DOI: https://doi.org/10.31392/NZ-udu-167.2026.06
УДК 37.012:004.6:37.091.3

Горбачук Василь Олександрович
кандидат педагогічних наук,
старший викладач кафедри інформаційних технологій і програмування
Українського державного університету імені Михайла Драгоманова
https://orcid.org/0000-0001-7858-5641
e-mail: gorbachuk.vas@gmail.com

DATAOPS У ПЕДАГОГІЧНИХ ДОСЛІДЖЕННЯХ З ІНФОРМАТИКИ: МЕТОДОЛОГІЯ ЗАБЕЗПЕЧЕННЯ ВІДТВОРЮВАНОСТІ ЕКСПЕРИМЕНТІВ

У статті актуалізовано фундаментальну проблему забезпечення відтворюваності результатів у сучасних педагогічних дослідженнях, зокрема в галузі методики навчання інформатики. Наголошено, що специфіка таких експериментів полягає у значній залежності підсумкових висновків від застосованого цифрового середовища, програмних інструментів, масивів навчальних даних, логів, а також процедур їх статистичної та аналітичної обробки. Зазначено, що традиційний підхід до висвітлення експериментів здебільшого обмежується загальним описом методики, залишаючи поза увагою відкритість даних, прозорість коду та алгоритми очищення вибірки, що суттєво ускладнює або унеможливлює незалежну перевірку наукових здобутків. Для розв'язання цієї проблеми запропоновано імплементувати методологію DataOps як інноваційний комплексний підхід до управління життєвим циклом даних педагогічного дослідження.

Теоретично обґрунтовано, що впровадження принципів DataOps у поєднанні з концепцією відкритої науки та FAIR-принципами дає змогу автоматизувати, тестувати й документувати всі етапи роботи з освітньою інформацією. У межах розвідки розроблено та представлено узагальнену модель DataOps-пайплайну педагогічного експерименту з інформатики. Запропонована модель охоплює логічну послідовність етапів: планування дослідження, збирання метрик із різних цифрових джерел (системи управління навчанням, Git-репозиторії, платформи автоматизованого тестування), перевірку якості, валідацію, очищення та етичну анонімізацію даних. Наступні кроки включають жорсткий контроль версій масивів і скриптів, відтворювану аналітичну обробку, візуалізацію, вичерпне документування й публікацію матеріалів у відкритому доступі. Також розроблено класифікацію рівнів зрілості педагогічного дослідження щодо його відтворюваності, яка масштабується від базового описового рівня до найвищого – повністю керованого на основі інфраструктури DataOps.

Підсумовано, що запропонований підхід виконує подвійну функцію: формує надійне технологічне підґрунтя для роботи з великими обсягами освітніх даних та слугує методологічною рамкою, яка докорінно підвищує прозорість, об'єктивність і наукову цінність педагогічних розвідок. Наголошено на потребі подальшої експериментальної апробації пайплайну в реальних навчальних курсах для формування дієвих методичних рекомендацій.

Ключові слова: *DataOps, педагогічні дослідження, навчання інформатики, відтворюваність, освітні дані, DataOps-пайплайн, навчальна аналітика.*

У сучасних педагогічних дослідженнях дедалі більшого значення набуває

проблема відтворюваності результатів. Це особливо актуально для досліджень з методики навчання інформатики, де експериментальні результати часто залежать не лише від педагогічних умов, а й від цифрового середовища, програмних інструментів, структури навчальних даних, способів оцінювання та процедур статистичної обробки. Відтворюваність педагогічного дослідження можна розглядати як здатність наукових результатів, отриманих під час педагогічного експерименту або спостереження, бути підтвердженими іншими дослідниками за умови використання аналогічних методів, умов, об'єктів і процедур аналізу.

