

УДК 373.3:37.015.31:57.081.1

DOI: [https://doi.org/10.33272/2522-9729-2023-6\(213\)-18-24](https://doi.org/10.33272/2522-9729-2023-6(213)-18-24)



**Чуб Ольга Ігорівна,**

кандидатка економічних наук, доцентка кафедри теоретичної та прикладної системотехніки, Харківський національний університет імені В. Н. Каразіна, Україна

**Olga Chub,**

Candidate of Economic Sciences, Associate Prof. of the Department of Theoretical and Applied System Engineering, V. N. Karazin Kharkiv National University, Ukraine

**E-mail:** [o.i.chub@karazin.ua](mailto:o.i.chub@karazin.ua)

**ORCID iD:** <https://orcid.org/0000-0002-1216-856X>



**Новожилова Марина Володимирівна,**

докторка фізико-математичних наук, професорка, завідувачка кафедри комп'ютерних наук та інформаційних технологій, Харківський національний університет міського господарства імені О. М. Бекетова, Україна

**Maryna Novozhylova,**

Doctor of Sciences (Mathematics), professor, Head at the Department of Computer Science and Information Technologies, O. M. Beketov National University of Urban Economy in Kharkiv, Ukraine

**E-mail:** [marina.novozhylova@kname.edu.ua](mailto:marina.novozhylova@kname.edu.ua)

**ORCID iD:** <https://orcid.org/0000-0002-9977-7375>

## СТВОРЕННЯ ХМАРНОГО СЕРЕДОВИЩА ВИВЧЕННЯ МАТЕМАТИЧНОГО ІНСТРУМЕНТАРІЮ DATA SCIENCE

**A** Статтю присвячено дослідженню та класифікації можливостей хмарних сервісів і ресурсів реалізації математичного інструментарію Data Science в умовах розвитку технологій онлайн-навчання. Метою дослідження цього питання – розкрити потенціал хмарних цифрових інструментів в освітньому процесі та визначити їх переваги в онлайн-навчанні за умов динамічного зовнішнього середовища. Аналіз попередніх досліджень щодо проблеми, що розглядається, виявив низку невирішених питань у даній предметній області. В якості основного елемента структури хмарного середовища пропонується використання хмарного засобу Google Colab. У роботі розглядаються принципи кооперації Google Colab із можливостями наукових бібліотек Python і таких солверів як OR-tools Google Colab та інших. Застосування даного підходу уможливорює досліднику даних зосередитися на етапі моделювання задачі, тоді як процес кодування значно спрощується.

**Ключові слова:** хмарні технології; онлайн-освіта; Data Science; soft skills; навчання впродовж життя

### Creation of a cloud environment for studying Data Science mathematical tools

**S** This article is devoted to the study and classification of the possibilities of cloud services and resources for the implementation of the mathematical tools of Data Science in the conditions of the development of online learning technologies. The purpose of the study of this issue is to reveal the potential of cloud-based digital tools in the educational process and to determine their advantages in online learning under the conditions of a dynamic external environment. The analysis of previous studies on the problem under consideration revealed a few unresolved issues in this subject area. As the main element of the structure of the cloud environment, the use of the cloud tool Google Colab is proposed. The paper examines the principles of Google Colab cooperation with the capabilities of scientific Python libraries and such solvers as OR-tools Google Colab and others. Using this approach allows the data scientist to focus on the modeling phase of the problem, while the coding process is greatly simplified.

*Considering knowledge of the fundamental mathematical tools of Data Science as one of the most important competencies of a data scientist, the task of comprehensive study of mathematical foundations and the corresponding software becomes of primary importance.*

*The article formulates the main requirements for the structure of the cloud information educational environment. Numerical experiments realized on multiple sets of input data for various classes of optimization problems and regression analysis problems showed that the OR-Tools toolkit in combination with the possibility of using scientific Python libraries based on Google Colab allows creating a flexible information environment.*

