

Додаток 1 до Договору № 234/0024
від « 01 » 08 2024 року
про надання грантової підтримки

ЗАТВЕРДЖУЮ
Проректор з наукової роботи
Київського національного університету
Імені Тараса Шевченка

Ганна ТОЛСТАНОВА



ТЕХНІЧНЕ ЗАВДАННЯ
до проміжного етапу проєкту з виконання наукового дослідження і розробки на 2024 рік
«Вирішення сучасних проблем хімії, біомедицини, фізики та матеріалознавства з
використанням центру високопродуктивних обчислень і машинного навчання»
(далі – Проєкт)

Назва конкурсу: «Дослідницькі інфраструктури для проведення передових наукових досліджень»

Реєстраційний номер Проєкту: 2023.05/0024

Науковий керівник проєкту (головний дослідник, principal investigator (PI)): Комаров Ігор Володимирович

Склад консультативної ради проєкту (ПІБ, науковий ступень, основне місце роботи)

- 1) Олександр ВАРНЕК (Alexandre Varnek, доктор хімічних наук (хабілітація), професор і керівник лабораторії хемоінформатики Страсбурзького університету, Франція);
- 2) Сергій АФОНІН (Sergiy Afonin, науковий співробітник Інституту органічної хімії Технологічного інституту Карлсруе, Німеччина);
- 3) Мирослав КАВАЦІЮК (Myroslav Kavatsyuk, викладач науково-технічного факультету Університету Гронінгена, Нідерланди);
- 4) Сергій НІРЕНБУРГ (Sergei Nirenburg, доктор філософії, професор когнітивних та комп'ютерних наук Політехнічного інституту Ренсселера в Трої, штат Нью-Йорк, США);
- 5) Сергій СВИСТУНОВ, кандидат технічних наук, завідувач відділу Комп'ютерного забезпечення наукових досліджень і науково-технічної інформації Інституту теоретичної фізики ім.М.М.Боголюбова НАН України.

Підстави для реалізації Проєкту з виконання наукових досліджень і розробок:

Рішення наукової ради Національного фонду досліджень України щодо визначення переможця конкурсу протокол від 27 червня 2024 року № 20

Декларація від установи грантоотримувача (вихідний номер, дата)

013/456 від 12 липня 2024 року

1. ЗАГАЛЬНА ІНФОРМАЦІЯ ПРО ПРОЄКТ

Короткий опис Проєкту (до 5000 знаків)

Сучасні наукові дослідження сьогодні важко собі уявити без застосування комп'ютерних технологій. На жаль, в забезпеченні ними науки в Україні спостерігається серйозне відставання від світового рівня. Даним проєктом передбачається створення нової дослідницької інфраструктурної одиниці – потужного комп'ютерного кластера КНУ ім. Т. Шевченка, обладнаного пакетами програм, та проведення на ньому наукових досліджень в галузі фізики ядра, елементарних частинок, високих енергій, хемо- та біоінформатики, штучного інтелекту та комп'ютерної лінгвістики. Зокрема, планується провести: 1. Симуляцію розпаду динуклонних систем (динейтрона) із застосуванням кодів GEANT, TALYS і скриптів Phyton, використовуючи машинне навчання, та проведення оптимізації відбору подій для визначення параметрів розпаду динейтрона, проведення теоретичного моделювання опису процесів розпаду двонуклонних систем динейтрона; 2. Напрацювання методів і підходів до прогнозування теплофізичних властивостей матеріалів і структур різної розмірності та морфології для підвищення наукового рейтингу команди виконавців і передумовою їхньої подальшої участі в престижних конкурсах наукових проєктів; 3. Інтеграцію VIRGO.UA у платформу Multi-Messenger Online Data Analysis (MMODA, <https://www.astro.unige.ch/mmoda/>) та отримання доступу до European Open Science Cloud (EOSC) чи його підструктур. VIRGO.UA виступатиме у ролі української сполучної ланки для MMODA. Центр буде обслуговувати веб-сервер, який дублює послуги з доступу до архівів даних та аналізу даних, а також моделювання від MMODA; 4. Визначення особливостей молекулярної структури, електронних взаємодій, реакційної здатності та механізмів реакцій синтезу молекул - потенційних будівельних блоків для біо- та медичної хімії. Будуть досліджені унікальні сполуки, синтезовані на основі діазоалканів із використанням сучасних технологій, таких як вітчизняні проточні (фото)реактори, які не мають аналогів у світі; 5. Прогнозування потенційної біологічної активності нових сполук, зокрема, сполук із структурних банків даних, використовуючи сучасні підходи хемоінформатики, молекулярного докінгу, молекулярної динаміки, розширеного семплінгу, штучного інтелекту; 6. Розроблення технології побудови великих нейронних моделей на прикладі великих лінгвістичних моделей з використанням потужностей створеного обчислювального кластера; 7. Адаптацію великих лінгвістичних моделей до інших мов, до розширеного лексикону інших предметних областей, до інших мовних реалій (жаргонізми, фігури мовлення, сталі вирази тощо); побудова методів тонкого налаштування великих лінгвістичних моделей під виконання прикладних задач, актуальних для держави і світу в цілому; дослідження побудови великих нейронних моделей нелінгвістичного спрямування.