Проблема полягає в тому, що значна частина педагогічних експериментів описується переважно на рівні загальної методики, тоді як дані, код аналізу, критерії очищення вибірки, способи обробки результатів, версії цифрових інструментів і повні матеріали дослідження часто залишаються недостатньо формалізованими. У дослідженнях з інформатики це створює додаткові труднощі, оскільки навчальні результати можуть фіксуватися у вигляді тестових балів, програмного коду, цифрових слідів у LMS, активності в репозиторіях, результатів автоматизованих перевірок або даних навчальної аналітики. Якщо ці дані не структуровані, не перевірені й не документовані, повторення експерименту або перевірка отриманих висновків стає ускладненою.

У цьому контексті перспективним є використання DataOps як методологічного підходу до управління життєвим циклом даних педагогічного дослідження. DataOps передбачає автоматизацію, тестування, моніторинг, контроль якості, контроль версій та узгоджену роботу з даними на всіх етапах їх використання. Для педагогічних досліджень з інформатики це відкриває можливість створення відтворюваних дослідницьких пайплайнів: від збирання освітніх даних до їх очищення, аналізу, візуалізації, інтерпретації та повторної перевірки результатів іншими дослідниками.

Проблема реплікації та відтворюваності в освітніх науках розглянута у праці M. Makel та J. Plucker [1]. Автори наголошують, що для розвитку освітніх досліджень важливими є не лише нові результати, а й повторна перевірка вже отриманих висновків. Підкреслено значення відтворюваних досліджень для підвищення надійності педагогічної науки, оскільки саме повторна перевірка дає змогу уточнювати, підтверджувати або спростовувати попередні висновки.

Безпосередньо проблему відтворюваного аналізу в освітніх дослідженнях розкрито у праці B. LeBeau, S. Ellison та A. Aloe [2]. Автори визначають відтворюваний аналіз як такий, за якого незалежний дослідник, використовуючи ті самі дані та той самий статистичний код, може отримати той самий результат. Важливим є акцент на script-based analysis, відкритих даних і прозорості аналітичних процедур. Це положення має принципове значення для педагогічних досліджень з інформатики, оскільки їх результати часто залежать від цифрових даних, програмних інструментів і процедур обробки, які мають бути доступними для перевірки.

Питання відкритого поширення даних в освітніх дослідженнях аналізують C. White, S. Estrera, C. Schatschneider та S. Hart [3]. Автори розглядають data sharing як важливий елемент відкритої науки і пропонують практичні поради для дослідників освіти щодо організації обміну даними протягом усього життєвого циклу дослідження. Це акцентує увагу не лише на підсумковому оприлюдненні результатів, а й на плануванні роботи з даними ще на етапі проєктування дослідження.

Класичним для осмислення відтворюваних досліджень в обчислювальній науці є дослідження R. Peng [4]. Проведений автором огляд показав, що відтворюване дослідження потребує доступності не лише кінцевих результатів, а й даних, коду та обчислювальних процедур, за допомогою яких ці результати були отримані. Це положення особливо важливе для педагогічних досліджень з інформатики, адже вони дедалі частіше спираються на цифрові інструменти, автоматизовану перевірку, програмний код і статистичне опрацювання навчальних даних.

Важливим методологічним підґрунтям для забезпечення відтворюваності є FAIR-принципи управління науковими даними, сформульовані M. Wilkinson та співавторами [5]. Відповідно до цих принципів, наукові дані мають бути findable, accessible, interoperable та reusable, тобто такими, що можуть бути знайдені, доступні, сумісні та придатні для повторного використання. У педагогічних дослідженнях з інформатики FAIR-принципи можуть бути покладені в основу вимог до освітніх даних: вони мають бути структурованими, описаними метаданими, придатними для повторного аналізу та коректної інтерпретації.

Питання підвищення надійності науки загалом розглянуто у праці M. Munafò та співавторів [6]. Автори пропонують комплекс заходів, спрямованих на підвищення відтворюваності: удосконалення методів, прозоре звітування, відкритість даних, коректне поширення результатів і зміни в системі оцінювання наукової діяльності. Для педагогічних досліджень ці положення є важливими, оскільки проблема відтворюваності пов’язана не лише з технічним збереженням даних, а й із культурою дослідницької прозорості.