**Keywords:** cloud technologies; online-education; Data Science; soft skills; lifelong learning

**Актуальність проблеми у загальному вигляді та її зв'язок із важливими практичними завданнями.** У другому десятиріччі XXI сторіччя в галузі освіти України відбуваються визначні зміни. Уже майже 3 роки Україна разом зі всім світом проходить через глобальні випробування. Зовнішні фактори сьогодення і пов'язані з цим глибокі зміни, що відбуваються в суспільстві, впливають на ринок праці в галузі інформаційних технологій (ІТ), його динаміку та тренди розвитку, викликають необхідність зміни форм і змісту навчання закладів вищої освіти. Сьогодні освіта в цілому, та ІТ-освіта зокрема [13], стрімко набувають як онлайн, так і дистанційну форму.

За даними сайту [8], із 2017 до 2027 року ринок онлайн-освіти виросте у чотири рази, при цьому необхідно відзначити, зростання обсягу (у грошовому вираженні) онлайн-освіти університетів обганяє зростання обсягу аудиторії онлайн освітніх платформ неформальної освіти майже втричі.

Ці тенденції безпосередньо торкнулися й реального сектору економіки в галузі інформаційних технологій, модифікуючи організаційні форми ведення бізнесу, вимоги до співробітників ІТ-компаній як уже працюючих, так і у перспективі.

Однією з таких тенденцій, яка формує довгострокову перспективу розвитку професійних вимог до професіонала в галузі ІТ, є аналіз даних за допомогою мов програмування високого рівня, а також інструментальних засобів інтелектуального аналізу з метою визначення корисної інформації для фірм, компаній, організацій різної форми власності. Ця сфера професійної діяльності належить до Data Science – науки про дані, яка є міждисциплінарною галуззю, що містить наукові методи, процеси і системи, які стосуються добування знань із даних, що зберігаються у різних формах як структурованих, так і неструктурованих [4].

При цьому, вибудовуючи організаційну структуру освітнього середовища, необхідно

зв'язати не тільки на відповідність попиту ринку праці щодо технічних навичок, професійних знань (hard skills), який залишається перманентно високим, змінюючись у деталях у розповсюджуючись на суміжні предметні галузі, але й на значне підвищення ролі гнучких навичок, «м'яких компетентностей» (soft skills).

Набуття професійних навичок (hard skills) у галузі Data Science передбачає володіння значним обсягом математичних знань, що складається з лінійної алгебри, теорії ймовірностей, статистики, дискретного аналізу, дослідження операцій, інтелектуального аналізу даних, штучного інтелекту та комп'ютерних методів моделювання даних.

Кожен проєкт Data Science містить етап формалізації потреб бізнесу, тобто побудови онтології предметної області дослідження, та вилучення вихідного набору даних, достатнього для подальшого оброблення.

**Аналіз попередніх досліджень і публікацій.** Приголомшливий прогрес у галузі науки про дані призвів до необхідності структуризації й упорядкування накопичених знань. У фундаментальному дослідженні [11] розглядається низка ключових для подальшого розвитку цієї галузі професійної діяльності питань: як навчити наступне покоління дослідників і науковців тим глибоким і нетривіальним дослідницьким моделям і методам, які вони мають використовувати при опрацюванні та керуванні даними? Як підтримати неперервний розвиток обчислювальних методів та інтеграцію даних у дослідження та практичну діяльність дослідника? У роботі на основі концепції життєвого циклу даних запропоновано інтелектуальну рамкову структуру для вирішення цих двох ключових питань і прийняття рішень щодо розвитку та провадження нових курсів Data Science (і яких саме), нових навчальних програм, визначення пріоритетів підтримки наукових досліджень даних у різних предметних областях. Розроблений у роботі Data Science цикл стосується не лише наборів даних, але й обчислювальних

дослідницьких інструментів, таких як системне та функціональне програмне забезпечення.

