.

Завдяки цьому модель може коректно інтерпретувати українські тексти, навіть якщо вона тренувалася на англійських. А згодом уже натренована на американських даних модель прогнозує рівень автоматизації для українських

професій. Це дає можливість використовувати знання з глобального ринку праці для аналізу ситуації в Україні, де подібних оцінок поки що немає. На рис. 2.6 представлений графічний результат розподілу можливості автоматизації професій.

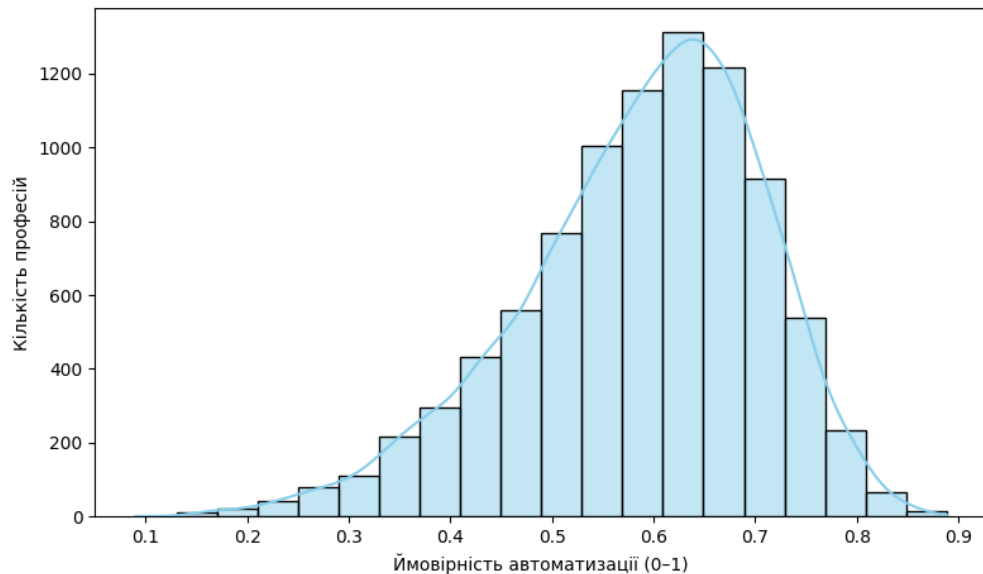


Рисунок 2.6 – Гістограма розподілу прогнозованого ризику автоматизації українських професій

*Джерело: розроблено автором*

Гістограма показує, що розподіл має форму, близьку до нормального, зі зміщенням у бік вищих значень. Пік розподілу припадає приблизно на область 0.55–0.65, що свідчить: найбільша частка українських професій має середньо-високий рівень ризику автоматизації. Швидше за все значна частина професій у майбутньому може автоматизуватись частково, а можливо і повністю. Для кращого розуміння чи відсутні викиди або нетипові значення результатів варто проаналізувати статистику утвореної змінної наведеної у таблиці 2.5

Таблиця 2.5 – Описова статистика нової змінної

	predicted_risk
count	8986.000000
mean	0.587075
std	0.119036
min	0.090087
25%	0.515143
50%	0.602664
75%	0.672996
max	0.888647

*Джерело: розроблено автором*

Середнє значення показника рівня автоматизації дещо вище 0,5, показник не надто перевищує середину. Можна вважати, що чіткої тенденції, що всі професії підпадають під автоматизацію або ж навпаки жодна професія не може бути автоматизованою не спостерігається. Стандартне відхилення дорівнює 0,119, що вказує на помірну варіативність оцінок. Мінімальні та максимальні значення входять у досліджуваний та адекватний проміжок, а отже будь-які нетипові значення відсутні. Межі кватилів вказують на певну асиметрію, а саме переважання показників і більшу сторону, тобто все ж таки оцінка високого рівня автоматизації переважає. Модель формує диференційовані прогнози, розподіл `predicted_risk` можна вважати інформативним для подальшої інтерпретації та прийняття рішень.

На даному етапі можна чітко побачити які професії підпадають під найвищий та найнижчий ризик автоматизації. Цей крок є вагомий у формуванні роботи, адже розуміння специфіки автоматизації дасть краще розуміння впливу штучного інтелекту на ринок праці. На рис. 2.7 наведені професії які можуть бути автоматизовані, а на рис. 2.8 ті спеціальності, які на даний момент часу не піддаються заміні.



Рисунок 2.7 – Топ-20 професій із найвищим ризиком автоматизації

Джерело: розроблено автором

Результати прогнозу свідчать, що найбільший ризик зосереджений у технічних, виробничих та рутинних професіях, де значна частина роботи пов'язана з механічними процесами, повторюваними операціями та використанням обладнання. Переважна більшість спеціалізацій у верхній частині рейтингу – це саме технічні ролі зі стандартними, чітко формалізованими завданнями. Така структура ризику вказує на високу ймовірність подальшої автоматизації та роботизації цих видів діяльності, оскільки їхні функції часто можуть бути відтворені машинами або роботизованими системами.



Рисунок 2.8 – Топ-20 професій із найнижчим ризиком автоматизації

*Джерело: розроблено автором*

Спеціалізації із найнижчим рівнем автоматизації належать до освітніх, педагогічних та культурно-просвітницьких професій. На відміну від технічних або виробничих спеціальностей, де значну частину завдань можна формалізувати й автоматизувати, робота педагогів, викладачів, наставників, тренерів та керівників навчальних закладів ґрунтується на міжособистісній взаємодії, творчості, адаптації до індивідуальних потреб учнів та прийнятті рішень у непередбачуваних ситуаціях.

Для аналізу структури простору професій було виконано зниження розмірності SBERT-ембедінгів до двох головних компонент методом PCA. На отриманому розсіювальному графіку, що наведений на рис. 2.9 кожна точка відповідає одній професії, а колір відображає прогнозований рівень ризику автоматизації.

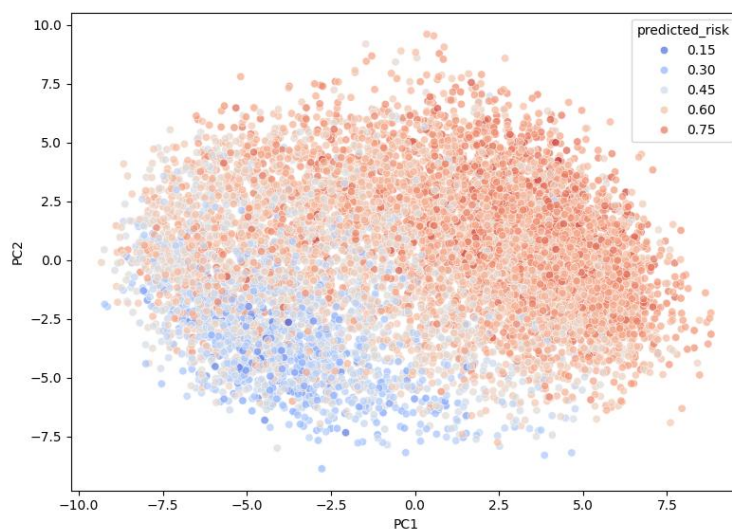


Рисунок 2.9 – Професії у 2D-просторі ембедінгів

*Джерело: розроблено автором*

Візуалізація демонструє, що професії формують неперервний семантичний простір, без чітких меж чи кластерів, що є типовим для контекстних мовних моделей. При цьому просторовий розподіл ризику має виражену структуру: професії з високим ризиком (червоний колір) концентруються переважно у правій частині графіка, тоді як низькоризикові (синій колір) – у лівій нижній зоні. Така закономірність свідчить про те, що SBERT-ембедінги коректно відображають змістовні відмінності текстових описів професій і їхню відповідність до ступеня автоматизації. Плавний перехід кольорів від синього до червоного свідчить про поступову зміну рівнів ризику автоматизації. Професії не формують окремих, чітко відмежованих груп, а розташовуються у вигляді безперервного спектра значень. Така поведінка узгоджується з тим, що професійні описи часто мають подібні або частково перетинні характеристики, що відображається у структурі їхніх векторних представлень.

Для виявлення структурних груп професій за характером їх сприйнятливості до автоматизації було проведено кластеризацію за допомогою алгоритму KMeans ( $k = 3$ ), яка продемонстрована на рис.2.10. Кластеризація виконувалася не на вихідних ембедінгах, а на компактному наборі інтерпретованих ознак, що відображають зміст та функціональну специфіку професій. До моделі кластеризації були включені такі показники: (1) прогнозований рівень ризику

автоматизації (*predicted\_risk*), (2) індекс технічної складової (*tech\_score*), (3) індекс соціально-комунікативної складової (*social\_score*), (4) індекс ручної/фізичної складової (*manual\_score*) та (5) індекс творчо-креативної складової (*creative\_score*).

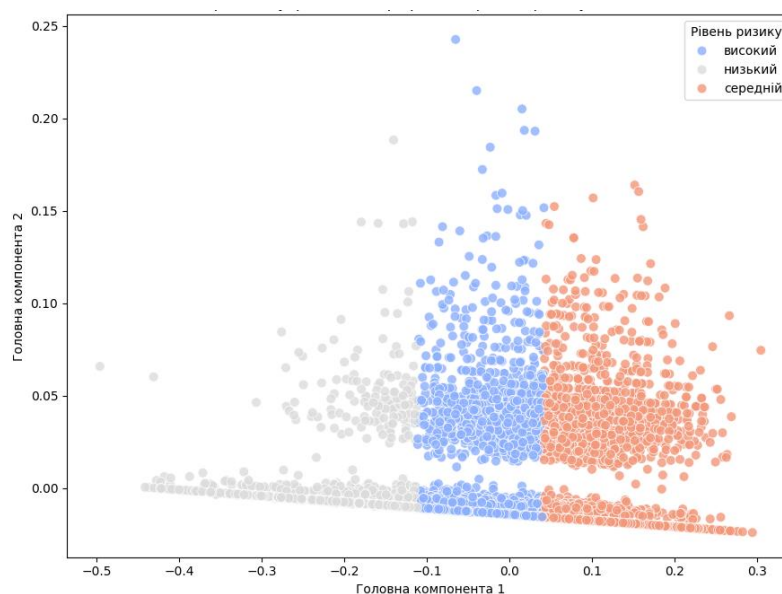


Рисунок 2.10 – Кластеризація українських професій за рівнем ризику автоматизації

Джерело: розроблено автором

Застосування PCA для двовимірної візуалізації показало наявність трьох чітко відокремлених груп професій. Характеристики кожного кластера були визначені на основі обчислення середніх значень показників *predicted\_risk*, *tech\_score*, *social\_score*, *manual\_score* та *creative\_score* для кожної групи, що дозволило встановити домінуючі ознаки в межах кожного кластера. Перший кластер об'єднує професії з низьким ризиком автоматизації, для яких характерні високі *social\_score* та *creative\_score*. Другий кластер охоплює професії зі змішаними характеристиками та середнім рівнем ризику. Третій кластер складається з професій, для яких притаманні високі значення *manual\_score* та підвищений *predicted\_risk*. Структура кластерів є логічно обґрунтованою, інтерпретованою та узгоджується з очікуваними тенденціями щодо впливу автоматизації на ринок праці, що підтверджує валідність проведеного групування.

Після побудови моделі прогнозування ймовірності автоматизації українських професій було виконано перевірку її робастності за трьома напрямками:

bootstrap-аналіз на рис. 2.11, абляційне тестування рис. 2.12 та перевірка стабільності при зміні random seed.

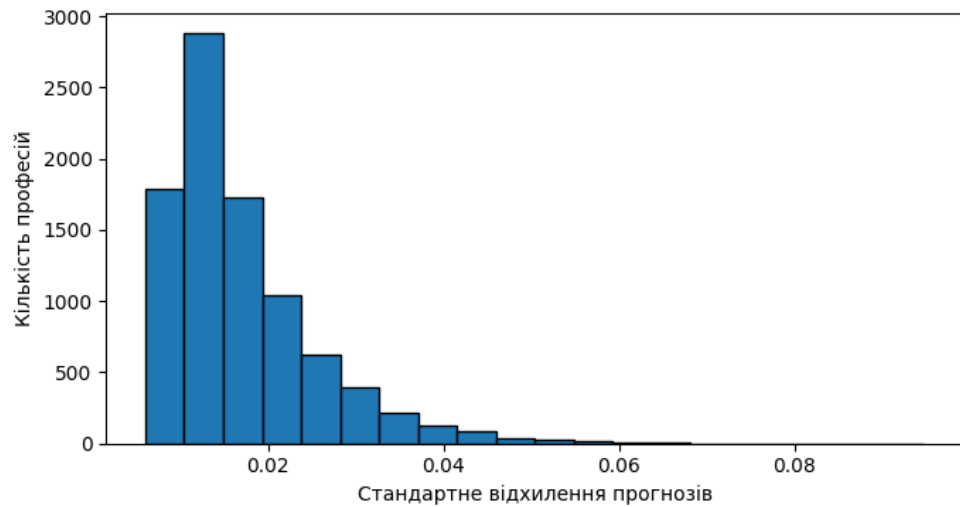


Рисунок 2.11 – Стабільність прогнозів (bootstrap-аналіз)

*Джерело: розроблено автором*

Bootstrap-аналіз реалізовано шляхом багаторазового формування вибірок із поверненням на основі вихідного набору даних та повторного навчання моделі XGBoost. Для кожної професії обчислено стандартне відхилення предикції на основі результатів усіх ітерацій. Розподіл значень стандартного відхилення сконцентрований поблизу нуля, що свідчить про низьку дисперсію прогнозів. Це означає, що модель демонструє стабільність результатів незалежно від конкретного складу підвибірки.

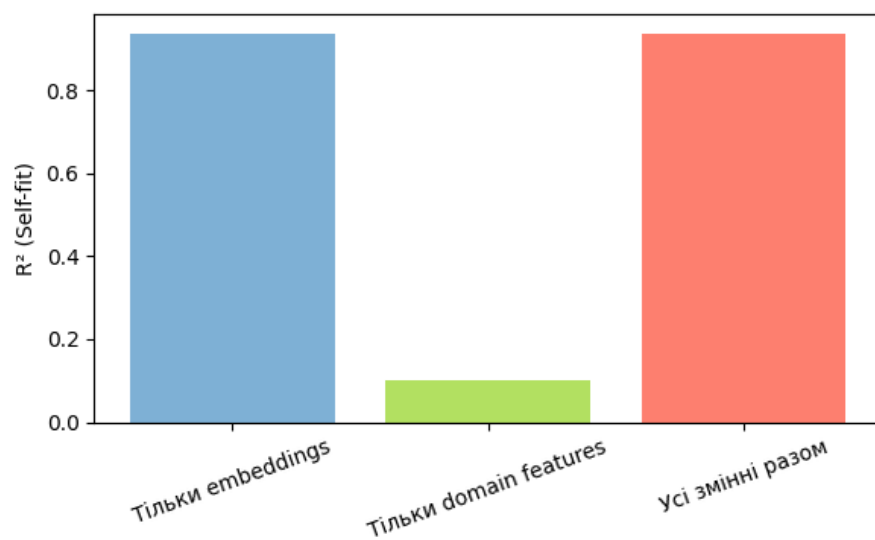


Рисунок 2.12 – Ablation Study

*Джерело: розроблено автором*

Проведено абляційний аналіз, у межах якого протестовано три варіанти вхідних ознак:

1. лише семантичні ембедінги професій,
2. лише доменні лінгвістичні характеристики (tech, social, manual, creative),
3. комбінація ембедінгів та доменних ознак.

Порівняння за метрикою  $R^2$  показало, що комбінований варіант має найвищу пояснювальну здатність, тоді як використання лише лінгвістичних індикаторів дає найнижчий результат. Це підтверджує, що модель ефективно інтегрує як семантичну інформацію, так і предметно-специфічні характеристики професій.

Таблиця 2.6 – Стійкість  $R^2$  до різних seed'ів

Seed	$R^2$
1	0.934996
7	0.934551
42	0.936162
101	0.935866
2024	0.936315

*Джерело: розроблено автором*

Окрім цього виконано аналіз стабільності до значення параметра random seed. Модель навчалася з кількома значеннями ініціалізації, після чого обчислювався коефіцієнт детермінації. Отримані значення  $R^2$  практично не відрізняються між собою, що вказує на низьку залежність результатів від стохастичних факторів навчання.

Модель прогнозування є робастною, а значення ризику автоматизації українських професій не є випадковими або нестійкими. Модель має достатній рівень стабільності для подальшого аналітичного аналізу й інтерпретації отриманих висновків. Абляційний аналіз підтвердив важливість комплексного представлення професій, де поєднання ембедінгів і доменних лінгвістичних ознак забезпечує найповнішу характеристику трудових функцій. Перевірка стабільності до різних random seed додатково довела, що модель не схильна до коливань результатів через випадковість ініціалізації, що є важливою умовою для її практичного застосування.

У фінальному блоці було реалізовано комплексну процедуру визначення професій, які потенційно можуть бути замінені системами штучного інтелекту. На початковому етапі з повного набору українських професій виділено підмножину спеціальностей із найвищими значеннями прогнозованого ризику автоматизації. До цієї групи потрапили всі професії, для яких попередня модель на основі XGBoost оцінила ймовірність автоматизації на рівні 0,6 або вище. Зоорієнтовую увагу на спеціальностях, які вже демонструють високий рівень вразливості відповідно до побудованої моделі. Формується вибірку професій, які з найбільшою ймовірністю потребуватимуть адаптації, перекваліфікації або трансформації функціоналу в умовах поширення ШІ.

Для кращої інтерпретації функціонального змісту професій були розроблені п'ять еталонних груп професій, що відображають основні типи діяльності:

- когнітивні (аналітичні) професії (*cognitive\_seed*), характерні для рутинної обробки інформації та офісних функцій;
- креативні професії (*creative\_seed*), що включають генерацію контенту, дизайн та творчі завдання;
- механічні (ручні) професії (*mechanical\_seed*), пов'язані з фізичною працею, операціями та технічними маніпуляціями;
- клієнтські (обслуговуючі) професії (*customer\_seed*), орієнтовані на безпосередню взаємодію з клієнтом;
- стратегічні професії (*strategic\_seed*), що потребують складного аналізу, експертності й прийняття рішень з високим рівнем відповідальності.

Кожна високоризикова професія була порівняна із зазначеними групами на основі косинусної схожості ембедінгів. Результатом стали кількісні індикатори семантичної подібності, які відображають близькість професії до когнітивного, креативного, механічного, клієнтського чи стратегічного типу діяльності.

Здійснена побудована окремої моделі логістичної регресії, навченої на *seed*-наборах, яка визначає ймовірність віднесення професії до когнітивного типу. Цей індикатор використовується разом із семантичною подібністю для формування інтегральної оцінки.