Планується вперше отримати характеристики спектрометра, необхідного для проведення експерименту з розпаду динейтрону через канал електронного розпаду в дейтрон; використання потенціалів міжатоної взаємодії, параметризованих методами машинного навчання, для молекулярно-динамічного моделювання процесів теплового транспорту в різних матеріалах. Дослідження в астрофізиці високих енергій будуть проводитися у новітньому напрямі гравітаційно-хвильової та нейтринної астрономії, де реєструються події одночасно в різних спектральних діапазонах та через різні канали спостережень, що є новітньою методикою фізики астрономічних джерел. За допомогою квантової хімії досліджуватимуться нові класи органічних молекул з унікальною структурою – потенційні будівельні блоки для медичної хімії та агрохімії, а також механізми ефективних фотоініційованих та проточних промислових процесів їх синтезу. Комп'ютерний аналіз потенційної взаємодії структур-кандидатів із біологічними мішенями з використанням молекулярної динаміки та молекулярного докінгу сприятиме цілеспрямованого синтезу нових біологічно активних речовин – потенційних лікарських засобів та препаратів для агрохімії.

В галузі штучного інтелекту та машинного навчання планується розробка, адаптація та тонке налаштування великих нейронних лінгвістичних моделей і прикладних систем на їх основі. Новими складовими даної технології буде оптимізація обчислювальних процесів, генерація зв'язних текстів заданого змісту, а також розпізнавання в текстах та соцмережах пропаганди, недостовірних повідомлень, текстів, що мають ознаки ІПСО, а також ботів та осіб з небезпечною поведінкою.

Довгострокова мета проекту - надання науковцям КНУ та інших наукових та освітніх установ доступу до ресурсів кластера для реалізації наукових та освітніх задач. Умови колективного використання кластера сприятимуть професійному обслуговуванню, своєчасному технічному та програмному оновленню, в тому числі, на основі використання коштів зацікавлених користувачів. Це сприятиме консолідації наукової спільноти в різних галузях комп'ютерних досліджень, їх популяризації та кращій підготовці молодих спеціалістів із комп'ютерних методів в хімії, фізиці та біології та започаткуванню спільних міждисциплінарних досліджень.

Тривалість виконання Проекту

Початок, рік	2024
Завершення, рік	2026

Тривалість виконання проміжного етапу Проекту у 2024 році

Початок (відповідно до Календарного плану), місяць, рік:	Серпень 2024 р.
Завершення, місяць, рік:	Грудень 2024 р.

2. ІНФОРМАЦІЯ ПРО ВИКОНАВЦІВ ПРОЄКТУ

Загальна кількість авторів проекту	З них		Допоміжний персонал
	Доктори наук	Кандидати наук	
19	9	10	-

3. ТЕМАТИЧНИЙ НАПРЯМ КОНКУРСУ ЗА ЯКИМ ЗДІЙСНЮЄТЬСЯ РЕАЛІЗАЦІЯ ПРОЄКТУ (заповнюється відповідно до поданої заявки):

1. Фундаментальні наукові дослідження в галузі природничих наук та математики.
2. Фундаментальні дослідження в галузі наук про життя.
3. Прикладні дослідження, що мають на меті створення технологій (включно з ІТ) та продуктів, які ґрунтуються на новітніх досягненнях фундаментальної науки.

4. ОПИС ПРОЄКТУ (заповнюється відповідно до поданої заявки)

4.1. Мета Проекту (до 500 знаків)

Даним проектом передбачається створення нової дослідницької інфраструктурної одиниці – потужного комп'ютерного кластера КНУ ім. Т. Шевченка, обладнаного пакетами програм, та проведення на ньому наукових досліджень в галузі фізики ядра, елементарних

частинок, високих енергій, хемо- та біоінформатики, штучного інтелекту та комп'ютерної лінгвістики.

4.2. Основні завдання Проєкту (до 1000 знаків)

1. Виконання симуляції експериментів з розпаду динуклонних систем, використовуючи машинне навчання, та проведення обробки й оптимізації відбору подій для визначення параметрів розпаду динейтрона, проведення теоретичного моделювання опису процесів розпаду двонуклонних систем динейтрона
2. Напрацювання методів і підходів до прогнозування теплофізичних властивостей матеріалів і структур різної розмірності та морфології
3. Інтеграція VIRGO.UA у платформу Multi-Messenger Online Data Analysis
4. Визначення особливостей молекулярної структури, електронних взаємодій, реакційної здатності та механізмів реакцій отримання сполук - потенційних будівельних блоків для біо- та медичної хімії
5. Передбачення потенційної біологічної активності нових сполук, зокрема, сполук із структурних банків даних, використовуючи сучасні підходи хемоінформатики
6. Розроблення технології побудови великих нейронних лінгвістичних моделей
7. Розроблення методів адаптації великих лінгвістичних моделей до різних мов

ЕВП № 1 (2024 рік)

Завдання 1.1 Розробка моделі спектрометру для спостереження динейтрону з використанням кодів GEANT, TALYS, а також налагодження взаємодії розрахунків на базі Python скриптів. Підготовка програмних кодів для молекулярно-динамічного моделювання процесів теплового транспорту в твердотільних структурах, виконання тестових розрахунків коефіцієнта теплопровідності модельних наноструктур.

Отримання даних спостережень віддалено гамма- та нейтринних телескопів, для побудови фізичних моделей прискорення космічних променів та генерування багатоканальних сигналів від астрофізичних джерел (таких як Cygnus X та MGRO J1908+06). (КР3)

Завдання 2.1 Встановлення та тестування пакетів квантово-хімічних та програм для молекулярної механіки та динаміки, забезпечення віддаленого доступу до них, підготовка вхідних завдань. (КР4)

Встановлення комерційних пакетів програм та пакетів із безкоштовною ліцензією на комп'ютерний кластер. Підготовка та виконання серії розрахункових тестів. Оцінка продуктивності вузлів кластера та створення оптимального режиму доступу та виділення ресурсів для окремих класів задач із врахуванням умов ліцензій. Проведення тестових розрахунків структур реальних молекул в різних наближеннях. Підготовка вхідних даних для виконання завдань 4.2-4.4.