Ідеї відкритої дослідницької культури розвинуто у праці B. Nosek та співавторів [7]. Автори наголошують на важливості відкритих даних, відкритих матеріалів, попередньої реєстрації досліджень, прозорого опису методів і доступності аналітичних процедур. У контексті педагогічних експериментів з інформатики ці підходи можуть сприяти підвищенню довіри до результатів, оскільки інші дослідники отримують можливість перевірити не лише висновки, а й сам шлях їх отримання.

Безпосередньо DataOps як підхід до управління даними розглянуто у систематичному огляді A. Fannouch, J. Gharib та Y. Gahi [8]. Автори аналізують DataOps через автоматизацію, інтеграцію даних, контроль якості, управління даними, agile-підходи та командну взаємодію. Підкреслено, що це дає змогу адаптувати DataOps до педагогічного експерименту: освітні дані мають не

просто збиратися, а проходити контроль якості, документування, контроль версій, тестування та аналітичну обробку.

Питання побудови DataOps-фреймворку для підвищення якості даних і ефективності пайплайнів розглянуто у праці C. Haertel та співавторів [9]. Дослідники зазначають, що DataOps поєднує практики DevOps із процесами автоматизації життєвого циклу даних, підвищення їхньої якості та надійності аналітичних пайплайнів. Це положення може бути безпосередньо використане для побудови моделі DataOps-пайплайну педагогічного дослідження з інформатики.

У дослідженні A. Haim, S. Shaw та N. Heffernan, присвяченому відкритій науці й практикам відтворюваності в освітньому дата-майнінгу, порушено питання відкритості даних, коду, методології та результатів у спільноті Educational Data Mining[10]. Цей огляд поєднує проблематику освітніх даних, цифрового аналізу, відкритої науки та відтворюваності.

Аналіз наукових джерел засвідчує, що проблема відтворюваності педагогічних досліджень має кілька взаємопов'язаних аспектів: методологічний, технічний, організаційний і етичний. Наявні праці достатньо обґрунтовують важливість відкритих даних, прозорого коду, FAIR-принципів і відтворюваного аналізу. Водночас недостатньо розкритим залишається питання використання DataOps саме як методології організації даних педагогічного експерименту з інформатики. Це зумовлює потребу в розробленні моделі DataOps-пайплайну, яка дозволить підвищити якість, прозорість і відтворюваність результатів педагогічних досліджень.

Мета статті: теоретично обґрунтувати можливості використання DataOps для забезпечення відтворюваності педагогічних досліджень з інформатики та запропонувати узагальнену модель DataOps-пайплайну педагогічного експерименту.

Для досягнення мети використано комплекс теоретичних методів: аналіз наукових джерел з проблем відтворюваності педагогічних досліджень, відкритої науки, FAIR-принципів, DataOps та освітнього дата-майнінгу; узагальнення підходів до організації життєвого циклу дослідницьких даних; порівняння традиційної та DataOps-орієнтованої організації педагогічного експерименту; моделювання DataOps-пайплайну педагогічного дослідження з інформатики.

Для забезпечення відтворюваності педагогічних досліджень з інформатики важливо розглядати експеримент не лише як сукупність педагогічних дій, а як повний дослідницький цикл, що охоплює планування, збирання даних, їх очищення, аналіз, інтерпретацію, документування та можливість повторної перевірки. Такий підхід узгоджується з принципами відтворюваного аналізу, відповідно до яких інший дослідник має змогу отримати ті самі результати, використовуючи ті самі дані та аналітичні процедури [2; 4]. У традиційній практиці педагогічного дослідження основна увага часто зосереджується на описі методики, вибірки, етапів експерименту

та підсумкових результатів. Водночас недостатньо деталізованими можуть залишатися процедури обробки даних, критерії виключення окремих записів, версії цифрових інструментів, структура наборів даних і статистичні скрипти [2].