У публікації [10] проведено аналіз необхідних компетенцій для IT-професіоналів у галузі науки про дані (Data Science), яка є найпоказовішою, міждисциплінарною, що об'єднує такі галузі, як інформатика, аналіз даних і статистика, і поєднує їх з іншими спеціалізаціями, які можуть включати функції корпоративної підтримки, такі як управління ланцюгом поставок, структурування та управління даними, візуалізацію даних, управління базами даних та аналіз процесів, управління конфіденційністю та безпекою тощо. Визначено також, що оскільки науковці, які займаються даними, задіяні в різних секторах ринку, пов'язаних як із виробництвом продукції, так і з наданням послуг, загальні обов'язки можуть включати видобуток даних, аналіз збережених цільових даних і надання діагностичної, наказової, проспективної та прогнозної статистики для підтримки корпоративного управління щодо створення цінності за допомогою бізнес-аналітики.

Автори дуже цікавої роботи [6] розглядають проблему профілювання дослідників даних на ринку праці в галузі IT, використовуючи базу даних оголошень в Інтернеті про роботу в енергетичному секторі США з 2010 по 2019 рр. Систематизуючи аналітичний процес, який лежить в основі цілеспрямованого відображення навичок, які шукають роботодавці професіоналів із оброблення даних, автори пропонують інструментарій змістовного аналізу поточних і майбутніх потреб ринку праці разом із стратегічним прогнозуванням розвитку людського капіталу.

Автори вводять термін «Employability», тобто можливість працевлаштування як сукупність раніше набутих знань, навичок, компетенцій, досвіду та інших кваліфікацій фахівця, що підкріплюють його здатність бути надійним джерелом ефективності, інновацій і продуктивності для роботодавця.

У дослідженні [12] розглядається ще один аспект розвитку компетенцій дослідника даних – це навчання впродовж життя. Автори наголошують, що для того, щоб відповідати викликам і новим нормам XXI сторіччя, навчання впродовж життя – від шкільної освіти до професійного розвитку – має стати навичкою кожного професіонала.

При цьому навчання впродовж життя має бути керівною ідеєю для вчителів, педагогів, викладачів закладів освіти всіх рівнів у секторах формальної та неформальної освіти. На основі проведеного множинного лінійного регресійного аналізу побудовано моделі регресії для компетентностей викладачів у навчанні впродовж життя, а дисперсійний аналіз також використовувався для порівняння результатів моделювання.

Зазначимо, що знання математики є середіншого першим із вимірів узагальненої чотиривимірної моделі навчання Центру редизайну навчальних планів (Center for Curriculum Redesign (CCR) in Boston, MA) [3; 5], яка забезпечує організаційні рамки компетентностей, необхідних у XXI сторіччі. Цей вимір включає наступні знання як базові, що допомагають вирішувати повсякденні завдання: математична та фінансова грамотність, природничі знання, інформаційно-комп'ютерна, культурна та громадянська грамотність, так і професійні знання сучасної прикладної математики, традиційні й нові галузі математики, міждисциплінарності.

У роботі [2] розглянуто особливості використання хмарних технологій у процесі підготовки вчителів інформатики як фахівців, здатних проводити всі етапи наукових досліджень, включаючи збирання інформації, аналіз й оброблення інформації математичними методами, та представляти результати досліджень у різних формах.

Стаття [1] присвячена теоретичному аналізу основних можливостей цифрових освітніх сервісів, цифрових платформ і ресурсів в умовах розвитку технологій змішаного навчання.

**Виділення невирішених раніше частин загальної проблеми.** Повертаючись до сучасних умов навчання та виконання професійних завдань у цілому, необхідно відзначити, що збереження неперервності освіти за умов пандемії COVID-19 та повномасштабної війни в Україні став можливим тільки завдяки прогресу у галузі хмарних технологій. Приймаючи до уваги, що знання фундаментального математичного інструментарію Data Science є однією з найважливіших компетентностей дослідника даних, першорядного значення набуває задача комплексного вивчення математичних основ і відповідного програмного забезпечення.

**Метою статті є** визначення особливостей та основних вимог до структури хмарного інформаційного освітнього середовища, що забезпечує розвиток процесу викладання та вивчення математичного інструментарію Data Science в умовах онлайн-навчання.