Оцінка потенційної заміни професій технологіями штучного інтелекту здійснювалась за допомогою інтегрального показника *AI Replacement Score*, який поєднує декілька незалежних джерел інформації та враховує як структурні, так і семантичні характеристики професій. Формування цього показника ґрунтувалося на припущенні, що заміність професій визначається не лише загальним рівнем ризику автоматизації, але й змістовою природою виконуваних функцій. Тому базовим компонентом інтегрального індексу став нормалізований прогнозований ризик автоматизації, отриманий на попередньому етапі дослідження. Разом із тим, для уточнення цього прогнозу було враховано семантичну подібність професій до еталонних груп, які представляють когнітивні, креативні, клієнтські, механічні та стратегічні типи діяльності.

Професії, які за своїм змістом ближчі до когнітивних та креативних видів діяльності, отримували вищі значення *AI Replacement Score*, що відображає їхню більшу потенційну заміність алгоритмічними системами. Натомість подібність до механічних, клієнтських та стратегічних типів діяльності знижувала інтегральний показник, оскільки такі види робіт традиційно вимагають фізичної присутності, міжособистісної взаємодії або глибокої експертної участі людини.

Крім того, додаткову вагу отримував показник, сформований за допомогою класифікаційної моделі логістичної регресії, що дозволив оцінити ймовірність належності професії до когнітивної групи. Комбінування цих компонентів дало змогу створити збалансований та інтерпретований індикатор, який відображає реальну структуру трудових функцій та потенційний рівень їх автоматизації. Після обчислення всі значення індексу були нормалізовані, що забезпечило їх порівнянність між професіями та можливість побудови чіткої шкали заміності, що продемонстрована на рис. 2.13

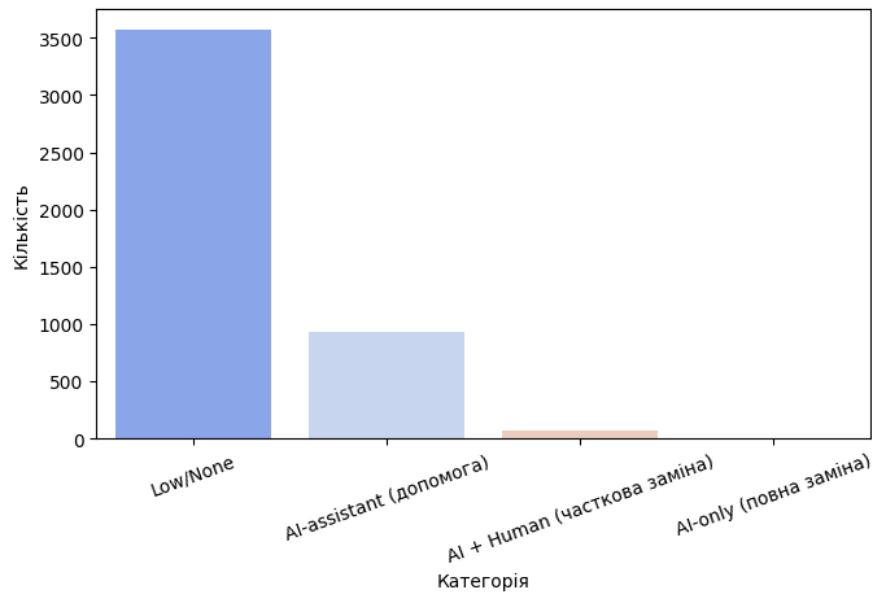


Рисунок 2.13 – Розподіл професій за рівнем заміни ШІ

*Джерело: розроблено автором*

У результаті було виокремлено чотири категорії, що відображають ступінь участі ШІ у виконанні відповідних трудових функцій. Професії з найвищими значеннями *AI Replacement Score* були віднесені до категорії повної заміни, яка означає, що значна частина або й усі характерні для цих професій завдання можуть бути виконані системами штучного інтелекту без безпосереднього залучення людини. У категорії часткової заміни опинилися професії, де ймовірно є гібридна модель «ШІ + людина», коли алгоритми здатні автоматизувати значну частину рутинних або структурованих операцій, однак контроль, прийняття складних рішень чи взаємодія з клієнтом залишаються у компетенції працівника. Професії з помірними значеннями індексу були віднесені до групи підтримки (AI-assistant), у якій штучний інтелект здатний підсилювати продуктивність працівника, надаючи рекомендації, автоматизуючи вторинні операції або допомагаючи в аналізі інформації, проте не може виконувати базові функції професії повністю. До категорії низької або відсутньої заміненості увійшли професії зі структурними ознаками, що роблять їхнє автоматизоване виконання малоімовірним у середньостроковій перспективі – це переважно ролі, які вимагають фізичної взаємодії, високоінтелектуальних стратегічних рішень або складної

міжособистісної комунікації. Більш наочна поведінку професій у групах представлена на рис. 2.14

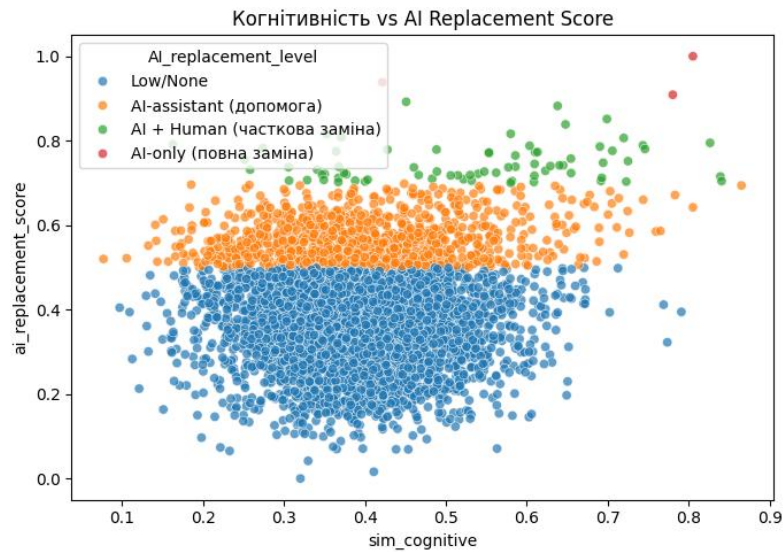


Рисунок 2.14 – Залежність індексу заміщення (AI Replacement Score) від рівня когнітивності професії

*Джерело: розроблено автором*

На графіку зображено залежність між показником семантичної когнітивної подібності професій (*sim\_cognitive*) та інтегральним індексом їх потенційної заміненості штучним інтелектом (*AI Replacement Score*). Спостерігається монотонна тенденція: зі зростанням значення *sim\_cognitive* збільшуються значення *AI Replacement Score*, оскільки сучасні системи штучного інтелекту найбільш ефективно автоматизують саме рутинні когнітивні, аналітичні та текстові завдання. У нижній частині діапазону заміненості (0–0.5) переважають професії з низькою когнітивною подібністю, що віднесено до категорії Low/None – це більш механічні професії які замінити важко та дорого. У середньому діапазоні (0.5–0.7) спостерігається концентрація професій категорії AI-assistant, а саме тих де ШІ виступає у ролі асистента, тобто допомагає, а не виконує всю роботу. У вищому діапазоні (0.7–0.9) розташовані переважно професії категорії AI + Human, де ШІ приймає на себе головні обов’язки, а людина лише відстежує, та контролює, щоб не була допущена помилка. Максимальні значення індексу (0.9–1.0) відповідають професіям, класифікованим як AI-only, чіткий перелік професій виведений у табл.2.7.

Таблиця 2.7 – Розподіл професій відповідно ризику заміни ІІІ

Назва	Рівень ризику	Частка когнітивності	Частка креативності	Частка обслуговування	Механічна частка	Стратегічна частка	Рівень заміни	Клас заміни
Офісний службовець (цінні папери)	0.794141	0.805242	0.406869	0.547168	0.416837	0.497982	1.000000	AI-only (повна заміна)
Апаратник виробництва рентгенівських екранів	0.881836	0.421276	0.314342	0.179398	0.366160	0.346689	0.938487	AI-only (повна заміна)
Офісний службовець (реєстрація та облік)	0.797733	0.780371	0.400370	0.608380	0.423100	0.569300	0.908849	AI-only (повна заміна)
Гравер шрифту	0.738896	0.450882	0.414663	0.296663	0.359662	0.236007	0.892208	AI + Human (часткова заміна)
Касир (на підприємстві, в установі, організації)	0.799181	0.638051	0.399214	0.535312	0.385595	0.519142	0.882395	AI + Human (часткова заміна)
Офісний службовець (кредит)	0.802957	0.698993	0.302787	0.542205	0.400102	0.594150	0.851536	AI + Human (часткова заміна)
Обліковець з реєстрації бухгалтерських даних	0.728837	0.647844	0.337077	0.497012	0.336787	0.520420	0.838634	AI + Human (часткова заміна)
Оцінювач (експертна оцінка майна)	0.720122	0.580059	0.351199	0.406084	0.237041	0.571345	0.816416	AI + Human (часткова заміна)
Апаратник одержання сполук для рентгенівських	0.793906	0.352845	0.329921	0.151621	0.363139	0.292117	0.815227	AI + Human (часткова заміна)
Розкатник-сортувальник паперу	0.741236	0.370815	0.331258	0.306706	0.346005	0.316630	0.808126	AI + Human (часткова заміна)
Касир-операціоніст	0.838308	0.719785	0.390743	0.574688	0.504554	0.543153	0.806449	AI + Human (часткова заміна)

Джерело: розроблено автором

Повноцінно модель виводить три професій, які може замінити ІІІ, для інших відсоток заміни значно знижується і модель не може стверджувати, що вони

також будуть замінені. Перша професія офісний службовець (цінні папери) та третя офісний службовець (реєстрація та облік) вони є дещо спорідненими, оскільки їхня діяльність пов'язана з виконанням рутинних, стандартизованих та чітко регламентованих операцій із великими масивами даних. Штучний інтелект здатний повністю автоматизувати ключові завдання цих професій: зчитування та обробку документів за допомогою OCR та NLP, верифікацію інформації, внесення й оновлення даних у реєстрах, формування звітності та виявлення помилок. Алгоритми забезпечують високу швидкість, точність і безперервність процесів, що значно перевищує людські можливості. Через низьку креативність і повторюваність операцій ці професії стають одними з найбільш вразливих до повної заміни цифровими системами та інтелектуальними автоматизованими платформами.

Проте до таких професій потрапив Апаратник виробництва рентгенівських екранів. Штучний інтелект може замінити професію апаратника завдяки можливості автоматично керувати технологічними процесами, забезпечувати контроль якості за допомогою комп'ютерного зору, оптимізувати параметри виробництва в реальному часі, а також прогнозувати поломки обладнання. Більшість операцій є рутинними, стандартизованими й легко формалізуються, що робить їх повністю придатними для автоматизації.

Всі подальші професії не підпадають до повноцінного ризику, тобто ШІ може виконувати роль асистента, допомагати у рутинних завданнях, що не потребують креативності та інноваційності проте повноцінно завершувати завдання на даний період дослідження ні.

Однак структурні зміни зайнятості – це лише один вимір трансформацій ринку праці. Поряд із технологічними змінами зберігається й інша важлива характеристика ринку – нерівність у доступі до винагороди за працю, зокрема гендерна різниця в оплаті. Навіть за умов активного впровадження ШІ, яке могло б сприяти більшій об'єктивності оцінювання роботи, питання рівної оплати праці не зникає. Навпаки, ці зміни роблять ще більш актуальним аналіз того, наскільки ринок праці є справедливим щодо жінок і чоловіків.

## РОЗДІЛ 3

### ОЦІНЮВАННЯ ГЕНДЕРНИХ ДИСПРОПОРЦІЙ В ОПЛАТІ ПРАЦІ ІЗ ЗАСТОСУВАННЯМ LOGIB МОДЕЛІ

#### **3.1 Алгоритм оцінювання гендерної нерівності в оплаті праці на основі моделі Logib**

Поняття еквівалентності не обмежується подібною трудовою діяльністю, а також може стосуватися абсолютно різних професій [54]. Із розумінням наявності нерівності у оплаті праці між гендерними групами допоможе модель Logib.

Logib – це швейцарський онлайн-інструмент вільного доступу, розроблений Федеральним офісом з питань гендерної рівності для аналізу заробітної плати та сприяння рівній оплаті праці між чоловіками та жінками. Він допомагає компаніям оцінювати свої практики оплати праці, виявляти гендерні відмінності [55].

Logib, стандартний інструмент Швейцарської Конфедерації для аналізу рівності оплати праці між жінками та чоловіками, складається з двох модулів. Обидва модулі розроблені як системи самооцінювання, які користувачі можуть виконувати самостійно, використовуючи надані документи, без необхідності спеціальних знань. Вони доступні всім роботодавцям безкоштовно як сучасні онлайн-застосунки.

Модуль 1 Logib був розроблений Федеральним управлінням зі сприяння гендерній рівності (FOGE: Federal Office for Gender Equality) [56] за підтримки приватних спеціалізованих інституцій на початку 2000-х років (Strub, 2004) [57]. Версія Logib Module 1 у форматі Excel доступна безкоштовно з 2004 року. Модуль 1 було підтверджено шляхом різних оцінок (INFRAS, 2011 [58]; Felfe, Trageser & Iten, 2015 [59]) та перетворено на онлайн-застосунок, який був запуснений 1 липня 2020 року.

Теоретичний підхід Модуля 1 базується на теорії людського капіталу (Becker, 1993) [60]. Використовується статистичний метод – багатофакторний регресійний аналіз, щоб оцінити, якою мірою об'єктивні та недискримінаційні індивідуальні та робочі (тобто пов'язані з посадою) характеристики визначають рівень заробітної плати та чи впливає на нього стать працівника. З огляду на статистичні вимоги цей модуль особливо підходить для великих компаній.

Модуль 2 Logib був розроблений за наукової підтримки Університету Берна та широко випробуваний у польових тестах протягом 2011–2019 років, щоб зробити стандартний аналітичний інструмент Швейцарської Конфедерації доступним також і для малих підприємств. Під малими підприємствами мається на увазі компанії, у яких працює від 2 до 49 осіб. У таких організаціях обсяг даних є надто малим для надійного статистичного аналізу, тому для них застосовується аналітичний підхід Модуля 2, здатний коректно врахувати структуру персоналу та специфіку малих колективів. Після завершення розроблення методологія пройшла наукову експертизу зовнішніми фахівцями з різних галузей (економіка, трудові науки, право), а також експертами Міжнародної організації праці (МОП) [61]. Результати підтвердили її наукову обґрунтованість та відповідність правовим вимогам.

Модуль 2 базується на науковому методі оцінювання робіт і може бути використаний для визначення того, в якій мірі вимоги та характеристики посади, а також індивідуальний досвід, освіта та професійне навчання впливають на заробітну плату, і чи існують при цьому відмінності між чоловіками та жінками.

Обидва модулі стандартної аналітичної моделі використовують однакові особистісні дані працівників та дані щодо заробітної плати. Дані, пов'язані з функціями та робочими завданнями, фіксуються по-різному в кожному модулі відповідно до його методологічного підходу. І метод OLS-регресії, який використовується в Модулі 1, і метод оцінювання робіт, що лежить в основі Модуля 2, були визнані Федеральним Верховним судом Швейцарії як юридично прийнятні методи для встановлення наявності чи відсутності дискримінації в оплаті праці [62].

Дана модель має ряд переваг:

- Програма є повністю безкоштовною. Будь-який роботодавець може розпочати перевірку у зручному для себе форматі, тобто як для особистої перевірки, так і для формування юридичного висновку.
- Інструмент побудований так, щоб бути доступним навіть без глибоких статистичних знань: первинний аналіз може зробити сама компанія. Програма на базі Excel, яка може швидко надати уявлення про поточну ситуацію [63].
- Завдяки статистичному моделюванню Logib може виявити ту частину різниці в оплаті праці, яка залишається після врахування об'єктивних факторів, відбувається так звана ідентифікація «невиправданих» гендерних розривів – це дозволяє говорити про потенційну дискримінацію [64].
- Logib визнаний EPIC (“Equal Pay International Coalition”) як «Good Practice» інструмент [65].
- Коли компанія публічно демонструє, що використовує Logib, це підсилює репутацію як справедливого роботодавця, що може позитивно вплинути на мотивацію персоналу та імідж. (наприклад, так компанія зменшує ризик звинувачень у дискримінації)
- Технічні методи, параметри аналізу та інструкції Logib публічно доступні кількома мовами (німецькою, французькою, італійською, англійською), що підвищує довіру та доступність для більшої групи населення.
- Повна конфіденційність, адже аналіз цих даних виконується на федеральному сервері. Статистична інформація передаються туди в зашифрованому вигляді, що унеможлиблює їх відстеження до окремих компаній. Після завершення аналізу дані видаляються із сервера. Результати аналізу можуть переглядати лише самі користувачі. Результати аналізу не зберігаються на сервері, їх можна завантажити у вигляді документів PDF та Excel. Дані не передаються третім особам, що знижує ризик витоку чутливої інформації.

- Позитивний досвід великих компаній, наприклад швейцарської пошти «Swiss Post» показує, що за допомогою Logib вони змогли значно зменшити «невиправдані» розриви в оплаті [66].
- Сприяє законодавчій відповідності, адже у Швейцарії компанії з більш ніж 100 співробітниками зобов'язані провести аналіз рівності оплати («Equal Pay Analysis») відповідно до змін у Законі про рівність [67].

Безпосередньо в Україні немає чіткого закону, що зосереджено вказує на необхідність дотримання гендерного балансу на підприємствах. Logib узгоджується з європейськими директивами та рекомендаціями Міжнародної організації праці, тому її впровадження сприяє наближенню України до практик Європейського союзу у сфері соціальної рівності.