У дослідженнях з інформатики ця проблема має особливу гостроту. Навчальні результати можуть бути представлені різнорідними даними: оцінками за лабораторні роботи, результатами автоматизованих тестів, кодом студентських програм, даними про кількість спроб, час виконання завдання, активність у репозитарії, результатами анкетування, логами LMS або показниками навчальної аналітики. Без єдиної системи управління такими даними дослідник може отримати статистично оброблені результати, але не забезпечити можливість їх повної перевірки іншими науковцями. Саме тому в освітніх дослідженнях особливого значення набувають відкритість даних, прозорість аналітичних процедур і можливість повторного використання дослідницьких матеріалів [3; 7; 10].

DataOps-орієнтований підхід дає змогу формалізувати роботу з даними педагогічного експерименту. Його доцільно розглядати як систему принципів і процедур, що забезпечують автоматизоване, контрольоване, документоване та відтворюване проходження даних через усі етапи дослідження. У педагогічному контексті DataOps може охоплювати: планування структури даних, визначення метаданих, автоматизоване збирання даних, перевірку їхньої повноти й коректності, очищення, анонімізацію, контроль версій, аналіз, візуалізацію, збереження коду й документації, а також підготовку матеріалів для повторної перевірки [8; 9].

Узагальнена модель DataOps-пайплайну педагогічного експерименту з інформатики передбачає послідовну організацію всіх етапів роботи з дослідницькими даними: планування дослідження, збирання освітніх даних, перевірку їхньої якості, очищення та анонімізацію, контроль версій даних і програмного коду, аналітичну обробку, візуалізацію результатів, документування, публікацію матеріалів і повторну перевірку отриманих висновків. Такий підхід дає змогу розглядати педагогічний експеримент не як одноразову процедуру збирання й аналізу даних, а як керований, документований і відтворюваний дослідницький процес, у межах якого кожен етап може бути перевірений, повторений або вдосконалений [2; 5; 8; 9].

Схематично запропоновану модель DataOps-пайплайну педагогічного експерименту з інформатики подано на рис. 1. Вона відображає основні етапи роботи з дослідницькими даними після планування експерименту: від збирання освітніх даних до перевірки їхньої якості, очищення, анонімізації, контролю версій, аналітичної обробки, візуалізації, документування, публікації матеріалів і повторної перевірки результатів. Така структура забезпечує керованість, прозорість і відтворюваність дослідницького процесу.

На етапі планування дослідження необхідно визначити мету експерименту, гіпотезу, вибірку, змінні, інструменти збору даних, критерії включення й виключення результатів, а також очікувані показники

ефективності. Для досліджень з інформатики такими показниками можуть бути рівень сформованості програмувальних умінь, якість коду, кількість успішних проходжень автотестів, динаміка спроб, швидкість виправлення помилок, успішність виконання проєктних завдань або рівень самостійності студентів.



Рис. 1. DataOps-орієнтована модель забезпечення відтворюваності педагогічного експерименту з інформатики

Етап збирання освітніх даних передбачає отримання інформації з різних джерел: електронного журналу, LMS, систем автоматизованої перевірки, Git-репозиторіїв, тестових платформ, анкет і спостережень. На цьому етапі важливо забезпечити узгодженість форматів, єдині ідентифікатори учасників, фіксацію часу збору даних і збереження первинних наборів без несанкціонованих змін. Саме збереження первинних даних дає змогу у майбутньому перевірити правильність їх очищення та обробки [3; 5].

Перевірка якості даних є одним із ключових етапів DataOps-пайплайну. Вона може включати контроль повноти записів, виявлення пропущених значень, перевірку діапазонів, пошук дублювань, виявлення технічних помилок, перевірку відповідності структури даних попередньо визначеній схемі. Наприклад, якщо дослідник аналізує результати лабораторних робіт з програмування, то необхідно перевірити, чи всі студенти мають однакову кількість записів, чи коректно збережені результати автотестів, чи не змішано дані різних груп або різних версій завдання [8; 9].

Етап очищення й анонімізації даних забезпечує етичну та методологічну

коректність дослідження. Освітні дані часто містять персональну інформацію, тому перед поширенням або повторним використанням вони мають бути анонімізовані. Водночас анонімізація не повинна руйнувати дослідницьку цінність даних: необхідно зберегти зв'язки між спробами одного студента, групою, етапом навчання, типом завдання та результатом. Саме тому важливим є документування всіх процедур очищення [3; 5; 7].