Завдання 3.1 Розроблення методів тонкого налаштування великих мовних моделей для виконання низки прикладних задач. (КР5)

Тонке налаштування великих мовних моделей є відносно простішою задачею, аніж побудова великих мовних моделей, тому виконання цих робіт доцільно планувати на перший етап проєкту. За базові мовні моделі будуть взяті наявні (BERT, GPT і т.п.). Роботи складаються із побудови навчальних наборів даних, відповідних прикладним задачам, навчання моделі та проведення експериментів із замірювання якості роботи моделі. Результатом цього етапу, окрім самих оптимізованих методів налаштування, буде низка прикладних програм обробки текстів, таких як системи класифікації та розпізнавання текстів пропаганди, недостовірних новин (fake-news) та інші.

Завдання 4.1 Запускові та тестові роботи по комплектації комп'ютерного кластеру. Проектування змін до архітектури обчислювального кластера ІОЦ з урахуванням встановлення нових вузлів

зберігання даних, вузлів віртуалізації та обчислювальних вузлів. Розробка, тестування та впровадження політик, системного програмного забезпечення та його конфігурації для забезпечення колективного використання нового типу вузлів з графічними прискорювачами. Встановлення та тестування коректної роботи у конкурентному режимі використовуваного у проєкті прикладного програмного забезпечення. Проєктування та реалізація схеми розміщення обладнання для забезпечення рівномірного навантаження на системи енергоживлення та забезпечення температурного режиму роботи обладнання. (KP1)

Завдання 5.1 Проведення установчої зустрічі з визначення порядку дії, підписання угод про нерозголошення. Проведення зустрічі з Консультативною радою. (KP2)

4.3. Детальний зміст Проєкту:

- Сучасний стан проблеми (до 2 сторінок)

Виконання наукових досліджень у різних галузях природничих наук вимагає ресурсоємного комп'ютерного моделювання з використанням потужних обчислювальних систем, потужність яких наближається до граничного можливої на даний час. Наявність високопродуктивного обчислювального ресурсу є необхідною умовою для кожної установи, яка претендує на отримання наукових результатів світового рівня у галузях фізики, хімії, біології, інформаційних технологій. У провідних країнах світу комп'ютерні дослідження є невід'ємною і надзвичайно важливою частиною наукового та технічного прогресу. Так, в Німеччині уже багато років існують і постійно оновлюються гігантські суперкомп'ютерні кластери загальноєвропейського рівня в м. Гархінг ("Rechenzentrum Garching", <https://www.mpie.de/4065158/Hardware>) та Юліх ("Jülich Supercomputing Centre", <https://jlsrf.org/index.php/lfs/article/view/167>). Не менш потужні кластери існують в Японії, США та Китаї. Тому створення в Київському національному університеті імені Тараса Шевченка (КНУ) власного потужного комп'ютерного кластера є важливим та невідкладним завданням. Воно стало би поштовхом для наукових досліджень, досвід яких вже є у науковців різних підрозділів університету і для виконання яких їм не треба було б шукати ресурсні можливості за кордоном. Такі дослідження вже започатковані у галузях ядерної фізики (фізичний факультет КНУ), комп'ютерної хімії та хемо- та біоінформатики (Навчально-науковий Інститут високих технологій КНУ), штучного інтелекту та комп'ютерної лінгвістики (факультет комп'ютерних наук та кібернетики КНУ).

На кафедрі ядерної фізики та високих енергій фізичного факультету вперше було отримано експериментальні дані щодо існування двонуклонних зв'язаних систем, динейтрона та дипротона, проф. І.М. Каденком [1-4]. Існування вказаних двонуклонних систем узгоджується з теоретичними передбаченнями А.Б. Мігдала [5]. Для подальших теоретичних досліджень потрібно буде розглянути багаточастинкову задачу з метою розрахунку значень рівнів енергій можливих енергетичних станів двонуклонних систем. Інша актуальна задача сучасної фізики і матеріалознавства - це створення нових матеріалів для контрольованого теплового транспорту в пристроях і схемах навітньої електроніки, оптотехніки, альтернативної енергетики [6-8]. Для вирішення цієї задачі необхідним є аналіз процесів теплового транспорту в матеріалах на атомарному рівні, реалізація якого можлива із залученням сучасних методів комп'ютерного моделювання та алгоритмів паралельних високопродуктивних обчислень. Також науковці КНУ імені Тараса Шевченка є залученими до проведення досліджень з використанням спостережень на провідних рентгенівських та гамма-телескопах сучасності з використанням Віртуальної Рентгенівської та Гамма-Обсерваторії (VIRGO.UA) – наукового центру астрофізики високих енергій. Зі створенням потужного комп'ютерного кластера КНУ буде повноправним учасником цих досліджень.

Комплекс задач у галузі теоретичної хімії та хемоінформатики, вирішення яких значно прискорилося б зі створенням кластера вже знаходиться у фокусі науковців Навчально-наукового Інституту високих технологій КНУ. З кінця 20-го – початку 21 сторіччя надзвичайно зросла роль теоретичної (квантової) хімії як джерела інформації про структуру сполук [9] та механізми

хімічних реакцій [10, 11]. Використання квантової хімії, молекулярної динаміки та молекулярного докінгу є все більш значущим та актуальним для моделювання взаємодій субстрат-рецептор, дослідження структури протеїнів, нуклеїнових кислот та інших наукових проблем біо- та медичної хімії, зрештою, для винайдення нових лікарських засобів та функціональних матеріалів [12]. Через відсутність власної розрахункової бази та ліцензійних програмних продуктів, для проведення таких досліджень на сучасному рівні викладачі та співробітники КНУ вимушено використовують власні обмежені можливості.