Збір таких даних у одному ексель файлі необхідний для того, щоб можна було розпочати аналіз. Безпосередньо ми заходимо на сторінку «Equal pay analysis with Logib» [61] після чого обираємо модуль відповідно розміру нашої компанії та отримуємо шаблон файлу на рис. 1.1 до заповнення.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Q	R	S	T	U	V	W
	Key (personal number, ID, etc.)	Age (date of birth, year of birth, age)	Gender	Years of service (date of entry, year of entry, number of years of service)	Education and training	Function (job carried out)	Level of requirements (relevant for module 1)	Professional position (relevant for module 1)	Employees paid monthly: Activity rate in %;	Employees paid hourly: number of paid hours during the reference month	Basic wage (according to activity rate)	Allwages (Shift, Sunday and night work plus other extra pay for difficult working conditions, 1/12 of annual total)	13th monthly wage (generally 1/12 of basic wage)	Special payments (1/12 of annual total)	Usual weekly working hours in company (fulltime employees, i.e. activity rate = 100%)	Statistical population	Remarks	Additional information1	Additional information2	Additional information3	Additional information4	Additional information5
1																						
2	1	55	F	31	1	Principal	1	1	100	0	11282	3384,5	940,13	0								
3	2	45	M	22	1	Vice Principal (Academic)	1	1	100	0	10718	3215,3	893,13	0								
4	3	48	M	26	1	Vice Principal (Vocational)	1	1	100	0	10718	3215,3	893,13	0								
5	4	38	F	16	1	Vice Principal (Pastoral & Educational)	1	1	100	0	10153	2030,7	846,12	0								
6	5	48	M	25	1	Head of Mechanical Engineering	2	2	100	0	7978,3	2393,5	664,86	0								
7	6	51	F	24	1	Head of Electrical Engineering	2	2	100	0	7978,3	2792,4	664,86	0								
8	7	34	M	11	1	Head of Agronomy & Agriculture	2	2	100	0	7978,3	797,83	664,86	0								
9	8	56	F	35	1	Head of Technology & Agriculture	2	2	100	0	7978,3	2393,5	664,86	0								
10	9	38	F	15	1	Placement and Apprenticeship	2	2	100	0	7450,3	2235,1	620,86	0								
11	10	46	F	34	1	Extracurricular Activities C	3	3	100	0	6396,5	639,65	533,04	0								
12	11	46	M	25	1	Head of Part-time & Distance Education	2	2	100	0	7978,3	2393,5	664,86	0								
13	12	44	F	15	1	Curriculum and Quality Assurance	2	2	100	0	7978,3	2393,5	664,86	0								
14	13	56	M	35	1	Head of Physical Education	3	3	100	0	7978,3	3988,6	664,86	0								
15	14	54	F	29	1	Student Mentor	3	4	100	0	6396,5	639,65	533,04	0								

Рисунок 1.1 – Приклад шаблону для оформлення даних моделі Logib

Джерело: розроблено автором

Стовпці А–Е містять персональні дані робітників, а саме:

- Колонка А містить унікальний ключ або ідентифікатор для кожного працівника.
- У колонці В – вік. Може записуватись як кількість років співробітнику, або вказувати рік народження у форматі 4 цифр, або повну дату народження. Це

залежить від вказаних налаштувань при завантаженні файлу. Вік, який введений, має відповідати референтному року.

- Колонка С відповідає за стать. Можна використовувати будь-яке зручне кодування, важливо його вказати при завантаженні заповненого файлу у систему. У власному випадку працювала із німецькою версією сайту, тому для чоловіків вказувала М (Mann) , а для жінок F (Frau).
- У колонці D вказувався стаж роботи працівника на досліджуваному підприємстві.
- Колонка E характеризувався рівнем освіти. Запропонована шкала від найвищою до найнижчого ступеня (від 1 до 8):
  - 1 – Магістерський ступінь і вище
  - 2 – Бакалаврський ступінь
  - 3 – Професійна освіта підвищеного рівня диплом майстра/техніка, диплом технічного училища
  - 4 – Педагогічний диплом різних рівнів (колишні педагогічні коледжі або еквівалент)
  - 5 – Академічна, професійна або спеціалізована середня освіта
  - 6 – Завершена професійна освіта і навчання
  - 7 – Професійне внутрішнє навчання від компанії, яке не визначається на державному рівні, але його достатньо для роботи на певній посаді
  - 8 – Обов’язкова шкільна освіта без професійної кваліфікації

Наступна категорія колонок (стовпці F–H) пов’язані із професійними посадами:

- Колонка F містить інформацію про конкретні види робіт («посади»), які можна відокремити від осіб, що їх виконують. Посади зі схожими завданнями та відповідальністю об’єднуються в одну функцію. Варіації додаткових завдань між людьми усередині функції ігноруються.
- У колонці G вказується рівень вимоги до роботи за кожною функцією. Класифікація може базуватися на профілі посади, описі функцій та вимогах до навичок. Від 1 (найвищі вимоги) до 4 (без особливих вимог):

- 1 – діяльність з найскладнішими завданнями прийняття рішень; дуже складні та вимогливі завдання
- 2 – діяльність із досить складними технічними або практичними завданнями; потрібна висока самостійність
- 3 – діяльність, що потребує професійних або спеціалізованих знань
- 4 – діяльність, що складається з простих ручних або рутинних завдань
- Колонка Н характеризує професійну позицію, у якому вводиться рівень відповідальності, пов'язаний із функцією: керівна відповідальність, відповідальність спеціаліста чи відповідальність за людське життя. Наводиться діапазон від 1 (найвища відповідальність) до 5 (без додаткової відповідальності):
  - 1 – Найвище керівництво
  - 2 – Середній менеджмент
  - 3 – Нижчий менеджмент
  - 4 – Найнижчий менеджмент
  - 5 – Співробітники без керівних функцій / функції без додаткової відповідальності

Колонки І та J наводять відсоток зайнятості працівників. Проте заповнення одночасно двох немає сенсу.

- Стовець І містить відсоткове позначення зайнятості
- Стовець J вказує на зайнятість працівників у годинах
- Стовці К–N містять інформацію про суму заробітної плати
- У колонці К вказується чиста заробітна плата без доплат та преміювальних
- Колонка L відповідає за доплати, різні надбавки
- У колонці M прописується 13-та зарплата, але вноситься як одну дванадцятую суми.
- Колонка N описує річні бонуси, премії тощо – також у вигляді однієї дванадцятої річної суми.

- Стовпець O вказує на особливості робочих годин. Тобто дана колонка актуальна для тих працівників, які працюють інакшу кількість годин від вказаної. Якщо таких немає, тоді ця колонка залишається пустою.
- Стовпець Q не є обов'язковим до заповнення, якщо на підприємстві немає особливих умов контрактів. Наприклад:
  - 1 – «Звичайний» контракт. Порожнє поле автоматично вважається 1.
 Для особливих контрактів:
  - 2 – Учні (Apprenticeship)
  - 3 – Стажери (Traineeship)
  - 4 – Експати (постійне місце роботи за кордоном, контракт зі швейцарським підрозділом)
  - 5 – Інші (напр., отримувачі інвалідності з обмеженою продуктивністю, погодинники без годин у референтному місяці)
- Стовпець R також не є обов'язковим до заповнення, містить примітки щодо окремих працівників.
- Стовпці S–W: Додаткові дані компанії (клас зарплат, місце роботи тощо) можуть залишатися порожніми.

Після формування такого файлу він завантажується у систему де розпочинається побудова регресійної моделі. Важливе уточнення особливостей, а саме кількість робочих годин на тиждень, календарний термін дослідження, і згодом розпочинається аналіз.

### **3.2 Оцінювання гендерних диспропорцій для навчального закладу в сільській місцевості**

Перший досліджуваний заклад освіти знаходиться у селищі із кількістю населення близько 1100 осіб. Даний освітній заклад готує фахівців у галузі сільського господарства, зокрема з агрономії та механізації, поєднуючи теорію з

практикою. Кількість працівників становить 214 особи, тому у цьому випадку буде застосований Модуль 1 [68].

Після завантаження даних система зорієнтовувати увагу на працівниках, вік яких більше 70. Офіційно в Україні немає обмежень щодо роботи людей старшого віку, тому для даного випадку це не є помилкою, просто система виявила це нетиповим та підозрілим. Далі заповнивши додаткову інформацію та погодившись з умовами проведення розпочала аналіз, після якого перше, що було виведено – це умовний спідометр на рис. 3.1.

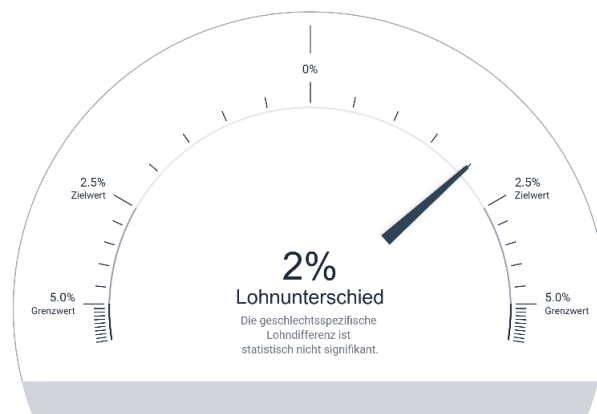


Рисунок 3.1 - Результат оцінки гендерної нерівності у першому навчальному закладі

*Джерело: реалізовано за допомогою моделі Logib [61]*

Ліва сторона графіку характеризує чоловічу стать, а права – жіночу. Так як стрілка переважає до правої сторони – це може вказувати, що жінки заробляють на 2% менше за чоловіків. Проте, враховуючи вказану інструкцію моделі, будь-яке відхилення до 5% відсотків не вважається значущим.

Після чого модель виводить повноцінно статистику даних, відповідається оцінка всіх вхідних значень. Тобто виділяється кількість осіб які увійшли у вибірку, якої статі, із якими посадами та заробітними платами. Візуально демонструється різниця між кількістю працівників із різними рівнями вимог на рис. 3.2 та професійними посадами на рис. 3.3 при цьому враховуючи стать.

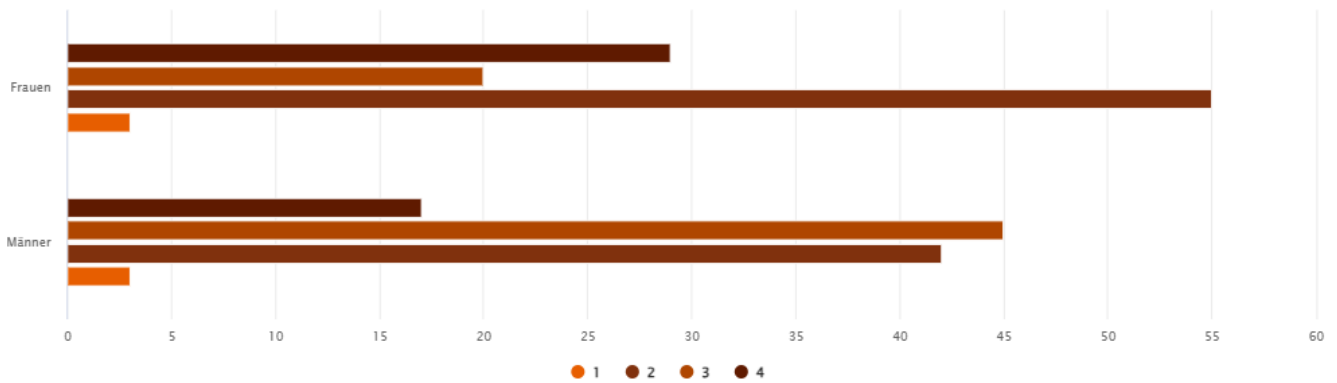


Рисунок 3.2 – Діаграма рівнів вимог за статтю

*Джерело: реалізовано за допомогою моделі Logib [61]*

Із діаграми чітко видно домінацію 2 рівня відповідальності, що вказує на переважання працівників жіночої статі, які виконують складні технічні або практичні завдання. Більшість чоловіків у свою чергу виконують завдання, які вимагають професійних або спеціалізованих знань.

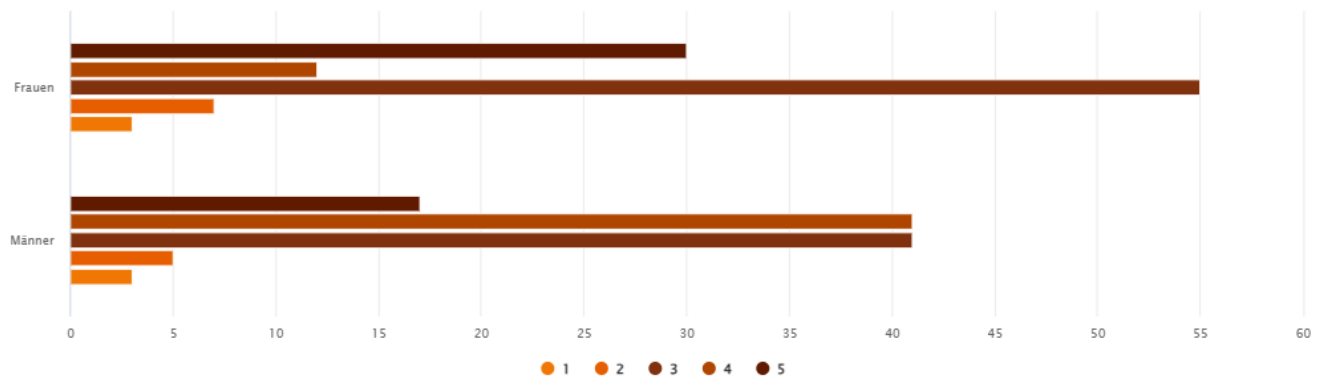


Рисунок 3.3 – Діаграма професійних посад за статтю

*Джерело: реалізовано за допомогою моделі Logib [61]*

На даному рисунку значно переважає кількістю жінок, які займають посаду нижчого рівня управління. Посади, що виконують безпосередній контроль за виконанням поточних завдань та організацію роботи конкретних груп або команд. У чоловіків рівна кількість працівників на різних категоріях. Тобто спостерігається баланс у кількості осіб які займають посади нижчого та середнього рівнів управління.

Модель безпосередньо запускає регресійний аналіз завдяки якому виводяться межі кватилів, що продемонстровано у табл. 3.1 та візуально зображено на рис.3.4. Сам результат регресії наводиться у табл. 3.2.

Таблиця 3.1 – Результати регресійного аналізу першого навчального закладу

	Q1		Q2		Q3		Q4		Разом	
	до 4 396		до 10 814		до 12 807		до 15 606			
	Кількіс ть	Част ка	Кількіс ть	Част ка	Кількіс ть	Част ка	Кількіс ть	Част ка	Кількіс ть	Част ка
<b>Чолов іки</b>	26	52.0 %	29	52.7 %	16	34.0 %	36	58.1 %	107	50.0 %
<b>Жінки</b>	24	48.0 %	26	47.3 %	31	66.0 %	26	41.9 %	107	50.0 %
<b>Разом</b>	50	100.0 %	55	100.0 %	47	100.0 %	62	100.0 %	214	100.0 %

Джерело: реалізовано за допомогою моделі Logit [61]

Аналіз розподілу стандартизованої заробітної плати за квантилями показує, що найбільша відносна частка жінок зосереджена у третьому квантілі (66,0%), тоді як у першому, другому та четвертому квантілях частка чоловіків переважає і становить відповідно 52,0%, 52,7% та 58,1%. При цьому у загальній вибірці кількість жінок і чоловіків є однаковою (по 50,0%).

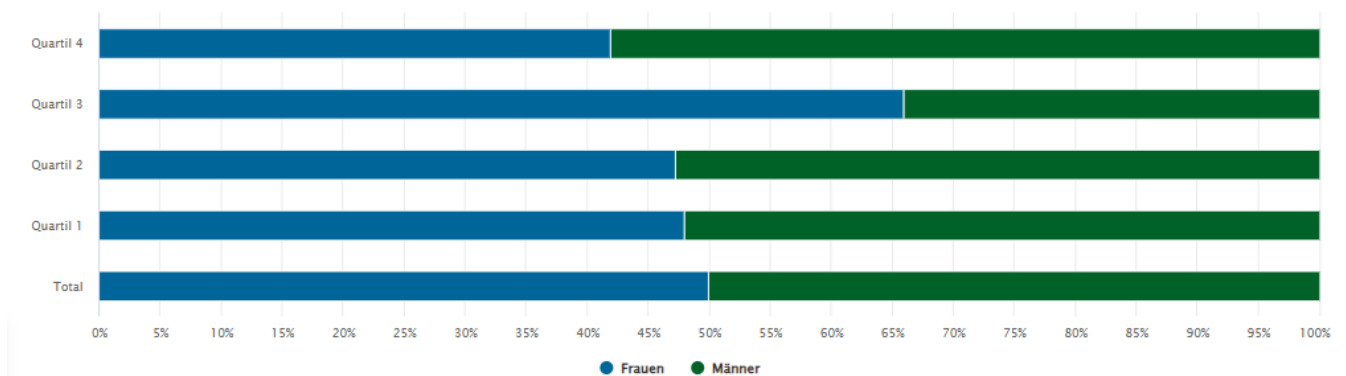


Рисунок 3.4 – Поділ на квантілі для аналізу першого навчального закладу

Джерело: реалізовано за допомогою моделі Logit [61]

Отриманий результат свідчить не про загальний гендерний перекис у фонді оплати праці, а про різну концентрацію жінок і чоловіків у межах окремих інтервалів заробітної плати, що є структурною особливістю внутрішнього розподілу доходу в межах установи. Це підтвердило поведінку показану на рис.3.1 та 3.2. Сам результат регресійного аналізу продемонстрований у табл. 3.2

Таблиця 3.2 – Результати регресійного аналізу першого навчального закладу

Коефіцієнт при змінній «стать» (гендерний коефіцієнт):	-0,019
Стандартна помилка гендерного коефіцієнта:	0,037
Кількість ступенів свободи:	202
Коефіцієнт детермінації ( $R^2$ ):	0,803
Нульова гіпотеза ( $H_0$ ): різниця в оплаті праці становить 0% Альтернативна гіпотеза ( $H_1$ ): різниця в оплаті праці відрізняється від 0% Критичне значення t-статистики ( $\alpha = 5\%$ , двосторонній тест, $df = 202$ ):	1,972
Обчислене значення t-статистики:	0,519
Статистична значущість	ні ( $p > 0,05$ )

*Джерело: реалізовано за допомогою моделі Logit [61]*

Результати регресійного аналізу засвідчили відсутність статистично значущого гендерного розриву в оплаті праці в межах досліджуваного навчального закладу за умови контролю ключових пояснювальних змінних. Оцінений коефіцієнт при змінній «стать» становить  $-0,019$ , що вказує на незначну та практично нейтральну різницю в рівнях оплати праці між чоловіками та жінками після урахування кваліфікаційних, посадових та організаційних характеристик. Обчислене значення t-статистики ( $0,519$ ) є меншим за критичне значення ( $1,972$ ) при рівні значущості  $5\%$ , що не дозволяє відхилити нульову гіпотезу про відсутність різниці в заробітній платі за статтю ( $p > 0,05$ ). Водночас високий коефіцієнт детермінації ( $R^2 = 0,803$ ) свідчить про те, що обрана модель добре описує варіацію заробітної плати за рахунок включених факторів, а отже, підтверджує, що саме структурні та функціональні характеристики посади, а не стать працівника, відіграють визначальну роль у формуванні рівня оплати праці в даній установі. Щоб краще зрозуміти роль змінних варто звернути увагу на рис. 3.5

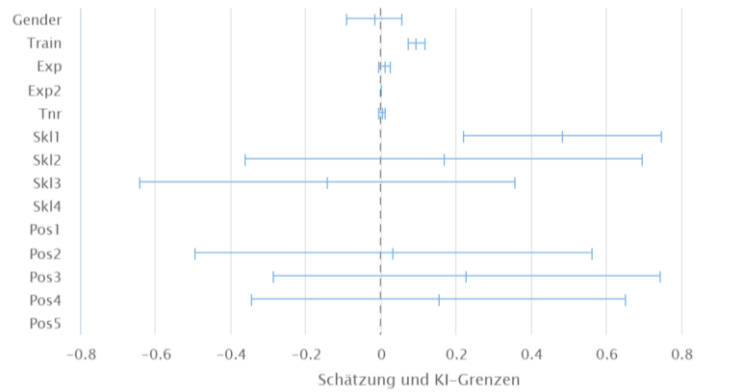


Рисунок 3.5 – Довірчі інтервали моделі 1

Джерело: реалізовано за допомогою моделі Logib [61]

Довірчий інтервал для змінної «стать» (Gender) перетинає нульову позначку, що свідчить про відсутність статистично значущого впливу гендерної приналежності на рівень оплати праці. Даний результат повністю узгоджується з попередніми кількісними оцінками ( $t = 0,519$  при критичному значенні  $1,972$ ;  $p > 0,05$ ), підтверджуючи неможливість відхилення нульової гіпотези щодо відсутності гендерного розриву в оплаті праці.