Контроль версій даних і коду забезпечує можливість відстежити, які саме дані й аналітичні скрипти були використані для отримання конкретних результатів. У дослідженнях з інформатики це може здійснюватися через Git-репозиторії, системи контролю версій, журнали змін, фіксацію версій програмного забезпечення та середовища виконання. Це важливо, якщо результати аналізу залежать від статистичних пакетів, мов програмування, бібліотек або платформ автоматизованої перевірки [2; 4].

Аналітична обробка має бути реалізована у максимально відтворюваній формі. Замість ручного виконання окремих операцій доцільно використовувати скрипти, які можна повторно запустити на тому самому наборі даних. Для цього можуть застосовуватися Python, R, Jupyter Notebook, Quarto, R Markdown або інші інструменти, що поєднують код, текстові пояснення, таблиці й візуалізації. Такий підхід дає змогу іншому досліднику не лише побачити підсумковий результат, а й перевірити логіку його отримання [2; 4].

Завершальними етапами DataOps-пайплайну є документування, публікація матеріалів і повторна перевірка результатів. Документація має містити опис вибірки, структури даних, змінних, процедур очищення, аналітичних скриптів, версій програмного забезпечення, обмежень дослідження та умов повторного використання матеріалів. Публікація може передбачати відкритий доступ до знеособлених даних, коду, інструкцій і навчальних матеріалів, якщо це не суперечить етичним вимогам і політиці конфіденційності [3; 5; 7; 10].

З метою оцінювання готовності педагогічного дослідження до відтворення доцільно запропонувати класифікацію рівнів зрілості. Перший рівень можна визначити як описовий: дослідження містить лише загальний опис методики та підсумкові результати. Другий рівень – частково документований: подано інструменти дослідження, критерії оцінювання та основні таблиці результатів. Третій рівень – дані доступні: дослідник надає знеособлений набір даних і пояснення до нього. Четвертий рівень – відтворюваний аналіз: разом із даними доступний код обробки, що дає змогу отримати ті самі результати. П'ятий рівень – DataOps-кероване дослідження: усі етапи життєвого циклу даних автоматизовані, задокументовані, проходять контроль версій, перевірені й придатні для повторного запуску. Запропонована класифікація узгоджується з ідеями відкритої науки, FAIR-принципами та підходами до відтворюваного аналізу [2; 5; 7; 10].

Застосування такої класифікації може бути корисним для педагогічних досліджень з інформатики, оскільки дає змогу оцінити не лише зміст експерименту, а й якість його дослідницької інфраструктури. Наприклад,

експеримент із впровадження нового підходу до навчання програмування може вважатися більш зрілим, якщо він містить не лише опис методики й результати тестування студентів, а й набір знеособлених даних, скрипти аналізу, опис автотестів, версії завдань і документацію до всіх етапів обробки.

Отже, DataOps у педагогічних дослідженнях з інформатики може виконувати подвійну функцію. По-перше, він забезпечує технологічну основу для якісного управління освітніми даними. По-друге, він виступає методологічною рамкою для підвищення прозорості та відтворюваності педагогічного експерименту. У цьому сенсі DataOps не замінює педагогічну методологію, а посилює її, забезпечуючи контрольовану, документовану й повторювану роботу з даними [8; 9].

Висновки. У статті теоретично обґрунтовано доцільність використання DataOps як методологічного підходу до забезпечення відтворюваності педагогічних досліджень з інформатики. З'ясовано, що проблема відтворюваності педагогічних експериментів пов'язана не лише з описом методики, а й із якістю даних, прозорістю процедур їх обробки, доступністю коду аналізу, документуванням змін і можливістю повторного виконання аналітичних процедур.

DataOps може бути адаптований до педагогічних досліджень як підхід до управління повним життєвим циклом освітніх даних. Він охоплює планування структури даних, їх збирання, перевірку якості, очищення, анонімізацію, контроль версій, аналітичну обробку, візуалізацію, документування та публікацію матеріалів дослідження. Такий підхід створює умови для підвищення прозорості педагогічного експерименту й полегшує повторну перевірку отриманих результатів.