Ще однією з галузей, що бурхливо розвивається останніми десятиріччями і є надзвичайно перспективною для прискорення науково-технічного прогресу, є машинне навчання та штучний інтелект. У сучасних системах штучного інтелекту широко застосовують великі нейронні моделі, створення яких вимагає потужних обчислювальних ресурсів. Основою інтелектуальних методів обробки текстів є великі нейромережеві лінгвістичні моделі (Large Linguistic Models) такі, як Bert [13], ALBERT [14], RoBERTa [15], GPT-4 [16], ChatGPT [17], побудовані на основі архітектури типу «трансформер», які навчаються на великих текстах обсягом у сотні гігабайтів. Для створення конкретних прикладних лінгвістичних систем використовують донавчання наперед навченої великої лінгвістичної моделі або її тонке налаштування (fine tuning) до конкретної прикладної задачі. Донедавна дослідження та розробка моделей і методів орієнтувалися лише на англійські тексти. Для створення систем обробки текстів українською мовою необхідно будувати великі лінгвістичні моделі на основі україномовних текстів, а для цього, в свою чергу, необхідне застосування машин з якомога більшою кількістю відеокарт типу V100/A100 або їх аналогів, що й планується за проектом.

Наразі, наявні технічні та фінансові можливості університетів та наукових установ в Україні не дозволяють створювати власні достатньо потужні комп'ютерні центри для проведення конкурентних наукових досліджень, що відповідали б сучасним стандартам, без залучення грантового фінансування.

- Новизна Проекту (до 1 сторінки)

Планується вперше отримати характеристики спектрометра, необхідного для проведення експерименту з виміру розпаду динейтрона через канал електронного розпаду в дейтрон. Ще одним елементом новизни проекту стане використанням потенціалів міжтомної взаємодії, параметризованих методами машинного навчання, для молекулярно-динамічного моделювання процесів теплового транспорту в матеріалах різної морфології. Дослідження в астрофізиці високих енергій, що плануються за проектом, також будуть проводитися у новітньому напрямі гравітаційно-хвильової та нейтринної астрономії, де реєструються події одночасно в різних спектральних діапазонах та через різні канали спостережень. Ефективне використання різних типів багатоканальних (multimessenger) даних (наприклад, дані про нейтрино IceCube з даними гамма-випромінювання Fermi-LAT, H.E.S.S., MAGIC) є новітньою методикою фізики астрономічних джерел.

Дослідження за проектом в галузі хемоінформатики націлюється на створення нових класів органічних молекул з унікальною структурою – потенційних будівельних блоків для медичної хімії та агрохімії і, зокрема, на розробку сучасних ефективних фотоініційованих та перспективних проточних промислових технологій. Встановлення за допомогою комп'ютерної (квантової) хімії механізмів досліджуваних процесів дозволить отримати нові фундаментальні знання про їх перебіг. Аналіз структур-кандидатів за допомогою створеного комп'ютерного кластера на предмет їх потенційної взаємодії з біологічними мішенями з використанням сучасних комп'ютерних програм для молекулярної динаміки та молекулярного докінгу створить передумови для цілеспрямованого синтезу нових біологічно активних речовин – потенційних лікарських засобів та препаратів для агрохімії. У разі знаходження сполук-лідерів з оптимізованою активністю проведені комплексні дослідження дадуть змогу провести швидке та ефективне масштабування їх синтезу та швидше впровадження у виробництво.

Новизною проекту у галузі штучного інтелекту та машинного навчання є нова технологія розробки, розширення, адаптації та тонкого налаштування великих нейронних моделей на

прикладі створення великих лінгвістичних моделей і прикладних систем на основі цих моделей. Новими складовими технології буде оптимізація всіх обчислювальних процесів з урахуванням можливостей наявного устаткування обчислювального кластера. Також новими елементами технології є низка актуальних для сучасного світу прикладних задач, що будуть виконуватися побудованими системами: розпізнавання пропаганди в текстах; розпізнавання у соцмережах ботів та осіб з небезпечною поведінкою; розпізнавання недостовірних повідомлень (fake-news); розпізнавання у соцмережах та ЗМІ текстів, що мають ознаки ІІСО; генерація зв'язних текстів заданого змісту.

- Методологія дослідження (до 2 сторінок)

На початку роботи буде здійснено закупівлю обладнання, побудову енергозабезпечення та віртуалізації кластера, оптимізацію роботи комп'ютерного кластеру та запуск його в експлуатацію.

Після цього буде впроваджено методи і засоби забезпечення доступної роботи комплексної інфраструктури обчислювального кластера для вирішення наукових завдань проєкту.

Симуляція експериментів з розпаду динуклонних систем (динейтрон) розпочнеться з розробки моделі спектрометру для спостереження динейтрону з використанням кодів GEANT, TALYS, а також налагодження взаємодії розрахунків на базі Python скриптів. Валідація експерименту з виміру спектру електронів та гамма-квантів з опроміненої мішені буде проведена за допомогою моделей розпаду динейтрона. Прогнозування теплофізичних властивостей матеріалів і структур різної розмірності та морфології виконання тестових розрахунків коефіцієнта теплопровідності модельних наноструктур розрахунок температурних залежностей коефіцієнта теплопровідності матеріалів, їх аналіз та оптимізація морфології для енергоефективних застосувань буде здійснено методами молекулярно-динамічного моделювання процесів теплового транспорту в твердотільних структурах, теплового транспорту в напівпровідникових структурах різної розмірності: нанонитки, нанокомпозити, мультишарові структури. інтеграція VIRGO.UA у платформу Multi-Messenger Online Data Analysis супроводжуватиметься розробкою програмних кодів на основі машинного навчання для інноваційного аналізу даних гамма- та нейтринних спостережень, для побудови фізичних моделей прискорення космічних променів та генерування багатоканальних сигналів від астрофізичних об'єктів, розробкою інноваційного (на основі машинного навчання) аналізу даних гамма- та нейтринних спостережень, для побудови фізичних моделей прискорення космічних променів та генерування багатоканальних сигналів від астрофізичних об'єктів., картографуванням регіонів для поєднання аналізу даних для вивчення ізольованих точкових і протяжних джерел у широких кутових областях гамма- та нейтринних спостережень астрофізичних об'єктів, побудовою фізичних моделей прискорення космічних променів та генерування багатоканальних сигналів на основі картографування регіонів і поєднання аналізу даних для вивчення ізольованих точкових і протяжних джерел у широких кутових областях гамма- та нейтринних спостережень астрофізичних об'єктів.