Натомість професійні, кваліфікаційні та посадові характеристики демонструють суттєвішу роль у формуванні заробітної плати. Зокрема, змінні, які відображають рівень освіти та підвищення кваліфікації (Train), професійний досвід (Exp), стаж роботи в установі (Tnr), а також рівень кваліфікації (Skl1–Skl4) та посадовий статус (Pos1 – Pos5), справляють більш помітний вплив на зміну рівня оплати праці. Включення квадрата досвіду ( $Exp^2$ ) дозволяє врахувати нелінійність цього зв'язку, що проявляється в ефекті спадної граничної віддачі.

### 3.3 Оцінювання гендерних диспропорцій для навчального закладу в місті

Другий досліджуваний заклад освіти знаходиться у місті із кількістю населення приблизно 244 тисячі осіб. Даний навчальний заклад пропонує здобути фахову передвищу освіту в галузі інформаційних технологій, готуючи фахівців за

відповідними спеціальностями. Кількість працівників становить 113 осіб, тому у цьому випадку буде застосований Модуль 1.

Заповнивши додаткову інформацію та погодившись з умовами проведення розпочала аналіз, після якого перше, що було виведено – це умовний спідометр на рис. 3.6.



Рисунок 3.6 – Результат оцінки гендерної нерівності у другому навчальному закладі

*Джерело: реалізовано за допомогою моделі Logib [61]*

У даній навчальному закладі жінки заробляють на 12,5% більше, ніж чоловіки. Існуючі відмінності в заробітній платі не можна пояснити, або принаймні не виключно, взаємодією різних об'єктивних факторів, таких як вік, освіта та професійна посада. Система вказує, що з цією цифрою заклад не відповідає вимогам у різних контекстах застосування, зокрема умовам Закону про гендерну рівність та вимогам щодо участі, що стосуються положень про рівне ставлення до жінок і чоловіків щодо рівної оплати праці у федеральних закупівлях. Тому у даному випадку відразу проаналізую результати регресії у табл.3.3

Таблиця 3.3 – Результати регресійного аналізу другого навчального закладу

Коефіцієнт при змінній «стать» (гендерний коефіцієнт):	0,119
Стандартна помилка гендерного коефіцієнта:	0,047
Кількість ступенів свободи:	101
Коефіцієнт детермінації (R <sup>2</sup> ):	0,737
Нульова гіпотеза (H <sub>0</sub> ): різниця в оплаті праці становить 0%	
Альтернативна гіпотеза (H <sub>1</sub> ): різниця в оплаті праці відрізняється від 0%	1,984
Критичне значення t-статистики (α = 5%, двосторонній тест, df = 202):	

Обчислене значення t-статистики:	2.527
Статистична значущість	так

*Джерело: реалізовано за допомогою моделі Logib [61]*

Після урахування всіх факторів – віку, освіти, посади, стажу, обсягу роботи та керівних обов’язків – жінки в середньому заробляють на 11,9% більше, ніж чоловіки (гендерний коефіцієнт = 0,119). Стандартна помилка цієї оцінки становить 0,047, що дозволяє з достатньою точністю стверджувати про наявність різниці. Обчислене значення t-статистики дорівнює 2,527, що перевищує критичне значення 1,984 для двостороннього тесту на рівні значущості 5%, тому різниця є статистично значущою і не випадковою. Коефіцієнт детермінації  $R^2 = 0,737$  свідчить про те, що модель добре пояснює варіацію зарплат на підприємстві. Результати регресійного аналізу однозначно вказують на наявність статистично значущої скоригованої гендерної різниці на користь жінок, яка не може бути пояснена включеними в модель об’єктивними характеристиками. Однак сам факт значущості коефіцієнта ще не розкриває повної картини надійності цієї оцінки.

Для детальнішого розуміння стабільності отриманого ефекту доцільно проаналізувати довірчі інтервали, які показують діапазон можливих значень гендерного коефіцієнта за умов статистичної невизначеності на рис. 3.7.

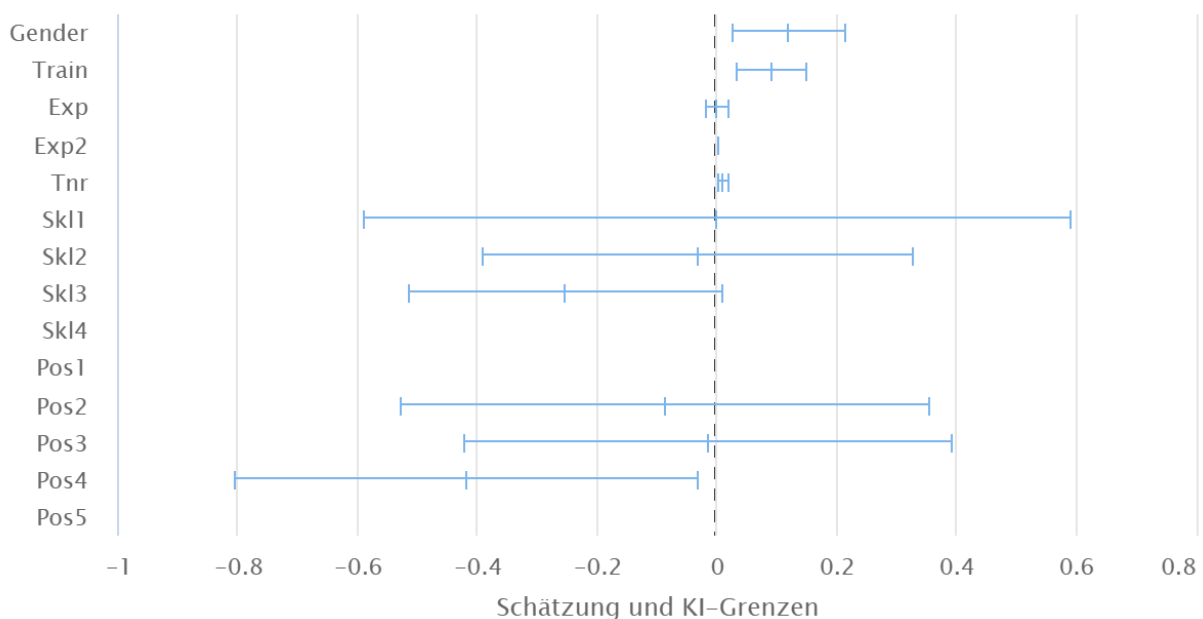


Рисунок 3.7 – Довірчі інтервали моделі 2

*Джерело: реалізовано за допомогою моделі Logib [61]*

Лише дві змінні *Gender* та *Train* мають статистично значущий вплив на заробітну плату. Їхні інтервали повністю лежать праворуч від нульової позначки, що вказує на достовірний позитивний ефект статі та проходження навчань. Натомість для змінних *Exp*, *Exp<sup>2</sup>* та *Tnr* інтервали охоплюють нуль, що свідчить про відсутність статистично визначеного впливу стажу та досвіду роботи. Кваліфікаційні (*Sk11–Sk14*) та посадові змінні (*Pos1–Pos5*) характеризуються широкими інтервалами, які повністю або частково перетинають нуль, тому їхній вплив на рівень заробітної плати не можна вважати статистично значущим. Таким чином, серед усіх предикторів моделі підтверджено значущість лише статі та проходження навчання.

Можливою причиною спостережуваної різниці у зарплаті є те, що чоловіки частіше займають нижчі посади порівняно з жінками, що автоматично відображається на їхньому рівні доходу. Щоб перевірити це припущення, доцільно проаналізувати розподіл співробітників за рівнем посад і функцій, що подано на рис. 3.8 та рис. 3.9.

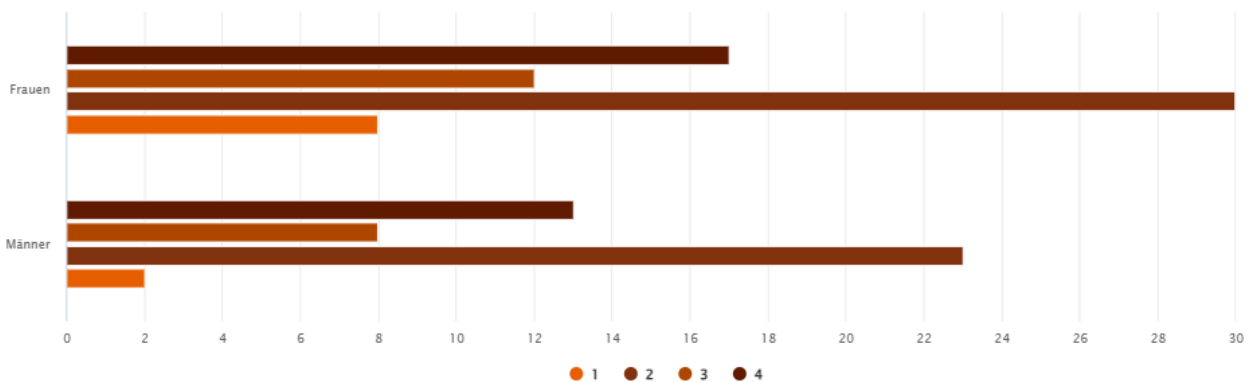


Рисунок 3.8 – Діаграма рівнів вимог за статтю у моделі два

Джерело: реалізовано за допомогою моделі Logib [61]

У найскладніших завданнях, які вимагають прийняття важких рішень жінки представлені майже частіше. У другій групі складних технічних та практичних завдань повторюється перевага жінок. У категоріях, де рівень складності нижчий (потреба базових навичок або рутинні дії), знову немає радикальних розривів, але жінки частіше трапляються у групі простих завдань. Структура завдань для чоловіків і жінок не симетрична – жінки помітно частіше виконують завдання верхнього рівня складності.

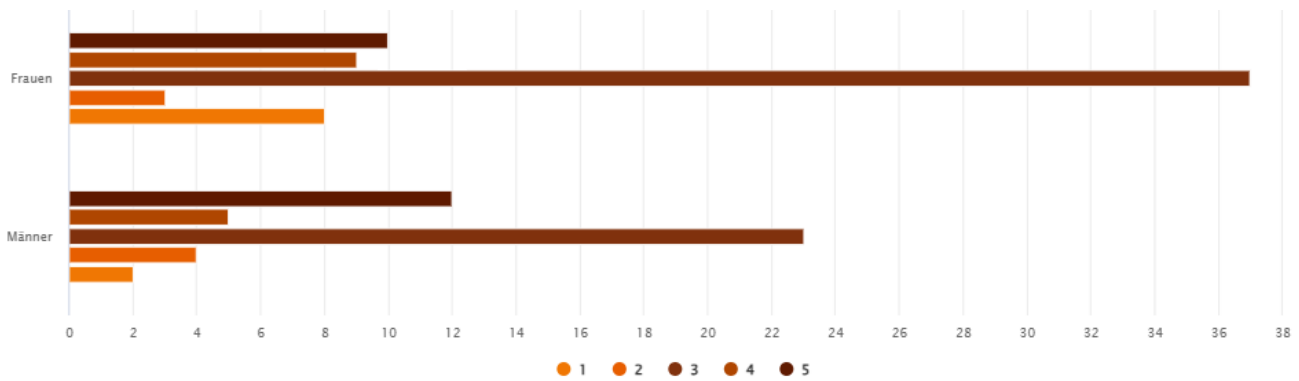


Рисунок 3.9 – Діаграма професійних посад за статтю у моделі два

*Джерело: реалізовано за допомогою моделі Logit [61]*

На даному графіку спостерігається лідируюча позиція жінок на керівних посадах, а на середньому рівня чоловіки переважають із незначним розривом. Але загалом кількість жінок у закладі є вищою, відповідно жінки в середньому займають трохи вищі позиції, ніж чоловіки у цій вибірці. Це і може бути поясненням чому відбувається дисбаланс у оплаті праці.

Порівняння двох навчальних закладів, розташованих у сільській та міській місцевості, не виявило системного зв'язку між територіальним розташуванням та рівнем гендерного розриву в оплаті праці. У сільському закладі скоригована різниця в оплаті праці є статистично незначущою, а модель показує, що стать не впливає на рівень заробітної плати після врахування посадових, кваліфікаційних та організаційних характеристик. Натомість у міському закладі зафіксовано статистично значущу перевагу в оплаті праці на користь жінок, яка не пояснюється структурою досвіду, віку чи стажу, хоча частково може бути пов'язана з їхньою більшою представленістю на складніших завданнях та керівних посадах. Місцевість сама по собі не є визначальним фактором гендерних відмінностей у заробітках. Виявлені розбіжності пов'язані насамперед зі структурою персоналу, посадовою ієрархією та розподілом рівнів вимог у кожній конкретній установі, а не з тим, чи знаходиться заклад у місті або в сільській місцевості.

## ВИСНОВКИ

Проведений аналіз показав, що вплив ІІІ на ринок праці залежить не стільки від самих алгоритмів, скільки від соціально-економічних механізмів адаптації. Рівень безробіття, пов'язаний із поширенням ІІІ, визначається швидкістю перекваліфікації працівників, готовністю освітніх інституцій оновлювати програми, а також державними політиками у сфері зайнятості та технологічного розвитку. Якщо ці елементи узгоджені, то ІІІ може виступати не джерелом руйнування, а засобом модернізації ринку праці та підвищення якості людського капіталу.

Дослідження показало, що ринок праці перебуває під потужним впливом технологічних зрушень. Професії суттєво відрізняються за рівнем ризику автоматизації: рутинні, технічні та адміністративні спеціальності виявляються найбільш вразливими, тоді як професії, що вимагають міжособистісної взаємодії, творчості чи педагогічних навичок, демонструють нижчий ризик заміщення автоматизованими системами. Розроблена модель дозволяє кількісно оцінювати технологічну вразливість професій і слугує основою для прогнозування майбутніх структурних змін у зайнятості.

Оцінка впливу нових технологій спирається не лише на аналіз ризиків автоматизації, але й на моделювання комплексних соціальних зрушень: появи нових ролей, зміни компетентнісних профілів, розвитку гібридних моделей праці «людина у поєднанні із штучним інтелектом».

Зосередження уваги на навичках майбутнього є фінальним і водночас ключовим акцентом роботи. Основними компетентностями, що будуть визначати конкурентоспроможність працівника, є не лише технічна грамотність, а й здатність до міждисциплінарної інтеграції, критичне мислення, креативність, цифрова комунікація та гнучкість у навчанні. У світі, де алгоритми виконуватимуть дедалі складніші операції, саме людська здатність до адаптації та синтезу інформації стане

визначальною. Ця теза безпосередньо перетинається з підходом NLP-аналізу, оскільки аналіз текстових стратегій, дискурсів та політичних документів дозволяє виявити, як суспільства концептуалізують ці навички та яким чином формують запит на нові професійні ролі.

Ефект «бульбашки ШІ» полягає не у знеціненні технології, а у необхідності критично оцінювати її реальний вплив, відрізняючи фактичний прогрес від риторики. Такий підхід дозволяє сформуванню збалансованого бачення майбутнього ринку праці, де штучний інтелект – не універсальна заміна працівникам, а важливий, але все ще обмежений інструмент, який потребує людської експертизи, регулювання і поступової інтеграції.

Тому вважаю, що найближчі роки професії штучним інтелектом не будуть замінені. ШІ вносить зміни у роботу, допомагаючи швидше впоратись з рутиною, але його допомога не завжди є точною та вірною, тому без аналізу людини даний процес не можливий. Також варто не забувати, що на даному етапі ШІ не може гарантувати нам повну конфіденційність інформації, тому більшість організацій які мають конфіденційну інформацію, яка при витоку у маси може мати негативні наслідки, не може залучати ШІ навіть для рутинних завдань. Для даної роботи будуть люди, які матимуть зосереджено виконувати роботу та не допускати розкриття критично важливої інформації.

Найважливішим працівникам, які вже зайшли на ринок праці важливо постійно вдосконалюватись, навчатись, намагатись йти в ногу із розвитком інформаційних технологій. Тим самим це дасть можливість залишатись працювати за своєю спеціалізацією ще довгий час, адже кваліфікованого працівника ШІ замінити ще не здатний.

Водночас соціальні аспекти ринку праці залишаються критично важливими. Аналіз гендерної рівності оплати показав, що навіть у межах однієї галузі спостерігаються різні тенденції розподілу заробітної плати між чоловіками та жінками. У деяких установах гендер не впливає на рівень оплати після врахування інших факторів, у інших – існують статистично значущі відмінності, що вказує на

локальні особливості та необхідність постійного моніторингу соціальної справедливості.

Поєднання технологічного та соціального аналізу дозволяє зробити висновок, що ринок праці змінюється одночасно на двох рівнях: технологічному та соціальному. Трансформації, спричинені цифровізацією, не усувають соціальних нерівностей, а лише змінюють їхній характер, тому для ефективного управління трудовими ресурсами необхідно інтегрувати інструменти прогнозування автоматизації з регулярним оцінюванням рівності оплати.

Комплексний підхід, що поєднує аналітику на основі даних та соціально-економічні методи, дозволяє приймати обґрунтовані рішення на рівні організацій і державної політики, формуючи ринок праці, який є водночас ефективним, технологічно адаптованим і соціально справедливим.

На мою думку поєднання двох підходів дає змогу комплексно і з двох протилежних сторін здійснити оцінку ринку праці. Вважаю у майбутньому якщо об'єднати два методи у одну модель, а не розглядати окремо – це може надати зовсім нові висновки та розуміння про структуру ринку праці в Україні.