Запропоновано узагальнену модель DataOps-пайплайну педагогічного експерименту з інформатики, що може бути використана під час дослідження ефективності навчання програмування, цифрових інструментів, автоматизованого оцінювання або навчальної аналітики. Окреслено також класифікацію рівнів зрілості педагогічного дослідження щодо відтворюваності: описовий, частково документований, рівень доступних даних, рівень відтворюваного аналізу та DataOps-керований рівень.

Перспективами подальших досліджень є розроблення тестового DataOps-пайплайну для конкретного курсу з інформатики, наприклад курсу програмування, експериментальна перевірка його ефективності, створення критеріїв оцінювання якості освітніх даних, а також розроблення методичних рекомендацій для дослідників і викладачів щодо підготовки відтворюваних педагогічних експериментів.

Використана література:

1. Makel M. C., Plucker J. A. Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 2014. P. 304-316. DOI : https://doi.org/10.3102/0013189X14545513
2. LeBeau B., Ellison S., Aloe A. M. Reproducible analyses in education research. *Review of Research in Education*, 45(1), 2021. P. 195-222. DOI : https://doi.org/10.3102/0091732X20985076

3. White C. M., Estrera S. A., Schatschneider C., Hart S. A. Getting started with data sharing: Advice for researchers in education. *Research in Special Educatio*n, 1, 2024. P. 1-15. DOI : https://doi.org/10.25894/rise.2604
4. Peng R. D. Reproducible research in computational science. *Science*, 334(6060), 2011. P. 1226-1227. DOI : https://doi.org/10.1126/science.1213847
5. Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 2016. Article 160018. DOI : https://doi.org/10.1038/sdata.2016.18
6. Munafò M. R., Nosek B. A., Bishop D. V. M., Button K. S., Chambers C. D., Percie du Sert N., Simonsohn U., Wagenmakers E.-J., Ware J. J., Ioannidis J. P. A. A manifesto for reproducible science. *Nature Human Behaviour*, 1, 2017. Article 0021. DOI : https://doi.org/10.1038/s41562-016-0021
7. Nosek B. A., Alter G., Banks G. C., Borsboom D., Bowman S. D., Breckler S. J., Buck S., Chambers C. D., Chin G., Christensen G., Contestabile M., Dafoe A., Eich E., Freese J., Glennerster R., Goroff D., Green D. P., Hesse B., Humphreys M., et al. Promoting an open research culture. *Science*, 348(6242), 2015. P. 1422-1425. DOI : https://doi.org/10.1126/science.aab2374
8. Fannouch A., Gharib J., Gahi Y. Enhancing DataOps practices through innovative collaborative models: A systematic review. *International Journal of Information Management Data Insights*, 5(1), 2025. Article 100321. DOI : https://doi.org/10.1016/j.jjimei.2025.100321
9. Haertel C., Sagavakar K. S., Staegemann D., Pohl M., Volk M., Turowski K. Designing a framework for DataOps: Improving data quality and pipeline efficiency in data science. *Procedia Computer Science*, 277, 2026. P. 2310-2319. DOI : https://doi.org/10.1016/j.procs.2026.02.268
10. Haim A., Shaw S., Heffernan N. How to open science: Promoting principles and reproducibility practices within the Educational Data Mining community. In Proceedings of the 16th International Conference on Educational Data Mining. 2023. DOI : https://doi.org/10.1145/3573051.3593398

R e f e r e n c e s :