**Викладення основного матеріалу дослідження.** Досвід авторів викладання множини дисциплін стосовно різних розділів сучасної прикладної математики в умовах забезпечення онлайн-освіти дозволив дослідним шляхом сформулювати основні вимоги до структури хмарного інформаційного освітнього середовища. Такими вимогами є:

- гнучкість у використанні;
- доступність сучасних програмних засобів моделювання та розв’язання задач дослідження;
- забезпечення неперервності навчання;

- мінімум програмних засобів, що використовуються;
- мінімум обчислювально-апаратних ресурсів учасників освітнього процесу;
- захист персональної інформації.

Виконання саме цих умов дозволяє забезпечити ефективність освітнього процесу, який проводиться в онлайн-форматі.

Нині спектр засобів програмно-апаратного забезпечення досить широкий, що уможлиблює різні конфігурації освітнього середовища. Керуючись досвідом використання різних налаштувань, можна запропонувати наступну структуру хмарного інформаційного освітнього середовища, що містить корпоративну систему закладу вищої освіти, середовище MS TEAMS, систему Moodle, які доповнюють один одного (рис. 1):

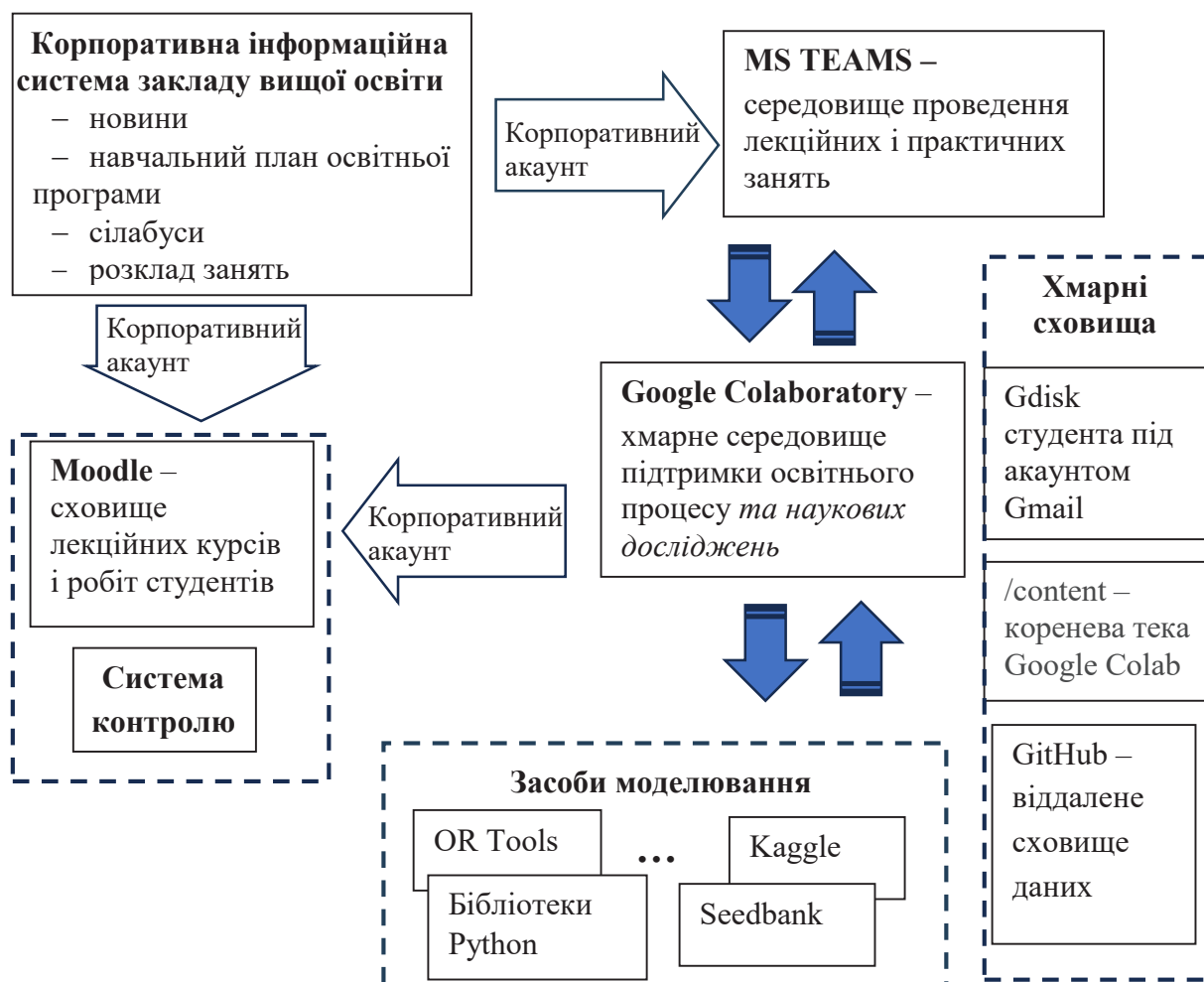


Рис. 1. Інфологічна модель хмарного освітнього середовища

Стороннім, але дуже важливим елементом цієї системи автори вважають застосування програмного середовища Google Colaboratory [7] як базового засобу забезпечення інтерактивності навчання, організації роботи в команді, набуття навичок розподілу задач та обов'язків при створенні програмного продукту за короткий визначений час.

Google Colaboratory або Colab [7] – це безкоштовний хмарний сервіс від Google Research, заснований на Jupyter Notebook (<http://www.jupyter.org>). Він містить довідник із документацією, а також кілька прикладів, щоб розпочати робити перші кроки, змінити вже створені коди та перейти до тестування.

Цей сервіс дозволяє будь-якому користувачеві з обліковим записом gmail писати та запускати вихідний код у своєму браузері. Основним робочим інструментом Google Colab є так званий блокнот (notebook).