## ВИКОРИСТАНІ ДЖЕРЕЛА

1. Мікроекономіка: підручник / А. П. Наливайко, Л. А. Азьмук, П. Г. Банщикова та ін. ; за наук. ред. А. П. Наливайка; М-во освіти і науки, молоді та спорту України, ДВНЗ "Київ. нац. екон. ун-т ім. Вадима Гетьмана". Київ : КНЕУ, 2011. 446 с.
2. Економіка праці та соціально-трудова відносини : підручник / [А. М. Колот, О.А. Грішнова, О.О. Герасименко та ін.] ; за наук. ред. д-ра екон. наук, проф. А.М. Колота. К. : КНЕУ, 2009. 711 с.
3. Дуляба О., Юрчик А. ДОСЛІДЖЕННЯ СУТНОСТІ НА ЗМІСТУ ЛЮДСЬКОГО ПОТЕНЦІАЛУ. Економіка та суспільство. 2024. № 68. URL:<https://doi.org/10.32782/2524-0072/2024-68-107> (дата звернення: 29.10.2025).
4. Pigou A. C. Unemployment. London : Macmillan, 1920. 344 с.
5. Маркс, К. *Капітал. Том I* [Електронний ресурс]. Hamburg : Otto Meissner Verlag, 1867. 704 с. Режим доступу: [https://www.marxists.org/ukrainian/marx-engels/kapital/Kapital\\_Tom\\_1\\_Knyha\\_1.pdf](https://www.marxists.org/ukrainian/marx-engels/kapital/Kapital_Tom_1_Knyha_1.pdf) (дата звернення: 29.10.2025).
6. Петюх В. М. Ринок праці: Навч. посібник. Київ: КНЕУ, 1999. 288 с.
7. Marshall A. Principles of economics. 4th ed. London : Macmillan and co., limited, 1898. 820 с.
8. Smith, A. *An Inquiry into the Nature and Causes of the Wealth of Nations* [Електронний ресурс]. London: W. Strahan and T. Cadell, 1776. Режим доступу: <https://www.gutenberg.org/ebooks/3300> (дата звернення: 29.10.2025).
9. Класифікатор професій: Загальні положення. Міністерство економіки, довкілля та сільського господарства України. URL: <https://me.gov.ua/Documents/Detail/90ffb288-8826-4914-8709-8815379a3fe1?lang=uk-UA&id=bd21c45a-a984-456e-b523-c03aa942dcb9&title=RekomendatsiiSchodoVidpovidalnogoVikoristanniaShi->

- PitanniaPravaIntelektualnoiVlasnosti&showMenuTree=true (дата звернення: 29.10.2025).
10. Місія, пріоритети, послуги. *Конфедерація роботодавців України*. URL: <https://employers.org.ua> (дата звернення: 29.10.2025).
11. Класифікатор професій: Пошук. *Міністерство економіки, довкілля та сільського господарства України*. URL: <https://me.gov.ua/Profession/List?lang=uk-UA&id=d4162ef8-2771-4ac5-99ef-1d4b6f5336af&tag=KlasifikatorProfesii-Poshuk> (дата звернення: 29.10.2025).
12. Олексій Мірошниченко: У який навчальний заклад потрібно вступати, щоб за 5 років штучний інтелект не відібрав роботу?. *Моя Школа. OBOZ.UA*. URL: <https://shkola.obozrevatel.com/ukr/news/u-yakij-navchalnij-zaklad-potribno-vstupati-schob-za-5-rokiv-shtuchnij-intelekt-ne-vidibrav-robotu.htm> (дата звернення: 29.10.2025).
13. Kaur S. Impact of ai on labor markets: intensifying or mitigating unemployment. *National Research Journal of Business Economics*. 2025. Т. 12, № 1. С. 88–97.
14. Piro, Pietro. «Artificial Intelligence (IA) and Labour Markets.» Preprint (May 2025).
15. Молчановський О. Бульбашка III: за чим треба слідкувати? м. Київ, 11.11.2025
16. Pilehvar, M. T., & Camacho-Collados, J. (2020). *Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning*. Springer. URL: <https://doi.org/10.1007/978-3-030-32381-1> (дата звернення: 29.10.2025).
17. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.org. URL: <https://arxiv.org/abs/1810.04805> (дата звернення: 29.10.2025).
18. Reimers N., Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT-networks. URL: <https://arxiv.org/abs/1908.10084> (дата звернення: 29.10.2025).

19. Chicco D. Siamese Neural Networks: An Overview. In: Artificial Neural Networks. Springer, 2021. P. 73–94.
20. Замлинна Я. В., Катуніна О. Прикладні задачі використання дифузійних моделей в економіці та управлінні. *Цифрова економіка* : Зб. матеріалів III Міжнар. науково-практ. конф., м. Київ, 5–6 черв. 2025 р. Київ, 2025.
21. TF-IDF. Ranktracker. URL: <https://www.ranktracker.com/uk/seo/glossary/td-idf/> (дата звернення: 29.10.2025).
22. Jacob Murel D., Kavlakoglu E. What Is Ridge Regression?. IBM. IBM. URL: <https://www.ibm.com/think/topics/ridge-regression> (дата звернення: 29.10.2025).
23. Економіко-математичні методи і моделі: економетрика підручник / В. В. Вітлінський, С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. К. : КНЕУ, 2013. – 502 с.
24. Методи та технології обчислювального інтелекту: навч. посіб. для студ. спеціальності 122 «Комп’ютерні науки» / І. В. Федорін; КПІ ім. Ігоря Сікорського. Київ, 2022. 314 с.
25. Моделювання та інформ. системи в економіці : зб. наук. праць / відп. ред. В. К. Галіцин. 2011. № 84. 1. 272 с.
26. Економіко-математичні методи і моделі: практикум / В. В. Вітлінський, Т. О. Білик, Г. І. Великоіваненко та ін. Київ: КНЕУ, 2014. 221 с.
27. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Incorporated, 2022.
28. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning Data Mining, Inference, and Prediction. 2nd ed. 2017.
29. An Introduction to Statistical Learning / G. James та ін. New York, NY : Springer US, 2021. URL: <https://doi.org/10.1007/978-1-0716-1418-1> (дата звернення: 05.11.2025).
30. Zhou Z.-H. Ensemble Methods: Foundations and Algorithms. 2nd ed. Taylor & Francis Group, 2025. 292 с.
31. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2017. 800 с.

32. Chollet F. Deep Learning with Python, Second Edition. Manning Publications Co. LLC, 2021. 400 с.
33. What company do words keep? Revisiting the distributional semantics of J.R. Firth & Zellig Harris. arXiv.org. URL: <https://arxiv.org/abs/2205.07750> (дата звернення: 15.11.2025).
34. Unsupervised cross-lingual representation learning at scale. arXiv.org. URL: <https://arxiv.org/abs/1911.02116> (дата звернення: 15.11.2025).
35. Krantz T., Jonker A. What Is Cosine Similarity? IBM. URL: <https://www.ibm.com/think/topics/cosine-similarity> (дата звернення: 15.11.2025).
36. Murphy K. P. *Probabilistic Machine Learning: Advanced Topics*. Cambridge, MA : MIT Press, 2023. 742 с.
37. Exploring Dimensionality Reduction Techniques in Multilingual Transformers / Á. Huertas-García та ін. *Cognitive Computation*. 2022. URL: <https://doi.org/10.1007/s12559-022-10066-8> (дата звернення: 15.11.2025).
38. Jolliffe I. T., Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016. Т. 374, № 2065. С. 20150202. URL: <https://doi.org/10.1098/rsta.2015.0202> (дата звернення: 15.11.2025).
39. Бізнес-аналітика багатовимірних процесів : навчальний посібник / Т.С. Клебанова, Л.С. Гур'янова, Л.О. Чаговець та ін. Харків : ХНЕУ ім. С. Кузнеця, 2018. 272 с.
40. Основи статистичного моделювання: навч. посібник / за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП «Рута», 2022. 604с.
41. Freiesleben T., Grote T. Beyond generalization: a theory of robustness in machine learning. *Synthese*. 2023. Т. 202, № 4. URL: <https://doi.org/10.1007/s11229-023-04334-9> (дата звернення: 25.11.2025).
42. Consistency of the bootstrap for asymptotically linear estimators based on machine learning. arXiv.org. URL: <https://doi.org/10.48550/arXiv.2404.03064> (дата звернення: 15.11.2025).

43. Ablation Studies in Artificial Neural Networks. *arXiv.org*. URL: <https://doi.org/10.48550/arXiv.1901.08644> (дата звернення: 25.11.2025).
44. Don't let your analysis go to seed: on the impact of random seed on machine learning-based causal inference - PMC. PMC Home. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12166761/> (дата звернення: 15.11.2025).
45. Big Data for Labour Market Intelligence: Challenges and Opportunities. European Training Foundation. Luxembourg: Publications Office of the European Union, 2019. – 136 с.
46. AI для HR: як штучний інтелект трансформує HR. SAP. URL: <https://www.sap.com/ukraine/resources/ai-for-hr> (дата звернення: 10.11.2025).
47. Як Big Data, III і метавсесвіт розвивають HR-індустрію Kyivstar Business Hub. Kyivstar Business Hub корпоративний блог для бізнесу. URL: <https://hub.kyivstar.ua/articles/yak-big-data-shtuchnyj-intelekt-i-metavsesvit-rozvyvayut-hr-industriyu> (дата звернення: 15.11.2025).
48. Equal pay analysis with Logib. Logib. URL: <https://logib.legacy.admin.ch/home> (дата звернення: 15.11.2025).
49. The Future of Employment: How susceptible are jobs to computerisation?. *Oxford Martin School*. URL: <https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-employment> (дата звернення: 05.11.2025).
50. SOC home. Bureau of Labor Statistics. URL: <https://www.bls.gov/soc/> (дата звернення: 05.11.2025).
51. Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages / J. Manyika та ін. McKinsey & Company. URL: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages> (дата звернення: 05.11.2025).
52. Brookings Institution. *Automation and Artificial Intelligence: How machines are affecting people and places*. URL: <https://www.brookings.edu/wp->

- content/uploads/2019/01/ES\_2019.01\_BrookingsMetro\_Automation-AI\_Report\_Muro-Maxim-Whiton-FINAL.pdf (дата звернення: 05.11.2025).
- 53.OECD. Adapting (to) Automation. URL: <https://www.oecd.org/publications/adapting-to-automation-905fdc2c-en.htm> (дата звернення: 05.11.2025).
- 54.Ozili P. K. The Acceptable R-Square in Empirical Modelling for Social Science Research. SSRN Electronic Journal. 2022. URL: <https://doi.org/10.2139/ssrn.4128165> (дата звернення: 05.11.2025).
- 55.Інспекція праці інформує: Дискримінація у сфері праці: види, причини, прояви. Східне міжрегіональне управління Державної служби України з питань праці. URL: <https://smu.dsp.gov.ua/inspektsiia-pratsi-informuie-dyskryminatsiia-u-sferi-pratsi-vydy-prychyny-proiavy/> (дата звернення: 10.11.2025).
- 56.CDE Network. *Welcome to the SDC Climate, DRR & Environment Network*. URL: <https://www.sdc-cde.ch/en/nsb?id=103518> (дата звернення: 10.11.2025).
- 57.Gender equality in all areas of life. Eidgenössisches Büro für die Gleichstellung von Frau und Mann (EBG). URL: <https://www.ebg.admin.ch/en>(дата звернення: 10.11.2025).
- 58.Strub, S. (2004). *Überprüfung der Einhaltung von Lohngleichheit zwischen Frauen und Männern bei Beschaffungen des Bundes. Bericht über die Pilotphase zur Umsetzung von Art. 8 Abs. 1 Bst. C des Bundesgesetzes über das öffentliche Beschaffungswesen*. Bern: Büro BASS.
- 59.INFRAS (2011). *Evaluation der Kontrollen im Beschaffungswesen. Schlussbericht*. Erstellt im Auftrag des Eidgenössischen Büros für die Gleichstellung von Frau und Mann EBG. Zürich, Infrac.
- 60.Becker G. S. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University Of Chicago Press, 1994. 412 с.
- 61.International Labour Organization. URL: <https://www.ilo.org/> (дата звернення: 10.11.2025).

62. Equal pay analysis with Logib. Eidgenössisches Büro für die Gleichstellung von Frau und Mann (EBG). URL: <https://www.ebg.admin.ch/en/equal-pay-analysis-with-logib> (дата звернення: 10.11.2025).
63. Logib is epic's 1st good practice. epic. *Home Page. epic*. URL: [https://www.equalpayinternationalcoalition.org/en/whats\\_new/logib-is-epics-1st-good-practice](https://www.equalpayinternationalcoalition.org/en/whats_new/logib-is-epics-1st-good-practice) (дата звернення: 15.11.2025).
64. Muth T. Switzerland leads the way in the fight against 'unexplained' wage gaps. *SWI swissinfo.ch*. URL: <https://www.swissinfo.ch/eng/workplace-switzerland/switzerland-leads-the-way-in-the-fight-against-unexplained-wage-gaps/82785211> (дата звернення: 15.11.2025).
65. Logib pay system recognised as EPIC's good practice. epic. *EPIC*. URL: [https://www.equalpayinternationalcoalition.org/en/whats\\_new/logib-pay-system-recognised-epics-good-practice](https://www.equalpayinternationalcoalition.org/en/whats_new/logib-pay-system-recognised-epics-good-practice) (date of access: 15.11.2025).
66. Equal pay: Swiss Post's commitment achieves results. Swiss Post. URL: <https://www.post.ch/en/about-us/news/2023/equal-pay-swiss-posts-commitment-achieves-results> (дата звернення: 15.11.2025).
67. Gender Equality Act: amendments that bring new opportunities. *PwC*. URL: <https://www.pwc.ch/en/insights/hr/gender-equality-act.html> (дата звернення: 15.11.2025).
68. Замлинна Я., О. Притоманова. Модель LOGIB як інструмент виявлення та оцінки гендерних розривів оплати праці. *Моделювання економіки: проблеми, тенденції, досвід* : XIII ВСЕУКР. НАУКОВО-ПРАКТ. КОНФ. ФОРУМ МОЛОДИХ ЕКОНОМІСТІВ-КІБЕРНЕТИКІВ, м. Львів, 21–22 листоп. 2025 р. / Львів, 2025. С. 348. URL: [https://econom.lnu.edu.ua/wp-content/uploads/2025/12/2025\\_Zbirnyk-OK-NEW-1.pdf](https://econom.lnu.edu.ua/wp-content/uploads/2025/12/2025_Zbirnyk-OK-NEW-1.pdf).

## ДОДАТКИ

### Додаток А

#### Sentence-BERT Embeddings & TF-IDF

```

MODEL_NAME = "paraphrase-multilingual-MiniLM-L12-v2"
sbert = SentenceTransformer(MODEL_NAME)
ukr_emb = sbert.encode(ukr_df["clean_text"].tolist(), show_progress_bar=True)
ox_emb = sbert.encode(oxford_df["occupation_clean"].tolist(), show_progress_bar=True)
np.save(ART / "ukr_emb.npy", ukr_emb)
np.save(ART / "ox_emb.npy", ox_emb)

# TF-IDF baseline
tfidf = TfidfVectorizer(max_features=2000)
tfidf_ukr = tfidf.fit_transform(ukr_df["clean_text"])
tfidf_ox = tfidf.transform(oxford_df["occupation_clean"])
joblib.dump(tfidf, ART / "tfidf_vectorizer.pkl")
joblib.dump(tfidf_ukr, ART / "tfidf_ukr.pkl")
joblib.dump(tfidf_ox, ART / "tfidf_ox.pkl")

# Порівняння розмірів
print(f"SBERT розмірність: {ukr_emb.shape[1]}, TF-IDF розмірність: {tfidf_ukr.shape[1]}")

```

#### Надання ознак

```

# Словники
tech_words = ["машина", "технолог", "робот", "виробництво", "обладнання", "інженер",
"система", "механізм", "техніка", "метал", "енергія", "станок", "пристрій", "автомат",
програмування", "комп'ютер", "електроніка", "алгоритм", "обчислення", "сервіс",
"технологія", "процесор", "електрик", "механіка", "матеріал", "CAD", "контролер", "дизайн-
система", "штучний інтелект"]
social_words = ["комунікація", "керування", "викладання", "освіта", "допомога",
"соціальний", "обслуговування", "клієнт", "людина", "співпраця", "психологія", "мова",
"учень", "порада", "громада", "лідерство", "коучинг", "навчання", "персонал", "колектив",
"пацієнт", "студент", "взаємодія", "учитель", "підтримка", "консультація", "зворотний
зв'язок"]
manual_words = ["ремонт", "збірка", "виготовлення", "монтаж", "налагодження",
"будівництво", "свердління", "фарбування", "покриття", "укладання", "зварювання", "різання",
"обробка", "шліфування", "встановлення", "заміна", "токарь", "механік", "монтажник",
"слюсар", "електрик", "водій", "оператор", "майстер", "технік", "верстат", "інструмент"]
creative_words = ["дизайн", "дизайнер", "художник", "мистец", "режисер", "копірайт",
"журналіст", "фотограф", "композитор", "актор", "письменник", "редактор", "ілюстратор",
"модельєр", "скульптор", "куратор", "сценарист", "стиліст", "бренд", "контент", "креатив",
"маркетолог", "аналітик", "архітектор", "копірайтер", "візуалізація", "реклама", "медіа",
"продюсер"]

# Функція підрахунку частки
def word_score(text, words):
    text = str(text).lower()
    count = sum(1 for w in words if re.search(rf"\b{w}\b", text))
    return count / (len(text.split()) + 1)

def clean_text_uk(text):
    text = str(text).lower()
    text = re.sub(r"[^a-яіієєа-z0-9\s]", " ", text)

```

```

text = re.sub(r"\s+", " ", text).strip()
return text

# Український датасет
if "Опис" in ukr_df.columns:
    ukr_df["full_text"] = ukr_df["Назва"].astype(str) + ". " + ukr_df["Опис"].astype(str)
else:
    ukr_df["full_text"] = ukr_df["Назва"].astype(str)
ukr_df["clean_text"] = ukr_df["full_text"].apply(clean_text_uk)

# Oxford датасет
ox_col = None
for c in oxford_df.columns:
    if "occupation" in c.lower():
        ox_col = c
        break

if ox_col:
    oxford_df["clean_text"] = oxford_df[ox_col].astype(str).apply(clean_text_uk)
    print(f" Колонка '{ox_col}' використана для очищення тексту в Oxford dataset.")
else:
    raise KeyError("Не знайдено колонки з назвою професії в Oxford dataset!")
    # Додання ознак до ukr_df
for df in [ukr_df, oxford_df]:
    df["tech_score"] = df["clean_text"].apply(lambda x: word_score(x, tech_words))
    df["social_score"] = df["clean_text"].apply(lambda x: word_score(x, social_words))
    df["manual_score"] = df["clean_text"].apply(lambda x: word_score(x, manual_words))
    df["creative_score"] = df["clean_text"].apply(lambda x: word_score(x, creative_words))
    # Масштабування
extra_ukr = ukr_df[["tech_score", "social_score", "manual_score",
"creative_score"]].fillna(0).to_numpy()
extra_ox = oxford_df[["tech_score", "social_score", "manual_score",
"creative_score"]].fillna(0).to_numpy()
X_ukr = np.hstack([ukr_emb, extra_ukr])
X_ox = np.hstack([ox_emb, extra_ox])
scaler = StandardScaler()
scaler.fit(X_ox)
X_ox_scaled = scaler.transform(X_ox)
X_ukr_scaled = scaler.transform(X_ukr)
joblib.dump(scaler, ART / "scaler.pkl")
np.save(ART / "X_ox.npy", X_ox_scaled)
np.save(ART / "X_ukr.npy", X_ukr_scaled)
# Метрики
def evaluate_model(model, X, y, label):
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42 )
    start = time.time()
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    duration = time.time() - start
    return {
        "Embedding": label,
        "R2": r2_score(y_test, y_pred),