1. Makel M. C., Plucker J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304-316. DOI : https://doi.org/10.3102/0013189X14545513 [in English].
2. LeBeau B., Ellison S., Aloe A. M. (2021). Reproducible analyses in education research. *Review of Research in Education*, 45(1), 195-222. DOI : https://doi.org/10.3102/0091732X20985076 [in English].
3. White C. M., Estrera S. A., Schatschneider C., Hart S. A. (2024). Getting started with data sharing: Advice for researchers in education. *Research in Special Education*, 1, 1-15. DOI : https://doi.org/10.25894/rise.2604 [in English].
4. Peng R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226-1227. DOI : https://doi.org/10.1126/science.1213847 [in English].
5. Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. DOI : https://doi.org/10.1038/sdata.2016.18 [in English].
6. Munafò M. R., Nosek B. A., Bishop D. V. M., Button K. S., Chambers C. D., Percie du Sert N., Simonsohn U., Wagenmakers E.-J., Ware J. J., Ioannidis J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, Article 0021. DOI : https://doi.org/10.1038/s41562-016-0021 [in English].
7. Nosek B. A., Alter G., Banks G. C., Borsboom D., Bowman S. D., Breckler S. J., Buck S., Chambers C. D., Chin G., Christensen G., Contestabile M., Dafoe A., Eich E., Freese J., Glennerster R., Goroff D., Green D. P., Hesse B., Humphreys M., et al. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. DOI : https://doi.org/10.1126/science.aab2374 [in English].
8. Fannouch A., Gharib J., Gahi Y. (2025). Enhancing DataOps practices through innovative collaborative models: A systematic review. *International Journal of Information Management Data Insights*, 5(1), Article 100321. DOI : https://doi.org/10.1016/j.jjimei.2025.100321 [in English].

9. Haertel C., Sagavakar K. S., Staegemann D., Pohl M., Volk M., Turowski K. (2026). Designing a framework for DataOps: Improving data quality and pipeline efficiency in data science. Procedia Computer Science, 277, 2310-2319. DOI : https://doi.org/10.1016/j.procs.2026.02.268 [in English].
10. Haim A., Shaw S., Heffernan N. (2023). How to open science: Promoting principles and reproducibility practices within the Educational Data Mining community. In Proceedings of the 16th International Conference on Educational Data Mining. DOI : https://doi.org/10.1145/3573051.3593398 [in English].

V. Horbachuk. DATAOPS in pedagogical research in computer science: methodology for ensuring reproducibility of experiments.

The article addresses the fundamental challenge of ensuring the reproducibility of results in contemporary pedagogical research, particularly within the field of computer science education. It highlights that the specificity of such experiments lies in the significant reliance of final conclusions on the digital environment, software tools, educational data sets, logs, and procedures for statistical and analytical processing employed. It is noted that the traditional approach to reporting experiments is largely confined to a general description of the methodology, often neglecting data openness, code transparency, and sample cleaning algorithms, which substantially hinders or renders independent verification of scientific findings impossible. To address this issue, the implementation of the DataOps methodology is proposed as an innovative, comprehensive approach to managing the data lifecycle in pedagogical research.

The study provides a theoretical substantiation that integrating DataOps principles with the concepts of Open Science and FAIR principles allows for the automation, testing, and documentation of all stages of working with educational information. Within the scope of this research, a generalized DataOps pipeline model for a computer science pedagogical experiment has been developed and presented. The proposed model encompasses a logical sequence of stages: research planning, collection of metrics from various digital sources (Learning Management Systems, Git repositories, automated testing platforms), quality assurance, validation, cleaning, and ethical anonymization of data. Subsequent steps include rigorous version control of datasets and scripts, reproducible analytical processing, visualization, comprehensive documentation, and open-access publication of research materials. Furthermore, a maturity classification for pedagogical research regarding its reproducibility has been developed, scaling from a basic descriptive level to the highest tier—fully managed through DataOps infrastructure.

It is concluded that the proposed approach serves a dual function: it creates a robust technological foundation for handling large volumes of educational data and serves as a methodological framework that fundamentally enhances the transparency, objectivity, and scientific value of pedagogical studies. The study emphasizes the necessity of further experimental validation of the pipeline in real-world educational courses to formulate actionable methodological recommendations.

Keywords: *DataOps, pedagogical research, computer science education, reproducibility, educational data, DataOps pipeline, learning analytics.*

Дата першого надходження рукопису до видання: 31.03.2026

Дата прийнятого до друку рукопису після рецензування: 13.04.2026

Дата публікації: 30.04.2026