Зокрема, Google Colaboratory підтримує мову програмування Python, яка має множину корисних бібліотек, орієнтованих на завдання машинного навчання, аналіз даних тощо.

Google Colaboratory – це інтегроване середовище розробки (Integrated development environment – IDE), що містить редактор початкового коду, а також інструменти для автоматизації складання та тестування програм.

Google Colaboratory не вимагає налаштування.

Необхідно зазначити, що коли користувач отримує доступ до Colab за допомогою свого облікового запису gmail, то одночасно реалізується віртуальна машина, на якій можна запускати свій код, ізольований від інших користувачів і ресурсів. Отже, можна відновити віртуальну машину до початкового стану у разі виникнення проблем. Однак при закритті браузера віртуальні машини будуть видалені після певного періоду бездіяльності, щоб звільнити ресурси. Тому потрібно зберігати свої блокноти з початковим кодом в Google Drive, щоб потім завантажити їх локально (формат Jupyter із відкритим кодом ipynb).

Головна особливість і перевага Google Colaboratory – можливість підключення безкоштовних потужних графічних процесорів GPU (забезпечує паралелізацію виконання завдань) і TPU (тензорний процесор, розробка Google, призначений для тренування

нейромереж), завдяки яким можна займатися не лише базовою аналітикою даних, а й складнішими дослідженнями в галузі оброблення великих об'ємів даних і машинного навчання.

Найвідомішими функціями та перевагами Google Colaboratory є:

- редагування та запуск коду мови Python і можливість працювати з Python-бібліотеками для аналізу даних онлайн;
- збереження програмних проєктів на Google Drive (GDrive);
- завантаження програмного коду з GitHub;
- можливість віддаленої сумісної роботи кількох користувачів у блокнотах із текстом, кодом, результатами та коментарями;
- імпорт блокнотів Jupyter або IPython;
- завантаження будь-яких блокнотів Colab локально з GDrive;
- потужні процесори для хмарних обчислень, у яких інтуїтивно зрозумілий інтерфейс, який дозволяє не перевантажувати комп'ютер файлами та процесами та робити всі обчислення швидко;
- доступ до облікового запису з файлами із будь-яких пристроїв.