```

```

    "MAE": mean_absolute_error(y_test, y_pred),
    "MSE": mean_squared_error(y_test, y_pred),
    "Spearman": spearmanr(y_test, y_pred).correlation,
    "TrainTime_sec": duration }
# Навчання Ridge на embedding
ridge = Ridge(alpha=1.0)
results = pd.DataFrame([
    evaluate_model(ridge, X_sbert, y, "SBERT"),
    evaluate_model(ridge, X_tfidf, y, "TF-IDF")])
display(results)
plt.figure(figsize=(8,5))
sns.barplot(
    data=results.melt(id_vars="Embedding",
                      value_vars=["R2", "Spearman", "MAE"]),
    x="variable", y="value", hue="Embedding", palette="Set2")
plt.title("TF-IDF vs SBERT - порівняння якості")
plt.ylabel("Score")
plt.xlabel("")
plt.tight_layout()
plt.show()

corr = ukr_df[["tech_score", "social_score", "manual_score", "creative_score"]].corr()
plt.figure(figsize=(6,5))
sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Кореляції професійними характеристиками (український датасет)")
plt.tight_layout()
plt.show()

# Додавання лінгвістичних індикаторів до Oxford dataset
def word_score(text, words):
    """Підрахунок частки ключових слів у тексті."""
    if not isinstance(text, str) or len(text.strip()) == 0:
        return 0
    count = sum(text.lower().count(w) for w in words)
    return count / (len(text.split()) + 1)
tech_words = ["machine", "robot", "automation", "engineer", "mechanic", "technology",
              "system", "equipment", "device", "programming", "metal", "construction",
              "energy", "manufacturing", "technician", "computer", "hardware"]
social_words = ["communication", "teacher", "education", "help", "social", "customer",
               "human", "team", "psychology", "language", "advice", "service", "training",
               "nurse", "care", "community", "interaction"]
manual_words = ["repair", "assembly", "construction", "maintenance", "installation",
                "cleaning", "operation", "manufacture", "fixing", "manual", "driver"]
creative_words = ["design", "designer", "art", "creative", "illustrator", "photographer",
                  "writer", "actor", "musician", "composer", "director", "editor",
                  "content", "brand", "marketing", "creator", "journalist"]
oxford_df["tech_score"] = oxford_df["occupation_clean"].apply(lambda x: word_score(x,
tech_words))
oxford_df["social_score"] = oxford_df["occupation_clean"].apply(lambda x: word_score(x,
social_words))
oxford_df["manual_score"] = oxford_df["occupation_clean"].apply(lambda x: word_score(x,
manual_words))

```

```

oxford_df["creative_score"] = oxford_df["occupation_clean"].apply(lambda x: word_score(x,
creative_words))
oxford_df[["occupation_clean", "tech_score", "social_score", "manual_score",
"creative_score"]].head()

```

### Порівняння моделей на SBERT

```

# SBERT embeddings + технічні/соціальні/ручні/креативні ознаки
X_base = np.hstack([
    ox_emb,
    oxford_df[["tech_score", "social_score", "manual_score",
"creative_score"]].fillna(0).to_numpy()])
y = oxford_df["automation_risk_norm"].to_numpy()
# Масштабування
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_base)
# Розбиття
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42)
models = {
    "LinearRegression": LinearRegression(),
    "Ridge": Ridge(alpha=1.0),
    "RandomForest": RandomForestRegressor(
        n_estimators=300, random_state=42, n_jobs=-1 ),
    "XGBoost": XGBRegressor(
        n_estimators=300, learning_rate=0.05, max_depth=6,
        subsample=0.9, colsample_bytree=0.8, random_state=42),
    "MLP": MLPRegressor(
        hidden_layer_sizes=(128, 64), activation="relu",
        max_iter=300, random_state=42)}
cv = KFold(n_splits=5, shuffle=True, random_state=42)
results = []
for name, model in models.items():
    print(f"Навчання {name} ...")
    start = time.time()
    # Крос-валідація
    cv_scores = cross_val_score(model, X_train, y_train, cv=cv, scoring='r2')
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    # Метрики
    mse = mean_squared_error(y_test, y_pred)
    mae = mean_absolute_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)
    rho, _ = spearmanr(y_test, y_pred)
    elapsed = time.time() - start
    results.append({
        "Model": name,
        "CV_R2_mean": cv_scores.mean(),
        "CV_R2_std": cv_scores.std(),
        "Test_R2": r2,
        "MAE": mae,
        "MSE": mse,

```

```

        "Spearman": rho,
        "TrainTime_sec": elapsed    })
results_df = pd.DataFrame(results).sort_values("Test_R2", ascending=False)
display(results_df)
plt.figure(figsize=(10,6))
sns.barplot(data=results_df.melt(id_vars="Model",
value_vars=["CV_R2_mean", "Test_R2", "Spearman"]),
            x="variable", y="value", hue="Model", palette="coolwarm")
plt.title("Порівняння моделей за якістю прогнозу ризику автоматизації")
plt.ylabel("Score")
plt.xlabel("")
plt.tight_layout()
plt.show()
best = results_df.iloc[0]["Model"]
best_model = models[best]
best_model.fit(X_train, y_train)
y_pred = best_model.predict(X_test)
    plt.figure(figsize=(6,6))
sns.scatterplot(x=y_test, y=y_pred, alpha=0.6)
plt.plot([0,1],[0,1],"--",color="red")
plt.title(f"{best}: Реальні vs Прогнозовані значення")
plt.xlabel("Реальні значення (automation_risk_norm)")
plt.ylabel("Прогноз моделі")
plt.tight_layout()
plt.show()

```

### Фінальне навчання найкращої моделі (XGBoost)

```

feature_cols = ["tech_score", "social_score", "manual_score", "creative_score"]
X_full = np.hstack([
    ox_emb,
    oxford_df[feature_cols].fillna(0).to_numpy()])
y_full = oxford_df["automation_risk_norm"].to_numpy()
best_model = XGBRegressor(
    n_estimators=500,
    learning_rate=0.05,
    max_depth=5,
    subsample=0.8,
    colsample_bytree=0.8,
    random_state=42,
    n_jobs=-1)
start = time.time()
best_model.fit(X_full, y_full)
train_time = time.time() - start
y_pred_train = best_model.predict(X_full)
r2 = r2_score(y_full, y_pred_train)
mae = mean_absolute_error(y_full, y_pred_train)
mse = mean_squared_error(y_full, y_pred_train)

plt.figure(figsize=(10,5))
sns.barplot(
    x=best_model.feature_importances_[-len(feature_cols):],

```

```

y=feature_cols,
palette="crest")
plt.title("Важливість інженерних ознак у XGBoost")
plt.xlabel("Feature importance")
plt.tight_layout()
plt.savefig(ART_DIR / "xgb_feature_importance.png", dpi=300)
plt.show()

```

### Експеримент: "Назва" vs "Назва + Опис"

```

# Варіант 1: лише назва
ukr_df["text_name"] = ukr_df["Назва"].astype(str)
# Варіант 2: назва + опис (якщо опис є)
if "Опис" in ukr_df.columns:
    ukr_df["text_name_desc"] = ukr_df["Назва"].astype(str) + ". " +
ukr_df["Опис"].astype(str)
else:
    ukr_df["text_name_desc"] = ukr_df["Назва"].astype(str)
model_name = "sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2"
model = SentenceTransformer(model_name)
ukr_emb_name = model.encode(
    ukr_df["text_name"].tolist(),
    show_progress_bar=True,
    convert_to_numpy=True)
ukr_emb_name_desc = model.encode(
    ukr_df["text_name_desc"].tolist(),
    show_progress_bar=True,
    convert_to_numpy=True)
y = ukr_df["predicted_risk"].to_numpy()
scaler = StandardScaler()
X_name_scaled = scaler.fit_transform(ukr_emb_name)
X_name_desc_scaled = scaler.fit_transform(ukr_emb_name_desc)
def evaluate_model(X, y, name):
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42 )
    model = XGBRegressor(
        n_estimators=300,
        learning_rate=0.05,
        max_depth=5,
        subsample=0.8,
        colsample_bytree=0.8,
        random_state=42,
        n_jobs=-1 )
    start = time.time()
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    elapsed = time.time() - start
    return { "Модель": name,
            "R2": r2_score(y_test, y_pred),
            "MAE": mean_absolute_error(y_test, y_pred),
            "MSE": mean_squared_error(y_test, y_pred),
            "Spearman": spearmanr(y_test, y_pred).correlation,

```

```

        "TrainTime_sec": round(elapsed, 2) }
results_exp = pd.DataFrame([
    evaluate_model(X_name_scaled, y, "Назва"),
    evaluate_model(X_name_desc_scaled, y, "Назва + Опис")])
display(results_exp)
plt.figure(figsize=(8,5))
sns.barplot(
    data=results_exp.melt(
        id_vars="Модель",
        value_vars=["R2", "Spearman"] ),
    x="variable",
    y="value",
    hue="Модель",
    palette="Set2")
plt.title("Порівняння моделей за R2 та Spearman (XGBoost, українські дані)")
plt.xlabel("Метрика")
plt.ylabel("Значення")
plt.ylim(0, 1)
plt.legend(title="Варіант тексту")
plt.show()

```

### Семантичне вирівнювання UKR - OXFORD

```

sim_matrix = cosine_similarity(ukr_emb, ox_emb)
top_k = 5
matches = []
for i, ukr_name in enumerate(ukr_df["Назва"]):
    sims = sim_matrix[i]
    top_idx = np.argsort(sims)[-top_k:][:-1]
    top_matches = oxford_df.iloc[top_idx][["occupation_clean",
"automation_risk_norm"]].copy()
    for j, row in top_matches.iterrows():
        matches.append({
            "ukr_profession": ukr_name,
            "eng_profession": row["occupation_clean"],
            "similarity": sims[j],
            "automation_risk_norm": row["automation_risk_norm"] })
matches_df = pd.DataFrame(matches)
matches_df.to_csv("artifacts/ukr_to_oxford_similarity.csv", index=False)
avg_sim = np.mean(np.max(sim_matrix, axis=1))
high_sim_share = np.mean(np.max(sim_matrix, axis=1) > 0.7)
subset = sim_matrix[:50, :50]
plt.figure(figsize=(10, 6))
sns.heatmap(subset, cmap="YlOrRd")
plt.title("Семантична подібність між UKR і OXFORD (приклад 50x50)")
plt.xlabel("Oxford професії")
plt.ylabel("Українські професії")
plt.tight_layout()
plt.savefig("artifacts/plots/similarity_heatmap.png", dpi=300)
plt.show()
examples = (
    matches_df.groupby("ukr_profession")

```

```

        .apply(lambda x: x.nlargest(1, "similarity"))
        .reset_index(drop=True)
        .sort_values("similarity", ascending=False)
        .head(10))
display(examples)

```

### Прогноз українських професій

```

feature_cols = ["tech_score", "social_score", "manual_score", "creative_score"]
ukr_features = np.hstack([
    ukr_emb,
    ukr_df[feature_cols].fillna(0).to_numpy()])
X_ukr_scaled = scaler.transform(ukr_features)
pred = best_model.predict(X_ukr_scaled)
ukr_df["predicted_risk"] = np.clip(pred, 0, 1)
plt.figure(figsize=(8,5))
sns.histplot(ukr_df["predicted_risk"], bins=20, kde=True, color="skyblue")
plt.title("Розподіл прогнозованого ризику автоматизації серед українських професій")
plt.xlabel("Ймовірність автоматизації (0-1)")
plt.ylabel("Кількість професій")
plt.tight_layout()
plt.savefig(PLOTS / "risk_distribution.png", dpi=300)
plt.show()
display(ukr_df["predicted_risk"].describe())
    # Топ-20 професій з найвищим ризиком
top20 = ukr_df.nlargest(20, "predicted_risk")[["Назва", "predicted_risk"]]
plt.figure(figsize=(8,6))
sns.barplot(data=top20, y="Назва", x="predicted_risk", palette="Reds_r")
plt.title("ТОП-20 професій із найвищим ризиком автоматизації")
plt.xlabel("Прогнозований ризик")
plt.ylabel("")
plt.tight_layout()
plt.show()
    # Середній ризик по розділах
group_col = "Код"
if group_col in ukr_df.columns:
    avg_by_section = (
        ukr_df.groupby(group_col)["predicted_risk"]
        .mean()
        .reset_index()
        .sort_values("predicted_risk", ascending=False)    )
    print(f"\n Середній рівень ризику автоматизації по розділах ({group_col}):")
    display(avg_by_section.head(10))
else:
    print("Колонки 'Код' немає – пропущено групування за розділами.")
    # Найменш ризикові професії
low20 = ukr_df.nsmallest(20, "predicted_risk")[["Назва", "predicted_risk"]]
sns.barplot(data=low20, y="Назва", x="predicted_risk", palette="Blues_r")
plt.title("ТОП-20 професій із найнижчим ризиком автоматизації")
plt.tight_layout()
plt.show()
    # PCA карта

```

```

from sklearn.decomposition import PCA
X_pca = PCA(n_components=2).fit_transform(X_ukr_scaled)
plt.figure(figsize=(8,6))
sns.scatterplot(x=X_pca[:,0], y=X_pca[:,1],
                hue=ukr_df["predicted_risk"],
                palette="coolwarm", alpha=0.7)
plt.title("Професії у 2D-просторі ембедінгів (кольором – ризик автоматизації)")
plt.xlabel("PC1"); plt.ylabel("PC2")
plt.tight_layout()
plt.show()

```

### Кластеризація українських професій за рівнем ризику

```

feature_cols = ["predicted_risk", "tech_score", "social_score", "manual_score",
               "creative_score"]
X_cluster = ukr_df[feature_cols].fillna(0).to_numpy()
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
clusters = kmeans.fit_predict(X_cluster)
ukr_df["cluster"] = clusters
cluster_summary = ukr_df.groupby("cluster")[feature_cols].mean().reset_index()
cluster_summary["risk_level"] = ["низький", "середній", "високий"]
risk_map = (
    cluster_summary.set_index("cluster")["risk_level"].to_dict())
ukr_df["risk_level"] = ukr_df["cluster"].map(risk_map)
pca = PCA(n_components=2, random_state=42)
X_pca = pca.fit_transform(X_cluster)
ukr_df["PC1"] = X_pca[:, 0]
ukr_df["PC2"] = X_pca[:, 1]
plt.figure(figsize=(9,7))
sns.scatterplot(data=ukr_df, x="PC1", y="PC2", hue="risk_level", palette="coolwarm",
                alpha=0.8, s=60)
plt.title("Кластеризація українських професій за рівнем ризику автоматизації")
plt.xlabel("Головна компонента 1")
plt.ylabel("Головна компонента 2")
plt.legend(title="Рівень ризику")
plt.tight_layout()
plt.savefig("artifacts/plots/clusters_pca.png", dpi=300)
plt.show()
top_high_risk = ukr_df[ukr_df["risk_level"]=="високий"].sort_values("predicted_risk",
                           ascending=False).head(20)
print("Приклади професій з високим ризиком автоматизації:")
display(top_high_risk[["Код", "Назва", "predicted_risk", "risk_level"]])

```

### Робастність і Абляційні експерименти (на українських професіях)

```

feature_cols = ["tech_score", "social_score", "manual_score", "creative_score"]
X = np.hstack([ukr_emb, ukr_df[feature_cols].fillna(0).to_numpy()])
y = ukr_df["predicted_risk"].to_numpy()
    #Bootstrap стабільність прогнозів
N_BOOT = 50
boot_mean_pred = []
for i in tqdm(range(N_BOOT)):
    X_b, y_b = resample(X, y, replace=True, random_state=42 + i)

```

```

    model = XGBRegressor(
        n_estimators=300, learning_rate=0.05, max_depth=5,
        subsample=0.8, colsample_bytree=0.8, random_state=42 + i, n_jobs=-1    )
    model.fit(X_b, y_b)
    y_pred = model.predict(X)
    boot_mean_pred.append(y_pred)
boot_mean_pred = np.array(boot_mean_pred)
pred_std = boot_mean_pred.std(axis=0)
print(f"Середнє std прогнозів: {np.mean(pred_std):.4f}")
plt.figure(figsize=(7,4))
plt.hist(pred_std, bins=20, edgecolor="black")
plt.title("Стабільність прогнозів (Bootstrap std)")
plt.xlabel("Стандартне відхилення прогнозів")
plt.ylabel("Кількість професій")
plt.tight_layout()
plt.savefig(ART / "plots/bootstrap_stability_ukr.png", dpi=300)
plt.show()

    # Ablation study - внесок domain features
X_emb = ukr_emb
X_domain = ukr_df[feature_cols].fillna(0).to_numpy()
def evaluate_feature_set(X_input, label):
    start = time.time()
    model = XGBRegressor(
        n_estimators=300, learning_rate=0.05, max_depth=5,
        subsample=0.8, colsample_bytree=0.8, random_state=42, n_jobs=-1    )
    model.fit(X_input, y)
    y_pred = model.predict(X_input)
    r2_self = r2_score(y, y_pred)
    return {"Feature_Set": label, "R2(Self-fit)": r2_self, "TrainTime_sec": time.time()-
start}
results_ablation = pd.DataFrame([
    evaluate_feature_set(X_emb, "Тільки embeddings"),
    evaluate_feature_set(X_domain, "Тільки domain features"),
    evaluate_feature_set(np.hstack([X_emb, X_domain]), "Усі змінні разом")])
display(results_ablation)
plt.figure(figsize=(6,4))
plt.bar(results_ablation["Feature_Set"], results_ablation["R2(Self-fit)"],
        color=["#7eb0d5", "#b2e061", "#fd7f6f"])
plt.title("Ablation Study (українські дані)")
plt.ylabel("R² (Self-fit)")
plt.xticks(rotation=20)
plt.tight_layout()
plt.savefig(ART / "plots/ablation_r2_bar_ukr.png", dpi=300)
plt.show()

    # Перевірка стабільності на різних seed'ах

seeds = [1, 7, 42, 101, 2024]
seed_results = []
for seed in seeds:
    model = XGBRegressor(

```

```

        n_estimators=300, learning_rate=0.05, max_depth=5,
        subsample=0.8, colsample_bytree=0.8, random_state=seed, n_jobs=-1)
model.fit(X, y)
y_pred = model.predict(X)
seed_results.append(r2_score(y, y_pred))
seed_df = pd.DataFrame({"Seed": seeds, "R2(Self-fit)": seed_results})
seed_df.to_csv(ART / "seed_stability_ukr.csv", index=False)
print(seed_df)
plt.figure(figsize=(6,4))
plt.plot(seeds, seed_results, marker="o", linewidth=2)
plt.title("Стійкість R2 до різних seed'ів (українські професії)")
plt.xlabel("Seed")
plt.ylabel("R2 (Self-fit)")
plt.tight_layout()
plt.savefig(ART / "plots/seed_stability_ukr.png", dpi=300)
plt.show()