Квантово-хімічні дослідження (в наближенні DFT та MCSCF) електронної будови та механізмів реакцій на їх основі буде здійснено за допомогою пакетів квантово-хімічних та програм для молекулярної механіки та динаміки, молекулярного докінгу та молекулярної динаміки.

Розроблення технології побудови великих нейронних лінгвістичних моделей буде здійснено шляхом побудови нових великих мовних моделей шляхом обробки текстових корпусів.

- Інформація про наявну матеріально-технічну базу, обладнання та устаткування, необхідні для виконання Проєкту (до 1 сторінки)

Обчислювальний кластер інформаційно-обчислювального центру (ІОЦ) Київського національного університету імені Тараса Шевченка <http://cluster.univ.kiev.ua/>, який буде розширено, а обладнання (яке ще не застаріло) – інтегровано у новий кластер, що буде створено за проєктом. Створений кластер гармонійно впишеться в існуючу кадрову та технічну інфраструктуру Університету, що забезпечить його подальший сталий розвиток.

Обчислювальний кластер ІОЦ КНУ має 10 обчислювальних вузлів Supermicro SYS- 6017R-TLF, E5-2620, DDR3-1600 32GB, 120 фізичних ядер (240 віртуальних ядер з технологією HyperThreading), виміряна обчислювальна потужність близько 1 ТФлопс, оперативної пам'яті 1.3 ГБайти на ядро. Кластер ІОЦ КНУ є першим обчислювальним ресурсом України для вирішення широкого кола задач (з 1999 р.) і незважаючи на застарілість і несправність частини обладнання цей кластер є діючим і постійно використовується для наукових досліджень та освітнього процесу науковцями, викладачами, аспірантами, студентами університету, інших наукових установ України, а також закордонними користувачами. Кластер університету є частиною Грід-інфраструктури для наукових та освітніх установ України (<http://grid.org.ua/monitor/>) і підключений до європейських Грід-інфраструктур. Наразі на кластері зареєстровано близько 60 локальних користувачів і ще більше 400 користувачів використовують кластер ІОЦ КНУ через Грід-інфраструктуру (<https://voms.grid.org.ua/voms/?vo=testbed.univ.kiev.ua>). На кластері встановлено і використовується більше 100 різних пакетів прикладних програм для моделювання у різних галузях і ці пакети періодично оновлюються.

- Очікувані результати виконання Проєкту (до 1 сторінки):

а) Опис наукової або науково-технічної продукції, яка буде створена в результаті виконання Проєкту (із зазначенням її очікуваних якісних та кількісних (технічних) характеристик).

Головним організаційно-науковим результатом проєкту буде створення в КНУ потужного обчислювального кластеру, доступного науковій спільноті України (KP1-KP2).

Наукові результати досліджень на створеному кластері будуть такими:

KP 3: Інформація про конформацію та електронну будову молекул – будівельних блоків для потреб медичної хімії, їх реакційну здатність, механізми відповідних хімічних реакцій та стабільність їх продуктів, взаємодію молекул із біологічними мішенями.

KP 4: Результати симуляції розпаду двонуклонних систем в пакеті GEANT (планується видання 1 або 2 публікацій квартилю Q1-Q2), а також знання про фізичні механізми теплового транспорту в твердотільних наноструктурах різної розмірності і морфології для енергоефективних застосувань та систем теплового контролю (планується не менше 2 статей в фахових виданнях, включаючи журнали квартилю Q1-Q2), спостереження астрофізичних джерел у режимі, наближеному до реального часу (заплановано публікацію не менше 2 статей у фахових виданнях, включаючи журнали квартилю Q1-Q2).

KP5: Буде отримана технологія розроблення, розширення, адаптації та тонкого налаштування великих нейронних моделей на прикладі створення великих лінгвістичних моделей і прикладних систем на основі цих моделей (планується заявка на патент).

б) Обґрунтування переваг очікуваної наукової або науково-технічної продукції порівняно з аналогами

На сьогодні аналогів такої дослідницької інфраструктури, як потужний комп'ютерний кластер, в Україні немає. Створений в процесі виконання проєкту кластер буде мати характеристики, що наближатимуть його до кращих обчислювальних центрів Європи.

Перевагою розрахунків на моделях, що плануються використати за проєктом (моделювання у ядерній фізиці, фізиці твердого тіла, хемоінформатиці) буде їх виконання на високому рівні теорії, що не властиво для більшості подібних робіт. Перевагою технології великих нейронних моделей буде максимальна оптимізація всіх обчислювальних процесів з урахуванням можливостей наявного устаткування побудованого обчислювального кластера. Розроблена технологія буде орієнтована на побудову та адаптацію великих лінгвістичних моделей до української мови, українських мовних реалій, смислового поля тощо. Тому технологія буде потужним інструментом для розроблення низки прикладних систем інтелектуальної обробки саме українськомовних текстів., що існують, на підставі порівняльного аналізу.

в) Обґрунтування практичної цінності запланованих результатів реалізації Проєкту для економіки та суспільства.