Google Colaboratory як хмарний додаток, вимагає мінімум апаратних ресурсів. Ця програма працює в інтернет-браузерах як на настільних ПК, так і на мобільних пристроях. Ознакою нашого часу є застосування студентами своїх мобільних пристроїв під час навчання, тому середовище Colab незамінне у цьому випадку, дозволяючи студентам керувати навчанням за допомогою свого мобільного телефону. Крім того, їм не потрібні зовнішні пристрої зберігання, оскільки вони отримують доступ до безкоштовного хмарного сховища.

**Результати дослідження.** Для математичного моделювання та реалізації математичних розрахунків на основі отриманої математичної моделі в середовищі Google Colab можна залучати такі засоби розроблення програмного забезпечення як, наприклад, Google OR-Tools [9]. OR-Tools – це набір програмного забезпечення з відкритим кодом (ліцензія рache 2.0), розроблений Google для вирішення задач оптимізації, у т. ч., математичного програмування, маршрутизації транспортних засобів, задач про призначення тощо.

OR-Tools містить набір компонентів, написаних мовою C++, але, й це дуже зручно, надає оболонки

і для таких поширених мов програмування, як Java, NET і Python.

Одним із ефективних є розв'язувач MIP. Оболонка MIPSolver є складником OR-Tools і призначена для задач лінійного програмування та змішаного цілочисельного програмування (MIP – mixed integer programming). MIPSolver взаємодіє з інструментальним засобом SCIP – платформою для розв'язання цілочисельних задач із обмеженнями (CIP) і змішаних цілочисельних нелінійних задач.

Крім того, Google Colaboratory дозволяє використовувати Seedbank [7] – репозиторій прикладів інтерактивного машинного навчання, які можна запускати у своєму браузері без налаштувань як базу для розширення та створення власних проєктів та ідей студентів, що є технічною та інформаційною базою розвитку креативного мислення студентів.

**Висновки та перспективи подальших розробок.** Створено та протестовано запропоновану структуру хмарного інформаційного освітнього середовища, яке є відкритим фундаментом вивчення математичного інструментарію Data Science. Застосування визначених програмних засобів у сукупності створює синергетичний ефект.

Проведені численні експерименти на множинах наборів вхідних даних для різних класів оптимізаційних задач і задач регресійного аналізу показали, що інструментарій OR-Tools у поєднанні з можливістю застосування наукових бібліотек Python на базі Google Colab дозволяє створювати гнучке інформаційне середовище.

#### Список використаних джерел

1. Пономарьов О. С. Реалізація можливостей цифрових освітніх сервісів і ресурсів в умовах розвитку технологій змішаного навчання. *Педагогічні науки: теорія та практика*. 2023. № 2. С. 136–140.
2. Роганов М. Л. Використання хмарних технологій у процесі професійної підготовки майбутніх вчителів інформатики. *Духовність особистості: методологія, теорія і практика*. 2018. № 4 (85). С. 198–205.
3. Akimov N., Kurmanov N., Uskelenova A., Aidargaliyeva N., Mukhiyayeva D., Rakhimova S., Raimbekov B., Utegenova Z. Components of education 4.0 in open innovation competence frameworks: Systematic review. *Journal of Open Innovation: Technology, Market, and Complexity*. 2023. Vol. 9, № 2. 100037.
4. Dhar V. Data Science and Prediction. *Communications of the ACM*. 2013. Vol. 56, № 12. P. 64–73.
5. Four-Dimensional Education: The Competencies Learners Need to Succeed. URL: [https://curriculumredesign.org/our-](https://curriculumredesign.org/our-work/four-dimensional-21st-century-education-learning-competencies-future-2030/)

[work/four-dimensional-21st-century-education-learning-competencies-future-2030/](https://curriculumredesign.org/our-work/four-dimensional-21st-century-education-learning-competencies-future-2030/) (дата звернення 10.10.2023).