```

### Виявлення професій, які може замінити ШІ

```

high_risk_ai = ukr_ai[ukr_ai["predicted_risk"] >= 0.6].copy()
print(f"Високий ризик автоматизації: {len(high_risk_ai)} "
      f"({len(high_risk_ai)/len(ukr_ai)*100:.1f}%)")
if "clean_text" not in high_risk_ai.columns:
    high_risk_ai["full_text"] = (
        high_risk_ai["Назва"].astype(str) + ". " +
        high_risk_ai.get("Опис", "").astype(str)
    )
    high_risk_ai["clean_text"] = high_risk_ai["full_text"].str.lower()
cognitive_seed = ["data analyst", "бухгалтер", "юрист", "офісний працівник",
                  "перекладач", "дослідник", "менеджер проєктів", "психолог", "програміст"]
creative_seed = ["дизайнер", "журналіст", "копірайтер", "маркетолог", "контент-креатор"]
mechanical_seed = ["вантажник", "зварювальник", "токарь", "будівельник", "оператор верстата",
                  "монтажник", "пакувальник", "знімач брикетів"]
customer_seed = ["продавець", "касир", "офіціант", "бармен", "менеджер з продажу",
                 "адміністратор", "рецепціоніст"]
strategic_seed = [
    # базові стратегічні
    "ринковий аналітик", "фінансовий аналітик", "бізнес-аналітик",
    "економіст", "консультант", "аудитор", "експерт з даних", "стратег",
    # розширені стратегічні (на основі економічних досліджень)
    "chief economist", "legal advisor", "податковий консультант",
    "судовий експерт", "регулятор", "compliance officer",
    "chief analyst", "public policy expert"]
model = SentenceTransformer("paraphrase-multilingual-MiniLM-L12-v2")
emb_high = model.encode(high_risk_ai["clean_text"].tolist(), show_progress_bar=True)
emb_cog = model.encode(cognitive_seed)
emb_cre = model.encode(creative_seed)
emb_mech = model.encode(mechanical_seed)
emb_cust = model.encode(customer_seed)
emb_strat = model.encode(strategic_seed)
high_risk_ai["sim_cognitive"] = cosine_similarity(emb_high, emb_cog).max(axis=1)
high_risk_ai["sim_creative"] = cosine_similarity(emb_high, emb_cre).max(axis=1)
high_risk_ai["sim_mechanical"] = cosine_similarity(emb_high, emb_mech).max(axis=1)

```

```

high_risk_ai["sim_customer"] = cosine_similarity(emb_high, emb_cust).max(axis=1)
high_risk_ai["sim_strategic"] = cosine_similarity(emb_high, emb_strat).max(axis=1)
seed_texts = cognitive_seed + creative_seed + mechanical_seed
seed_labels = [1] * (len(cognitive_seed) + len(creative_seed)) + [0] * len(mechanical_seed)
clf_ai = LogisticRegression(max_iter=1000).fit(model.encode(seed_texts), seed_labels)
high_risk_ai["ai_class_prob"] = clf_ai.predict_proba(emb_high)[: , 1]
def normalize(s):
    return (s - s.min()) / (s.max() - s.min() + 1e-9)
prisk_norm = normalize(high_risk_ai["predicted_risk"])
ai_score = (
    0.22 * prisk_norm + # базовий ризик – основа
    0.30 * high_risk_ai["sim_cognitive"] +
    0.25 * high_risk_ai["sim_creative"] +
    0.20 * high_risk_ai["ai_class_prob"] -
    0.20 * high_risk_ai["sim_customer"] -
    0.40 * high_risk_ai["sim_mechanical"] - # ручна робота зменшує заміщення
    0.25 * high_risk_ai["sim_strategic"] ) # стратегічні ролі – найменше замінні
high_risk_ai["ai_replacement_score"] = normalize(ai_score)
def classify(score):
    if score >= 0.9: return "AI-only (повна заміна)"
    if score >= 0.7: return "AI + Human (часткова заміна)"
    if score >= 0.5: return "AI-assistant (допомога)"
    return "Low/None"
high_risk_ai["AI_replacement_level"] = high_risk_ai["ai_replacement_score"].apply(classify)
export_cols = [
    "Назва", "predicted_risk",
    "sim_cognitive", "sim_creative", "sim_customer",
    "sim_mechanical", "sim_strategic",
    "ai_class_prob", "ai_replacement_score",
    "AI_replacement_level"]
high_risk_ai[export_cols].to_csv(ART / "ai_replacement_final.csv", index=False)
summary = high_risk_ai["AI_replacement_level"].value_counts().reset_index()
summary.columns = ["Категорія", "Кількість"]
plt.figure(figsize=(7,5))
sns.barplot(data=summary, x="Категорія", y="Кількість", palette="coolwarm")
plt.title("Розподіл професій за рівнем заміни ШІ (фінальний)")
plt.xticks(rotation=20)
plt.tight_layout()
plt.savefig(PLOTS / "ai_replacement_levels_final.png", dpi=300)
plt.show()
plt.figure(figsize=(7,5))
sns.scatterplot(
    data=high_risk_ai,
    x="sim_cognitive", y="ai_replacement_score",
    hue="AI_replacement_level", alpha=0.7)
plt.title("Когнітивність vs AI Replacement Score")
plt.tight_layout()
plt.savefig(PLOTS / "ai_scatter_final.png", dpi=300)
plt.show()

```



## Звіт подібності

### Метадані

Назва організації

**Kyiv National Economic University named after Vadym Hetman KNEU**

Заголовок

**Замлинна\_Притоманова\_плагіат**

Автор Науковий керівник / Експерт

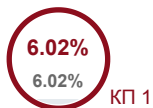
**ЗамлиннаПритоманова**

підрозділ

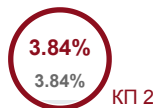
**кафедра штучного інтелекту, моделювання та статистики**

### Обсяг знайдених подібностей

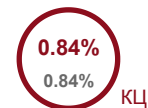
Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.


**25**

Довжина фрази для коефіцієнта подібності 2


**15792**

Кількість слів


**122075**

Кількість символів

### Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		3
Інтервали		0
Мікропробіли		2
Білі знаки		0
Парафрази (SmartMarks)		45

### Джерела

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Колір тексту означає в якому джерелі він був знайдений. Ці джерела і значення Коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

#### 10 найдовших фраз

Колір тексту

ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	<a href="http://ir.kneu.edu.ua/bitstream/handle/2010/11412/ekon_pratsi_09.pdf?sequence=1&amp;isAllowed=y">http://ir.kneu.edu.ua/bitstream/handle/2010/11412/ekon_pratsi_09.pdf?sequence=1&amp;isAllowed=y</a>	85 0.54 %
2	<a href="https://dovidnyk.in.ua/directories/profesii">https://dovidnyk.in.ua/directories/profesii</a>	55 0.35 %
3	<a href="https://dovidnyk.in.ua/directories/profesii">https://dovidnyk.in.ua/directories/profesii</a>	47 0.30 %
4	<a href="http://ir.kneu.edu.ua/bitstream/handle/2010/11412/ekon_pratsi_09.pdf?sequence=1&amp;isAllowed=y">http://ir.kneu.edu.ua/bitstream/handle/2010/11412/ekon_pratsi_09.pdf?sequence=1&amp;isAllowed=y</a>	43 0.27 %
5	<a href="https://dovidnyk.in.ua/directories/profesii">https://dovidnyk.in.ua/directories/profesii</a>	41 0.26 %



# CERTIFICATE

is awarded to

**YANA ZAMLYNNA**

for being an active participant in  
III International Scientific and Practical Conference  
"DIGITAL ECONOMY"

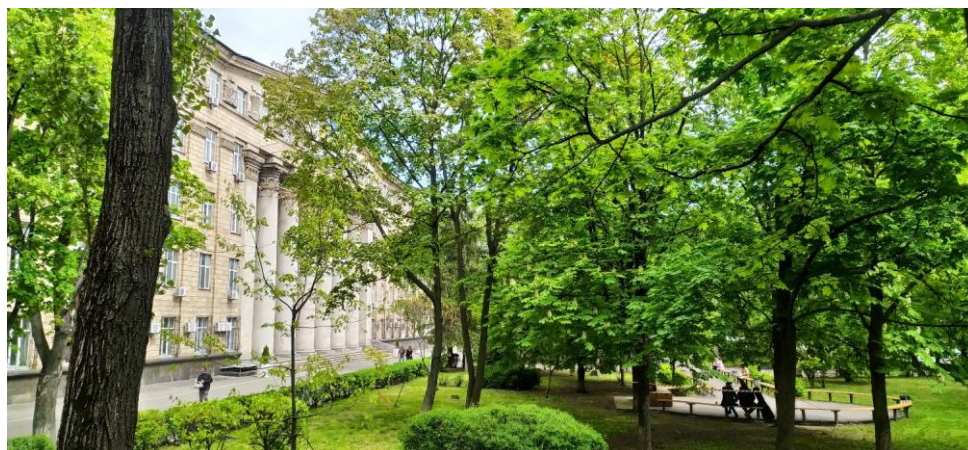
*15 hours of participation (0,5 ECTS credits)*



Kyiv, Ukraine, June 5 - 6, 2025

№ DE-III/2025/050

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВАДИМА ГЕТЬМАНА  
ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ В ЕКОНОМІЦІ  
НАУКОВИЙ ПАРК КНЕУ  
PHILIPPS UNIVERSITÄT MARBURG (НІМЕЧЧИНА)  
OLD DOMINION UNIVERSITY (США)  
SUSTAINABILITY INNOCENTER (ШВЕЦІЯ)  
ARENGUKOOSTÖÖ ÜMARLAUD (ЕСТОНІЯ)  
AMERICAN UNIVERSITY OF ARMENIA (ВІРМЕНІЯ)  
UNIwersytet w Białymstoku (ПОЛЬЩА)**



## **ЦИФРОВА ЕКОНОМІКА**

### **Збірник матеріалів III Міжнародної науково-практичної конференції**

**5-6 червня 2025 р.**

Нікіфорчук А.І., Кулікова Є.С. ФІНАНСОВІ РІШЕННЯ НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ: МОЖЛИВОСТІ ДЛЯ УКРАЇНСЬКИХ ПІДПРИЄМСТВ.....	683
Олійник К. Д. ШТУЧНИЙ ІНТЕЛЕКТ ЯК ТРИГЕР ТРАНСФОРМАЦІЙНИХ ЗМІН: ВПЛИВ ЦИФРОВОЇ ЕКОНОМІКИ НА РИНОК ПРАЦІ .....	686
Петренко Л.М. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ДІЯЛЬНОСТІ ЮРИСТІВ .....	691
Попівняк Ю.М. ЧАТ-БОТИ ЯК ІНСТРУМЕНТ АВТОМАТИЗАЦІЇ БУХГАЛТЕРСЬКОГО ОБЛІКУ В УМОВАХ ЦИФРОВОЇ ТРАНСФОРМАЦІЇ .....	695
Семенов С.Г., Єнгалічев С.О. АДАПТИВНЕ ПЛАНУВАННЯ ЗАДАЧ У РОЗПОДІЛЕНИХ СИСТЕМАХ НА ОСНОВІ АЛГОРИТМУ ОПТИМІЗАЦІЇ ПОЛІТИК ІЗ ПРОКСИМАЛЬНИМ ОНОВЛЕННЯМ.....	698
Скляренко П.А., Вілянський А.В. ШТУЧНИЙ ІНТЕЛЕКТ ЯК ІНСТРУМЕНТ СТРАТЕГІЧНОГО УПРАВЛІННЯ ТА ЦИФРОВОЇ ТРАНСФОРМАЦІЇ БІЗНЕСУ .....	701
Старенький Ю.В., Кудельський В. Е. ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В КІБЕРБЕЗПЕЦІ .....	705
Терещенко А.С., Орленко О.М. ОСОБЛИВОСТІ ФОРМУВАННЯ ЦІН ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ .....	709
Торгонська Д. І., Матвійчук Н. М. ІНТЕГРАЦІЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ В СИСТЕМУ ОБЛІКУ ТА АУДИТУ: ТРАНСФОРМАЦІЯ ПРОФЕСІЙНИХ КОМПЕТЕНЦІЙ .....	711
Троян К.М. ПЕРСПЕКТИВИ ВИКОРИСТАННЯ АРХІТЕКТУРИ ТРАНСФОРМЕР В АНАЛІЗІ ФІНАНСОВИХ ЧАСОВИХ РЯДІВ.....	715
Хортів К., Сухорукова А. Л. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА МАШИННОГО НАВЧАННЯ В ЕКОНОМІЦІ: МОЖЛИВОСТІ, ВИКЛИКИ ТА ПЕРСПЕКТИВИ.....	719
Чеверда С.С. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ В ОПТИМІЗАЦІЇ РОЗПОДІЛУ РЕСУРСІВ ПРОЄКТНОГО ОФІСУ: ІНТЕГРАЦІЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ.....	722
Чорненький М.В. ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ФІНАНСОВИХ ПРОЦЕСАХ ERP-СИСТЕМИ ПІДПРИЄМСТВА.....	725
Чубань В.С. ШТУЧНИЙ ІНТЕЛЕКТ У СФЕРІ ПОСЛУГ: ЧАТ-БОТИ, ВІРТУАЛЬНІ АСИСТЕНТИ ТА ПЕРСОНАЛІЗАЦІЯ СЕРВІСУ .....	728
Шафалюк О. К., Ахмедов Р. Р. АВТОГЕНЕРАЦІЯ ТЕКСТІВ У МАРКЕТИНГОВІЙ ДІЯЛЬНОСТІ: ЕФЕКТИВНІСТЬ, ВИКЛИКИ, ПЕРСПЕКТИВИ.....	731

**СЕКЦІЯ 3. МОДЕЛІ ГЛИБОКОГО НАВЧАННЯ  
В ЗАДАЧАХ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ ТА ОБРОБКИ МОВИ**

Єрмоленко С.В. РОЗПІЗНАВАННЯ ТЕКСТОВИХ ТА ВІЗУАЛЬНИХ ОБ'ЄКТІВ ДЛЯ МУЛЬТИМОДАЛЬНОГО АНАЛІЗУ КОНТЕНТУ .....	735
Замлинна Я.В., Катуніна О.С. ПРИКЛАДНІ ЗАДАЧІ ВИКОРИСТАННЯ ДИФУЗІЙНИХ МОДЕЛЕЙ В ЕКОНОМІЦІ ТА УПРАВЛІННІ .....	737

**Замлинна Я.В.**

«Економічна кібернетика та Дата Сайнс», 5 курс  
Київський національний економічний  
Університет імені Вадима Гетьмана  
**Катуніна О.С.**, кандидат економічних наук,  
доцент, Київський національний економічний  
Університет імені Вадима Гетьмана

## **ПРИКЛАДНІ ЗАДАЧІ ВИКОРИСТАННЯ ДИФУЗІЙНИХ МОДЕЛЕЙ В ЕКОНОМІЦІ ТА УПРАВЛІННІ**

До основних інструментів генеративного моделювання належать варіаційні автоенкодера (Variational Autoencoder, VAE), генеративно-змагальні мережі (Generative Adversarial Networks, GAN) [1] та дифузійні моделі (Diffusion Models). Варіаційні автокодера (VAE) використовують варіаційний байєсівський підхід у латентному просторі та здебільшого застосовуються для відтворення даних і дослідження їхньої внутрішньої структури. GAN реалізують змагальну взаємодію генератора і дискримінатора, що дозволяє досягати високої якості зображень, хоча часто супроводжується нестабільністю навчання. Дифузійні моделі, засновані на зворотному стохастичному процесі, демонструють стабільність і здатність до генерації реалістичних зображень. У таблиці 1 наведено узагальнене порівняння зазначених підходів:

Таблиця 1. Порівняння основних підходів у генеративному моделюванні

Характеристика	<b>VAE (Variational Autoencoder)</b>	<b>GAN (Generative Adversarial Network)</b>	<b>Diffusion Models (DM)</b>
Принцип роботи	Байєсівський латентний простір, реконструкція	Змагання генератора та дискримінатора	Зворотний процес денойзингу (відновлення з шуму)
Якість генерації	Помірна, часто змазана	Висока, але нестабільна	Дуже висока, фотореалістична
Стабільність навчання	Висока	Часто нестабільна (mode collapse)	Висока, масштабована

Контрольованість	Через латентний простір	Обмежена, потребує кастомізації	Висока (текстові промпти, умови)
Швидкість генерації	Швидка	Швидка	Повільна (багато кроків)
Вимоги до обчислень	Помірні	Середні–високі	Високі, добре масштабується на GPU/TPU
Основні застосування	Реконструкція, аномалії, генерація латентного простору	Deepfakes, креативний контент, перетворення стилів	Текст-до-зображення, медицина, дизайн, симуляція
Приклади моделей	$\beta$ -VAE, VQ-VAE	DCGAN, StyleGAN, CycleGAN	DDPM, Stable Diffusion, DALL·E 2, Imagen

Дифузійні моделі набули популярності як інструмент ефективного створення мультимедійного контенту — зображень, аудіо й відео на базі зворотного дифузійного процесу [2]. Дифузійні моделі - це клас генеративних моделей, які вчаться поступово знімати шуми в даних шляхом реверсування процесу дифузії. Основна ідея полягає в тому, щоб почати з чистого шуму та ітеративно вдосконалити його до високоякісної вибірки з цільового розподілу [2]. В основі методу лежить поступове перетворення шуму в остаточне зображення за допомогою серії дифузійних кроків. Кожен крок має на меті наблизити шум до реального розподілу даних, що сприятиме видаленню шуму з зображення [4].

Створення реалістичних зображень на базі дифузійних моделей відбувається при використанні нейронної мережі типу U-Net, яка побудована за допомогою двовимірних CNN. Зображення, яке попадає в дифузійну модель, зменшується в розмірах, переходячи в так званий latent space, після чого до нього додається шум певною кількістю кроків. Маючи текстову вказівку, або запит, з блоку «Conditioning» шум видаляється, формуючи матрицю, яка потім, виходячи з latent space, перетворюється в повноцінне зображення [4].

Відомими дифузійними алгоритмами є Stable Diffusion, DALL-E та Midjourne, які поєднують як текстове представлення інформації, що притаманне великим мовним моделям (LLMs), так і здатність згорткових шарів знаходити та локалізувати деталі на зображеннях [4]. Перевагами дифузійних моделей є такі:

- Дифузійні моделі здатні генерувати дуже реалістичні та деталізовані зображення. У багатьох випадках вони перевершують інші методи, зокрема GAN, за візуальною якістю [5].
- На відміну від GAN, де генератор і дискримінатор змагаються між собою, що часто призводить до нестабільного навчання, дифузійні моделі не мають цієї змагальної динаміки, завдяки чому навчання проходить стабільніше [5].
- Обчислення в дифузійних моделях можна легко розпаралелювати, особливо на сучасних графічних процесорах (GPU — *Graphics Processing Unit*) або тензорних процесорах (TPU — *Tensor Processing Unit*). Це дозволяє тренувати великі моделі швидше [6].
- Дифузійні моделі можна побудувати на основі різних нейромережових архітектур, наприклад, U-Net, що дозволяє адаптувати дифузійні моделі до різних типів даних: зображення, аудіо, текст тощо [6].