Створення в КНУ власного потужного комп'ютерного кластера позитивно вплине на якість публікацій в міжнародних наукових виданнях, освітній рівень студентів та докторантів, а також конкурентоздатність науковців КНУ при подачі заявок на фінансування в рамках європейських та міжнародних наукових проєктів, дозволить суттєво підвищити рівень наукових досліджень в галузі комп'ютерних хімії, біології, фармакології та викладання студентам та докторантам відповідних навчальних дисциплін, які вимагають залучення розрахункових ресурсів та надання студентам (докторантам) доступу до кластера (KP1-KP2).

Дослідження хімічних сполук, що планується за проєктом, створить передумови для оптимізації умов реакцій цих сполук, реакційної здатності продуктів, їх хімічної модифікації. Подальше вивчення взаємодії молекул із біологічними мішенями методами молекулярної динаміки та молекулярного докінгу дозволить виділити молекули-лідери, перспективні з точки зору медичної та агрохімії для їх подальшої розробки як кандидатів на лікарські засоби.

- Опис шляхів та способів подальшого використання результатів виконання Проєкту в суспільній практиці (до 1 сторінки).

Високопродуктивний обчислювальний кластер буде ресурсом колективного користування. Таке користування забезпечуватиметься через віддалений доступ з мережі Інтернет з використанням стандартних протоколів (ssh, https, тощо), за допомогою веб-браузера для реєстрації (<http://cluster.univ.kiev.ua/>), клієнта ssh для роботи (plus.hpc.knu.ua). Також кластер буде доступний для використання засобами Грід (<http://testbed.univ.kiev.ua/>, <https://voms.grid.org.ua/>, <http://grid.org.ua/test>). Робоча група кластера отримуватиме і виконуватиме заявки на реєстрацію користувачів кластера на основі цієї процедури. Для забезпечення справедливого розподілу ресурсів між проєктами та користувачами буде розроблена спеціалізована система планування.

Кластер буде залучений до широкого кола дослідницьких проєктів, які виконуватимуться як в Україні, так і міжнародними командами дослідників. Успішне виконання проєкту дозволить розвивати та утримувати обчислювальні ресурси власними силами після закінчення робіт за проєктом, за рахунок участі у спільних проєктах.

Розширення тематики наукових задач, які можуть бути профінансовані у майбутньому Українськими чи міжнародними грантодавчими організаціями, включає, зокрема, наступні пункти.

А) З початку повноцінного функціонування кластеру будуть відновлені розрахунки Грід-задач у промисловому режимі для участі у спільних проєктах з Європейським центром ядерних досліджень та інших.

Б) Починаючи з 2 етапу, буде налагоджено застосування кластера у навчальному процесі університету для викладання таких навчальних курсів: Паралельні обчислення, Фізика біомолекул, Сучасні комп'ютерні технології у медицині, Грід-системи та технології, Математичні методи обробки діагностичних даних, Комп'ютерна розробка лікарських препаратів, Структурна біологія та моделювання, Структурна біоінформатика, Хемоінформатика, Обчислювальна біологія. Частина студентів, як передбачається, будуть здобувати освіту в цих галузях на умовах платного навчання, що також буде непрямым джерелом підтримки кластера.

В) Виконання високопродуктивних обчислень різними підрозділами університету та інших організацій в рамках спільних проєктів: Молекулярна динаміка, Квантова хімія, Моделювання взаємодії іонізуючого випромінювання методом Монте-Карло, Нелінійна динаміка і біологічні нейронні мережі, Штучний інтелект, Астрономія, тощо. Виконання таких проєктів буде здійснюватися в рамках програм МОН, НФДУ та за рахунок приватних організацій – клієнтів кластера (в Україні), та на умовах фінансування в рамках Європейських (Горизонт Європа) та американських (Національний Інститут здоров'я) проєктів, участь в конкурсах на згадане фінансування буде здійснюватися на 1, 2, та 3 етапах проєкту.

Г) Університет вже передбачив використання потужностей кластеру, і відповідно його фінансування за кошти ЄС, в межах Європейських цифрових інноваційних хабів на 2024-2027 рр (31 січня 2024 року Університет став членом консорціуму на створення Heath Tech Hub Ukraine мета якого полягає у становленні та розвитку системи сталої та доступної підтримки цифровізації та інновацій у сфері охорони здоров'я України загалом та Київської області зокрема).

5. НАУКОВІ АБО НАУКОВО-ТЕХНІЧНІ РЕЗУЛЬТАТИ, ЯКІ ОЧІКУЄТЬСЯ ОТРИМАТИ У 2024 РОЦІ (до 2 сторінок)

ЕВП № 1 (2024 рік)

Завдання 1.1 Розробка моделі спектрометра для спостереження динейтрону з використанням кодів GEANT, TALYS, а також налагодження взаємодії розрахунків на базі Python скриптів. Підготовка програмних кодів для молекулярно-динамічного моделювання процесів теплового транспорту в твердотільних структурах, виконання тестових розрахунків коефіцієнта теплопровідності модельних наноструктур.

Отримання даних спостережень віддалено гамма- та нейтринних телескопів, для побудови фізичних моделей прискорення космічних променів та генерування багатоканальних сигналів від астрофізичних джерел (таких як Cygnus X та MGRO J1908+06). (КР3)

Завдання 2.1 Встановлення та тестування пакетів квантово-хімічних та програм для молекулярної механіки та динаміки, забезпечення віддаленого доступу до них, підготовка вхідних завдань. (КР4)

Встановлення комерційних пакетів програм та пакетів із безкоштовною ліцензією на комп'ютерний кластер. Підготовка та виконання серії розрахункових тестів. Оцінка продуктивності вузлів кластера та створення оптимального режиму доступу та виділення ресурсів для окремих класів задач із врахуванням умов ліцензій. Проведення тестових розрахунків структур реальних молекул в різних наближеннях. Підготовка вхідних даних для виконання завдань 4.2-4.4.