6. Google Colaboratory. URL: <https://colab.research.google.com>. (дата звернення 10.10.2023).
7. Google's new 'Seedbank' is home to ML project examples for developers to run & edit. URL: <https://9to5google.com/2018/07/16/seedbank-google-machine-learning-tensorflow/> (дата звернення 10.10.2023).
8. Online Education – Worldwide. URL: <https://www.statista.com/outlook/dmo/eservices/online-education/worldwide#revenue> (дата звернення 10.10.2023).
9. Route. Schedule. Plan. Assign. Pack. Solve. OR-Tools is fast and portable software for combinatorial optimization. URL: <https://developers.google.com/optimization> (дата звернення 10.10.2023).
10. Smaldone F., Ippolito A., Lagger J., Pellicano M. Employability skills: Profiling data scientists in the digital labour market. *European Management Journal*. 2022. Vol. 40, № 5. P. 671–684.
11. Stodden V. The Data Science Life Cycle: A Disciplined Approach to Advancing Data Science as a Science. *Communications of the ACM*. 2020. Vol. 63, № 7. P. 58–66.
12. Thwe W. Ph., Kálmán A. The regression models for lifelong learning competencies for teacher trainers. *Heliyon*. 2023. Vol. 9, № 2. P. 13749.
13. 50 Online Education Statistics: 2023 Data on Higher Learning & Corporate Training. URL: <https://research.com/education/online-education-statistics> (дата звернення 10.10.2023).

#### References

1. Ponomarov, O. S. (2023). Realizatsiia mozhlyvostei tsyfrovyykh osvitynykh servisiv i resursiv v umovakh rozvytku tekhnolohii zmishanoho navchannia [Realization of the possibilities of digital educational services and resources in the conditions of the development of mixed learning technologies]. *Pedahohichni nauky: teoriia ta praktyka [Pedagogical sciences: theory and practice]*, 2, 136-140 [in Ukrainian].
2. Rohanov, M. L. (2018). Vykorystannia khmarnykh tekhnolohii u protsesi profesiinoy pidhotovky maibutnykh vchyteliv informatyky [The use of cloud technologies in the process of professional training of future computer science teachers]. *Dukhovnist osobystosti: metodolohiia, teoriia i praktyka [Personal spirituality: methodology, theory and practice]*, 4 (85), 198-205 [in Ukrainian].
3. Akimov, N., Kurmanov, N., Uskelenova, A., Aidargaliyeva, N., Mukhiyayeva, D., Rakhimova, S., Raimbekov, B., & Utegenova, Z. (2023). Components of education 4.0 in open innovation competence frameworks: Systematic review. *Journal of Open Innovation: Technology, Market, and Complexity*, 9, 2, 100037.
4. Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM*, 56, 12, 64-73.
5. *Four-Dimensional Education: The Competencies Learners Need to Succeed*. Retrieved from <https://curriculumredesign.org/our-work/four-dimensional-21st-century-education-learning-competencies-future-2030/>.
6. *Google Colaboratory*. Retrieved from <https://colab.research.google.com>.

7. Google's new 'Seedbank' is home to ML project examples for developers to run & edit. Retrieved from <https://9to5google.com/2018/07/16/seedbank-google-machine-learning-tensorflow/>

8. Online Education – Worldwide. Retrieved from <https://www.statista.com/outlook/dmo/eservices/online-education/worldwide#revenue>

9. Route. Schedule. Plan. Assign. Pack. Solve. OR-Tools is fast and portable software for combinatorial optimization. Retrieved from <https://developers.google.com/optimization>

10. Smaldone, F., Ippolito, A., Lager, J., & Pellicano, M. (2022). Employability skills: Profiling data scientists in the

digital labour market. *European Management Journal*, 40, 5, 671-684.

11. Stodden, V. (2020). The Data Science Life Cycle: A Disciplined Approach to Advancing Data Science as a Science. *Communications of the ACM*, 63, 7, 58-66.

12. Thwe, W. Ph., & Kálmán, A. (2023). The regression models for lifelong learning competencies for teacher trainers. *Heliyon*, 9, 2, 13749.

13. 50 Online Education Statistics: 2023 Data on Higher Learning & Corporate Training. Retrieved from <https://research.com/education/online-education-statistics>

Дата надходження до редакції  
авторського оригіналу: 01.11.2023