У економіці та управлінні дифузійні моделі демонструють високу ефективність (Таблиця 2) у низці прикладних задач, зокрема: автоматизованій генерації візуального контенту для бізнес-презентацій і звітності; створенні рекламних матеріалів і маркетингових креативів на основі текстових запитів; візуалізації складних фінансових або логістичних даних у доступній та естетичній формі; симуляції сценаріїв ринкової поведінки з метою стратегічного планування; генерації *synthetic data* для навчання або тестування економетричних моделей; а також побудові інтерактивних середовищ для освіти, аналітики та віртуальної співпраці.

Таблиця 2. Прикладні напрями використання дифузійних моделей в економіці та управлінні

Завдання / Галузь	Застосування дифузійної моделі	Очікуваний ефект / Результат
Бізнес-аналітика	Візуалізація аналітичних звітів за текстовими описами	Швидке створення інфографіки, підвищення сприйняття даних
Маркетинг і реклама	Генерація банерів, постерів і візуальних креативів за описом кампанії	Економія часу і витрат на дизайн, А/В тестування образів

Фінансове планування	Симуляція ринкових сценаріїв у візуальній формі	Краще розуміння ризиків і трендів
Електронне навчання та EdTech	Створення навчальних візуалізацій за короткими описами тем	Персоналізація навчання, зростання залученості
Управління персоналом	Візуалізація корпоративних структур або сценаріїв співпраці	Оптимізація управлінських процесів, комунікація змін
Розробка продукту	Швидке створення концептуальних зображень продуктів за текстовим описом	Прототипування ідей без участі дизайнерів
Економетричне моделювання	Генерація синтетичних даних для тренування моделей	Більше даних для точніших прогнозів

Проведене практичне дослідження включало генерацію зображень на основі текстових запитів за допомогою моделі Stable Diffusion v1.5, яка є представником дифузійних моделей нового покоління. Модель була завантажена з відкритого репозиторію Hugging Face та розгорнута у середовищі Python із використанням бібліотеки `diffusers`. У якості прикладного сценарію було задано текстову інструкцію (*prompt*): «*A futuristic city at sunset, flying cars, neon lights, ultra detailed, 4K*». Робочий процес включав такі етапи:

- ініціалізацію моделі та підключення авторизаційного токена;
- вибір відповідного обчислювального пристрою (CPU або GPU);
- генерацію зображення з подальшим його візуальним аналізом і збереженням.

Генерація здійснювалася за допомогою латентного дифузійного процесу: випадковий шум поступово трансформувався у семантично змістовне зображення, згідно з умовою текстового запиту. Ключовим компонентом архітектури моделі виступала неймережа на базі U-Net, яка реалізує зворотну трансформацію з латентного простору шляхом ітеративного усунення шуму.

Результатом серії експериментів стала низка альтернативних варіантів зображень, що якісно та деталізовано відтворюють зміст текстового запиту. На Рис. 1 представлено одне з синтезованих зображень з ознаками футуризму, неоновим підсвічуванням та виразною атмосферністю, що демонструє здатність дифузійної моделі ефективно інтерпретувати високорівневі текстові дескриптори та відображати їх у візуальній формі.



Рисунок 1. Згенероване зображення промту «*A futuristic city at sunset, flying cars, neon lights, ultra detailed, 4K*»

*Джерело: розроблено автором*

Крім основної генерації, було додатково протестовано можливість варіювання параметрів генерації, зокрема `num_inference_steps` та `guidance_scale`, що безпосередньо впливають на чіткість, деталізацію та відповідність зображення текстовому опису. Це дало змогу оцінити гнучкість моделі у адаптації під різні цілі: від швидкого чернеткового ескізу до високоякісного художнього рендеру. Збільшення кількості кроків покращувало деталізацію, але вимагало більше обчислювальних ресурсів, що підтверджує необхідність балансування між якістю результату та продуктивністю системи.

У дослідженні продемонстровано, що зміна навіть одного ключового слова у запиті («*futuristic*» на «*cyberpunk*»), або додавання стилістичних маркерів як-от «*digital art*», «*isometric*», «*concept art*») кардинально змінює композицію сцени. Це відкриває широкі можливості для інтерактивного дизайну: користувач може поступово коригувати промт і миттєво отримувати візуальний фідбек, фактично створюючи адаптивне середовище для візуального прототипування — без потреби у залученні художника. Такий підхід може знайти застосування у швидкій розробці контенту та ідей для економічних застосувань, а також створенні ігрових рівнів, інтерфейсів, у галузі освіти та маркетингу тощо.

Практичне застосування дифузійних моделей засвідчує їх високу ефективність у завданнях генерації зображень на основі текстових інструкцій. Висока якість, стабільність навчання та універсальність дифузійних моделей забезпечують хню значущість у сучасних дослідницьких і прикладних проєктах, особливо в креативних індустріях, дизайні, освіті та мультимедійній продукції.

Очікується, що дифузійні моделі й надалі розширюватимуть сфери свого використання, стаючи невід'ємною частиною повсякденної реальності.[3].

На основі проведеного дослідження можна зробити висновок про широкі перспективи використання дифузійних моделей у таких галузях, як:

- генерація візуального контенту у сфері дизайну, ігор та кіноіндустрії;
- покращення медичних зображень (наприклад, відновлення томограм);
- автоматизоване створення навчальних матеріалів;
- створення нових засобів взаємодії у віртуальній та доповненій реальності.

Таким чином, дифузійні моделі є не лише інструментом генерації зображень, а й ефективною технологічною платформою для побудови інтелектуальних систем для широкого використання в економіці та управлінні.

#### ***Список використаних джерел:***

1. Gupta, R., Tiwari, S., & Chaudhary, P. (2025). *Generative AI: Techniques, Models and Applications*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-82062-5>
2. Unite.AI. (n.d.). *Розуміння дифузійних моделей: глибоке занурення в генеративний ШІ*. Retrieved May 16, 2025, from <https://www.unite.ai/uk/understanding-diffusion-models-a-deep-dive-into-generative-ai/>
3. Unite.AI. (n.d.). *Дифузійні моделі в ШІ – усе, що вам потрібно знати*. Retrieved May 16, 2025, from <https://www.unite.ai/uk/diffusion-models-in-ai-everything-you-need-to-know/>
4. Малиш, О. (n.d.). *Використовуємо CNN для обробки зображень. Частина перша*. DOU. Retrieved May 16, 2025, from <https://dou.ua/forums/topic/48368/>
5. Kaddoura, S. (2023). *A Primer on Generative Adversarial Networks*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-32661-5>
6. Raut, R., et al. (2023). *Generative Adversarial Networks and Deep Learning*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003203964>



# СЕРТИФІКАТ

№ 13-2025/003

що засвідчує участь

## Яни Замлинної

у XIII ВСЕУКРАЇНСЬКІЙ НАУКОВО-ПРАКТИЧНІЙ КОНФЕРЕНЦІЇ  
ФОРУМІ МОЛОДИХ ЕКОНОМІСТІВ-КІБЕРНЕТИКІВ

«МОДЕЛЮВАННЯ ЕКОНОМІКИ: ПРОБЛЕМИ, ТЕНДЕНЦІЇ, ДОСВІД»

Професійна програма стажування  
12 год (0,4 кредиту ECTS)



Декан економічного факультету  
Ростислав МИХАЙЛИШИН

21-22 листопада 2025 року  
м. Львів

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ІВАНА ФРАНКА  
Економічний факультет  
Кафедра економічної кібернетики



# ТЕЗИ ДОПОВІДЕЙ

ХІІІ ВСЕУКРАЇНСЬКА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ  
ФОРУМ МОЛОДИХ ЕКОНОМІСТІВ-КІБЕРНЕТИКІВ

**«МОДЕЛЮВАННЯ ЕКОНОМІКИ:  
ПРОБЛЕМИ, ТЕНДЕНЦІЇ, ДОСВІД»**

21-22 листопада 2025 року, Львів

## ЗМІСТ

Підхормний О.М. Деркач Д.О.	Кібернетичні принципи стійкості фінансової системи в умовах інформаційної асиметрії	11
Замлинна Я.В.	Модель LOGIB як інструмент виявлення та оцінки гендерних розривів оплати праці	13
Бойко А.І.	Моделі поведінки вступників при виборі ЗВО: типологія і соціально-демографічні детермінанти	15
Дубина І.О. Чеверда С.С.	Цифровізація малих транспортних підприємств як фактор підвищення економічної стійкості в умовах воєнного стану	17
Синицький А.П.	Оптимізація маркетингових кампаній у ритейлі	19
Зьола М.Б.	Впровадження програмних рішень як інструмент оптимізації діяльності місцевих адміністрацій	21

### СЕКЦІЯ 1 МОДЕЛЮВАННЯ ЕКОНОМІКО-ВИРОБНИЧИХ СИСТЕМ

Богданова Г.С.	Довіра і соціально-економічний розвиток: статистичний аналіз взаємозв'язків показників	23
Буртяк І.В. Леськів В.В.	Оцінка досягнень циркулярної економіки за допомогою реляційного аналізу Грея	25
Джоші О.І.	Логічна модель бази даних цифрової платформи для набуття навичок обізнаності та реалізації цілей сталого розвитку	27
Гой Г.В. Юнькова О.О.	Структурне моделювання впливу податкових інструментів та інновацій на ефективність бізнесу	29
Нікітенко О.Ю. Гур'янова Л.С.	Розробка моделей аналізу ефективності контент стратегії Youtube каналу	31
Moloko Y.R. Paslavska I.M.	Analysis of tacit collusion in ukraine's oligopolistic fuel market using game theory	32
Косован О.В.	Перспективи застосування імітаційного моделювання у роздрібній торгівлі	35
Троневиц К.А.	Побудова матриці трасування вимог і тест-кейсів в бізнес-аналізі на прикладі підприємства «Coca-Cola»	36
Алексеев О.В.	Оптимізація структури виробництва у фермерському господарстві засобами економіко-математичного моделювання	39
Музика І.П.	Побудова моделей рекомендаційних систем в електронній комерції	41
Мирка П.В.	Особливості побудови концептуальної моделі бізнес-аналізу на прикладі Netflix	43
Гречина А.Р.	Імітаційне моделювання оптимізації виробничих ресурсів підприємства в умовах невизначеності	44
Тіцька А.Я.	Моделювання фінансового стану підприємства в умовах економічної нестабільності	46
Телішевська Р.О. Мельничук В.А.	Аналіз чинників формування лояльності споживача	48
Чеверда С.С. Бондарь Л.С.	Гібридний підхід до планування ресурсів: комбінація евристичних та точних методів оптимізації для мультипроектного середовища	50

**Замлинна Я.В.**

**Електронна пошта:** [yana.zamlyinna@gmail.com](mailto:yana.zamlyinna@gmail.com)

**Науковий керівник:** д.ф.-м.н, професор Притоманова О.М

**Електронна пошта:** [prytomanova.olga@kneu.edu.ua](mailto:prytomanova.olga@kneu.edu.ua)

*Київський національний економічний університет імені Вадима Гетьмана*

**МОДЕЛЬ LOGIB ЯК ІНСТРУМЕНТ ВИЯВЛЕННЯ ТА ОЦІНКИ  
ГЕНДЕРНИХ РОЗРИВІВ ОПЛАТИ ПРАЦІ**

**Zamlyinna Y.**

**THE LOGIB MODEL AS A TOOL FOR IDENTIFYING  
AND ASSESSING GENDER PAY GAPS**

Logib — це безкоштовний швейцарський онлайн-інструмент, розроблений Федеральним офісом з питань гендерної рівності для аналізу заробітної плати та сприяння рівній оплаті праці між чоловіками та жінками. Він допомагає компаніям оцінювати свої практики оплати праці, виявляти гендерні відмінності в заробітній платі та забезпечувати відповідність вимогам щодо рівної оплати[1].

Застосування Logib дозволяє підвищити прозорість у сфері винагороди працівників, зміцнити довіру між роботодавцем і персоналом, а також створити передумови для справедливого та обґрунтованого прийняття управлінських рішень.

Дана модель не просто цікавий та сильний інструмент. Вона також дозволяє дотримуватись законодавчих вимог. Для прикладу у Швейцарії компанії з кількістю працівників, що перевищує 100 зобов'язані провести аналіз рівності оплати (equal pay analysis) з огляду на нову редакцію закону про гендерну рівність [2]. За даними ОЕСР запровадження обов'язкового звітування про оплату праці діє у 21 країні. Раннє виявлення дисбалансу допомагає уникнути судових ризиків чи штрафів.

Безпосередньо в Україні немає чіткого закону, що зосереджено вказує на необхідність дотримання гендерного балансу на підприємствах. Logib узгоджується з європейськими директивами та рекомендаціями Міжнародної організації праці, тому її впровадження сприяє наближенню України до практик Європейського союзу у сфері соціальної рівності.

У своїй роботі я досліджувала один із навчальних закладів України. Оцінка гендерної нерівності починається із збору інформації та підготовки її відповідно наданому шаблону, який продемонстрований на сайті Logib. Заповнений варіант продемонстровано на рис. 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Q	R	S	T	U	V	W		
Key	Personnel number (ID, etc.)	Age	Gender	Years of service	Education and training	Function	Level of experience	Professional position	Employee pay monthly	Employee paid hourly	Basic wage	Allowance	13th monthly wage	Weekly working hours	Usual weekly working hours in company	Statistical population	Remarks	Additional information1	Additional information2	Additional information3	Additional information4	Additional information5		
1	1	55	F	31	1	Principal	1	1	100	0	11282	3384,5	940,13	0										
2	2	45	M	22	1	Vice Principal (Academic)	1	1	100	0	10718	3215,3	893,13	0										
3	3	48	M	26	1	Vice Principal (Vocational)	1	1	100	0	10718	3215,3	893,13	0										
4	4	38	F	16	1	Vice Principal (Pastoral & Educational)	1	1	100	0	10153	2030,7	846,12	0										
5	5	48	M	25	1	Head of Mechanical Engin	2	2	100	0	7978,3	2393,5	664,86	0										
6	6	51	F	24	1	Head of Electrical Engineer	2	2	100	0	7978,3	2792,4	664,86	0										
7	7	34	M	11	1	Head of Agronomy & Agri	2	2	100	0	7978,3	797,83	664,86	0										
8	8	56	F	35	1	Head of Technology & App	2	2	100	0	7978,3	2393,5	664,86	0										
9	9	38	F	15	1	Placement and Apprentice	2	2	100	0	7450,3	2235,1	620,86	0										
10	10	46	F	34	1	Extracurricular Activities C	3	3	100	0	6396,5	639,65	533,04	0										
11	11	46	M	25	1	Head of Part-time & Dista	2	2	100	0	7978,3	2393,5	664,86	0										
12	12	44	F	15	1	Curriculum and Quality Ass	2	2	100	0	7978,3	2393,5	664,86	0										
13	13	56	M	35	1	Head of Physical Educatic	3	3	100	0	7978,3	3988,6	664,86	0										
14	14	54	F	29	1	Student Mentor	3	4	100	0	6396,5	639,65	533,04	0										

Рисунок 1 – Заповнений шаблон даних для моделі Logib

21-22 листопада 2025 року, Львів

Джерело: розроблено автором на основі даних штатного розпису навчального закладу

У колонках вказується ключовий номер працівника (у своєму випадку взяла порядковий), вік, гендер; стаж роботи на досліджуваному місці праці. Також вказується назва посади, відсоток робочого часу, заробітна плата та можливі доплати. Вільні колонки у даному випадку можна заповнити при наявності особливої інформації, що потребує особливого аналізу.

Після завантаження файлу в систему та уточненні певних деталей розпочинається запуск регресії. Запуск регресії - це статистичний метод, який поміщає всю інформацію в "коробку", аналізує кореляції та взаємодії між цими змінними та визначає вагу (важливість), яку має кожна змінна для пояснення результату, в даному випадку вагу, надану таким змінним, як "освіта", для пояснення "заробітної плати".

Після проведеного аналізу формується декілька файлів із описами та висновками. Один із результатів візуалізації представлено на рис. 2.

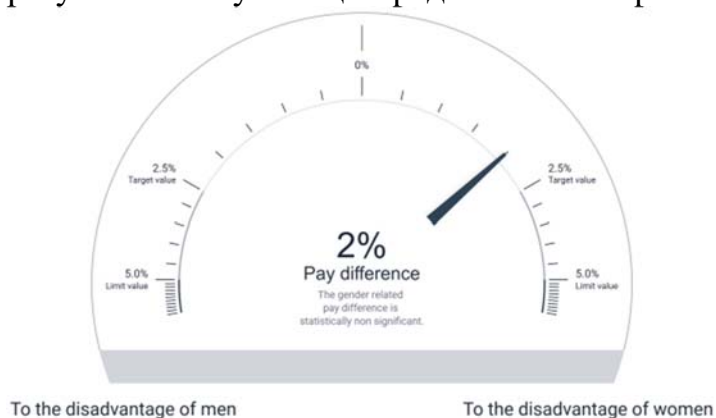


Рисунок 2 – Результат оцінки гендерної нерівності у навчальному закладі  
Джерело: розроблено автором за допомогою моделі Logib [3]

Усі розрахунки виконано за стандартною аналітичною моделлю Конфедерації (модуль 1) відповідно до офіційної методики [3]. Вони базуються на стандартизованому показнику заробітку 214 працівників, що працюють повний робочий день, з яких 107 (50,0%) жінок та 107 (50,0%) чоловіків у звітному місяці вересні 2025 році. У середньому жінки заробляють на 0,1% менше, ніж чоловіки. З урахуванням особистої кваліфікації та характеристик робочого місця, жінки заробляють на 2,0% менше. Ця різниця не є статистично значущою. Виявлені відмінності в оплаті праці можна пояснити виключно взаємодією різних об'єктивних факторів, таких як вік, освіта та професійна посада.

#### Література:

1. CDE Network. *Welcome to the SDC Climate, DRR & Environment Network*. URL: <https://www.sdc-cde.ch/en/nsb?id=103518> (дата звернення: 10.11.2025).
2. Federal Constitution and Gender Equality Act. Eidgenössisches Büro für die Gleichstellung von Frau und Mann (EBG). URL: <https://www.ebg.admin.ch/en/federal-constitution-and-gender-equality-act> (дата звернення: 10.11.2025).
3. Equal pay analysis with Logib. Logib. URL: <https://logib.legacy.admin.ch/home> (дата звернення: 11.11.2025).