Завдання 3.1 Розроблення методів тонкого налаштування великих мовних моделей для виконання низки прикладних задач. (КР5)

Тонке налаштування великих мовних моделей є відносно простішою задачею, аніж побудова великих мовних моделей, тому виконання цих робіт доцільно планувати на перший етап проекту. За базові мовні моделі будуть взяті наявні (BERT, GPT і т.п.). Роботи складаються із побудови навчальних наборів даних, відповідних прикладним задачам, навчання моделі та проведення експериментів із замірювання якості роботи моделі. Результатом цього етапу, окрім самих оптимізованих методів налаштування, буде низка прикладних програм обробки текстів, таких як системи класифікації та розпізнавання текстів пропаганди, недостовірних новин (fake-news) та інші.

Завдання 4.1 Запускові та тестові роботи по комплектації комп'ютерного кластеру. Проектування змін до архітектури обчислювального кластера ІОЦ з урахуванням встановлення нових вузлів зберігання даних, вузлів віртуалізації та обчислювальних вузлів. Розробка, тестування та впровадження політик, системного програмного забезпечення та його конфігурації для забезпечення колективного використання нового типу вузлів з графічними прискорювачами. Встановлення та тестування коректної роботи у конкурентному режимі використовуваного у проєкті прикладного програмного забезпечення. Проектування та реалізація схеми розміщення обладнання для забезпечення рівномірного навантаження на системи енергоживлення та забезпечення температурного режиму роботи обладнання. (КР1)

Завдання 5.1 Проведення установчої зустрічі з визначення порядку дії, підписання угод про нерозголошення. Проведення зустрічі з Консультативною радою. (КР2)

Індикатори виконання у 2024 році:

1. Результати тестування у вигляді звіту. Модель спектрометру для спостереження динейтрону з використанням кодів GEANT, TALYS. Дані спостережень віддалено гамма- та нейтринних телескопів. Звітна документація (Звіт про проміжні результати реалізації проєкту, висновок Консультативної ради проєкту).
2. Встановлені та протестовані пакети квантово-хімічних та програм для молекулярної механіки та динаміки. Встановлені комерційні пакети програм та пакетів із безкоштовною ліцензією на комп'ютерний кластер. Результати тестових розрахунків структур реальних молекул в різних наближеннях.
3. Методи тонкого налаштування великих мовних моделей для виконання низки прикладних задач. Прикладні програми, результати тестування.
4. Проєкт змін до архітектури обчислювального кластера IOI з урахуванням встановлення нових вузлів зберігання даних, вузлів віртуалізації та обчислювальних вузлів. Спроектвані та реалізовані схеми розміщення обладнання для забезпечення рівномірного навантаження на системи енергоживлення та забезпечення температурного режиму роботи обладнання.
5. Проведення установчої зустрічі з визначення порядку дії, підписання угод про нерозголошення.

5.1. Заплановані наукові або науково-технічні результати, які очікуються отримати в рамках проміжного етапу виконання Проєкту:

№ з/п	Індикатори виконання	Заплановано (відповідно до КП), кількість
1.	Публікація результатів:	
1.1.	Статті у журналах, що індексуються наукометричними базами даних:	-
	- Scopus та/або Web of Science Core Collection, всього, од.	-
	з них із квантилем Q1 і Q2 на момент опублікування, од.	-
	з них із квантилем Q3 і Q4 на момент опублікування, од.	-
1.2.	Статті у наукових журналах (без квантилю), збірниках наукових праць, матеріалах конференцій тощо, що індексуються наукометричними базами даних Scopus або Web of Science Core Collection (крім тих, що увійшли до п.1.1), од.	-
1.3.	Статті у фахових виданнях України категорії «Б», од.	-
1.4.	Статті у періодичних виданнях інших країн, що мають ISSN, од.	-
1.5.	Публікації у матеріалах конференцій, тезах доповідей та виданнях, що не включені до переліку наукових фахових видань України та не індексуються наукометричними базами даних Scopus або Web of Science Core Collection, од.	-
1.6.	Монографії та розділи монографій, опубліковані (або підготовлені і подані до друку) у закордонних видавництвах іноземними мовами,	-

	од.		
1.7.	Монографії та розділи монографій, опубліковані (або підготовлені і подані до друку) в українських видавництвах, од.	-	
1.8.	Монографії та розділи монографій, опубліковані (або підготовлені і подані до друку) з обмеженим доступом (для проєктів оборонного та/або подвійного призначення), од.	-	
1.10	Словники, довідники, енциклопедії, видані українськими та/або закордонними видавництвами, од.	-	
1.12	Інші публікації, які не описані у пп. 1.1-1.10, од.	-	
2.	Презентація та дисемінація результатів:	3	
2.1.	Міжнародні науково-комунікативні заходи, конференції, од.	1	
2.2.	Всеукраїнські та регіональні науково-технічні/промислові виставкові заходи, од.	-	
2.3.	Представлення розробки/бізнес-плану/результатів проєкту на:		
	інноваційних фестивалях, од.	-	
	конкурсах стартапів, од.	-	
	акселераційних програмах, од.	-	
	хакатонах, од.	-	
2.4.	Науково-популярні публікації з метою поширення інформації про результати проєкту для загальної (широкої) аудиторії, од.	-	
2.5.	Представлення інформації про результати проєкту на науково-популяризаційних заходах (Дні науки, Наукові пікніки тощо), од.	2	
2.6.	Інші заходи, які не описані у пп. 2.1-2.5, од.	-	
3.	Охоронні документи на об'єкти права інтелектуальної власності (ОПІВ)		
3.1.	Отримано патентів України на винахід, од.	-	
3.2.	Отримано свідоцтв України на промисловий зразок, од.	-	
3.3.	Отримано патентів України на корисну модель, од.	-	
3.4.	Отримано охоронний документ на ОПІВ інших країн, од.	-	
3.5.	Інші ОПІВ, які не описані у пп. 3.1-3.4, од.	-	
3.6.	Подано заявок на отримання охоронного документу на ОПІВ України та/або інших країн, од.	-	
4.	Впровадження та використання наукових або науково-практичних результатів:		
4.1.	Підписання (укладання) договорів (угод) на впровадження (використання) результатів проєкту (окрім індивідуальних), серед них:	-	
4.1.1	Господарських договорів/контрактів, од./тис.грн.	-	
4.1.2	Ліцензійних договорів/договорів на ноу-хау, од./тис.грн.	-	
4.1.3	Грантових угод, од./тис.грн.	-	
4.1.4	Інші договори (угоди), які не описані у пп. 4.1.1-4.1.3, од./тис.грн.	-	

4.4	Документально підтверджено використання результатів у практиці органів державної/місцевої влади, суспільних практиках тощо, од.	-	
4.5	Проведено маркетингові дослідження, перемовини з потенційними замовниками із підписанням протоколу (меморандуму, угоди) про наміри комерційного впровадження результатів, од.	-	
4.6	Подано заявок на державні, міжнародні наукові гранти (окрім індивідуальних), од.	-	
5.	Створено чи істотно удосконалено/покращено існуючі:	3	
5.1.	Пристрої (макет, експериментальний/дослідний зразок), од.	1	
5.2.	Матеріали, процеси, технології, технологічні регламенти, цифрові продукти та електронні сервіси, од.	2	
5.3.	ТУ, ДСТУ, будівельні норми, зареєстровані проекти законодавчих актів, од.	-	
5.4.	Наукові (науково-технічні) послуги, од.	-	
5.5	Іншу продукцію, яка не описана у пп. 5.1-5.4, од.	-	

5.2. Передбачувана наукова цінність і актуальність отриманих результатів (науково-технічної продукції), їх порівняння з українськими та/або кращими закордонними аналогами (до 30 рядків).

Оскільки перший етап проєкту націлений в основному на створенні дослідницької інфраструктури, він передбачає науково-організаційні заходи. Отримання наукових результатів і науково-технічної продукції планується на наступні етапи.

5.3. Опис запланованих робіт зі створення дослідницької інфраструктури на етап (до 20 рядків).

Для забезпечення роботи нового обчислювального ресурсу на даному етапі організаційно-наукових робіт планується побудувати інфраструктуру енергозабезпечення та віртуалізації, яка буде основою для подальшої побудови високопродуктивного обчислювального ресурсу. Буде придбано і інтегровано у кластер: джерело безперебійного живлення; блок розподілу живлення; сервер збереження даних; жорсткі диски; 4 мережевих адаптера; обчислювальний сервер; мережеві адаптери з підтримкою Linux; графічний акселератор.

5.4. Опис практичної цінності результатів для потреб оборони, безпеки, економіки та/або суспільства України (у разі наявності) (до 30 рядків).

Після закінчення періоду виконання проєкту комп'ютерний кластер буде використовуватися в межах України та їх міжнародними партнерами на умовах спільного доступу за фінансування українськими та міжнародними грантодавчими організаціями. Потенційне співробітництво з комерційними установами відбуватиметься в рамках договорів про виконання науково-технічного дослідження. Наявність розрахункового кластера дозволить науковцям не тільки КНУ, а і всієї України використовувати комп'ютерний розрахунок як елемент дослідження в складних, комплексних міждисциплінарних проєктах. Це значно посилить роль України в міжнародних консорціумах та дозволить брати в них участь у статусі повноцінного партнера. З іншого боку, в комбінації з якісними розрахунками, експериментальні дослідження теж вийдуть на новий рівень розуміння, що значно підвищить як їх ефективність, так і практичну значимість для економіки країни. Крім цього, наявність розрахункового кластера дозволить готувати фахівців міжнародного рівня в Україні без потреби їх виїзду за кордон. У галузі використання штучного інтелекту існує можливість залучення майбутніх споживачів як для створення великих

лінгвістичних моделей, так і споживачів окремих розроблених прикладних лінгвістичних систем. Буде можливим також виконання замовлень клієнтів щодо побудови спеціальних прикладних систем «під ключ». Потенційними споживачами таких систем є органи державної влади (МОН), силові відомства (СБУ, ГУР, МО, МВС), високотехнологічні ІТ-компанії, освітні та наукові заклади України.

Технічне завдання до Проєкту із виконання наукових досліджень і розробок не містить відомостей, заборонених до відкритого опублікування.

Науковий керівник Проєкту
Директор Навчально-наукового інституту високих технологій, професор



Igor KOMAROV



ПОГОДЖЕНО:

Перший заступник виконавчого директора з питань
грантової підтримки Грантонадавача

(підпис)

(Власне ім'я та ПРІЗВИЩЕ)

Андрій Кудряков

Начальник управління грантового забезпечення
Грантонадавача

(підпис)

(Власне ім'я та ПРІЗВИЩЕ)

Тетяна Березова

Начальник відділу

грантової підтримки у сфері соціально-гуманітарних наук

управління грантового забезпечення Грантонадавача

(підпис)

(Власне ім'я та ПРІЗВИЩЕ)

Аліона Демченко

Головний спеціаліст відділу

грантової підтримки у сфері соціально-гуманітарних наук

управління грантового забезпечення Грантонадавача

(підпис)

(Власне ім'я та ПРІЗВИЩЕ)

Лєв Сахаренко