

Пучков В.О.<https://orcid.org/0009-0004-2152-4353>

Національний університет «Одеська юридична академія»

ПОСТАНОВКА ПРОБЛЕМИ ДОСЛІДЖЕННЯ ПРОГНОСТИЧНИХ AR-СИСТЕМ НА ОСНОВІ МУЛЬТИМОДАЛЬНОГО АНАЛІЗУ

У статті розглянуто проблему розробки прогностичних систем доповненої реальності на основі мультимодального аналізу поведінкових даних користувача. Актуальність дослідження зумовлена зростанням ролі доповненої реальності у забезпеченні ефективної людино-машинної взаємодії та необхідністю переходу від реактивних до проактивних інтерфейсів. Використання мультимодальних інтерфейсів, що поєднують обробку жестів, мовних команд, рухів очей та тактильних сигналів, дозволяє підвищити точність, швидкість і природність взаємодії користувача з системою, а також зменшити когнітивне навантаження за рахунок використання природних каналів сприйняття та керування. У роботі запропоновано концепцію проактивної взаємодії, відповідно до якої система здатна прогнозувати дії користувача та адаптувати інтерфейс на основі аналізу мультипотоків даних, отриманих із різних сенсорних джерел. Розглянуто застосування методів машинного навчання та когнітивного моделювання для передбачення намірів користувача, що створює передумови для формування адаптивних та інтелектуальних AR-систем. Проведено огляд підходів, які демонструють ефективність використання мультимодальних інтерфейсів у прикладних сферах, зокрема в освіті, промисловості та робототехніці. Виконано аналіз обмежень сучасних AR-систем, серед яких виділено недостатній рівень інтеграції даних різних модальностей, обчислювальні обмеження при обробці великих обсягів даних у режимі реального часу, а також вузьку спеціалізацію існуючих алгоритмів. Визначено, що подолання цих обмежень потребує комплексного підходу до поєднання мультимодальних сигналів, когнітивних моделей та методів інтелектуальної обробки даних. Результати дослідження підтверджують доцільність подальшої розробки адаптивних прогностичних AR-систем, орієнтованих на персоналізацію взаємодії з користувачем та проактивну підтримку його дій у реальному часі.

Ключові слова: доповнена реальність, мультимодальна взаємодія, прогностичні системи, прогнозування, людино-машинна взаємодія, когнітивні моделі, машинне навчання, інтерактивні інтерфейси.

Постановка проблеми. Розвиток інформаційних технологій і мобільних платформ сприяв поширенню систем доповненої реальності (Augmented Reality, AR), що інтегрують віртуальні об'єкти у фізичне середовище в реальному часі та розширюють можливості взаємодії користувача. AR розглядається як напрям розвитку людино-машинної взаємодії (Human-Machine Interaction, HMI), оскільки забезпечує інтуїтивні інтерфейси, що поєднують реальний та віртуальний простір.

Еволюція AR-систем передбачає перехід до мультимодальних інтерфейсів, що використовують жести, голос, рухи очей і тактильні сигнали, підвищуючи природність взаємодії та знижуючи когнітивне навантаження. Активно розробляються архітектури AR-систем із сенсорними пристроями, алгоритмами обробки сигналів та інтеграцією даних із різних джерел, що дозволяє створювати інтерактивні інформаційні серед-

овища для освіти, промисловості, медицини та робототехніки.

Перспективним є використання штучного інтелекту для аналізу поведінки користувача та створення адаптивних систем, здатних прогнозувати дії та наміри на основі мультимодальних даних. Незважаючи на прогрес, більшість AR-систем залишаються реактивними і не враховують динамічні наміри користувача, що обмежує ефективність взаємодії.

Незважаючи на значний прогрес у дослідженнях доповненої реальності та мультимодальних інтерфейсів, більшість наявних AR-систем реалізують переважно реактивні моделі взаємодії, реагуючи на команди користувача пост-фактум. Питання аналізу мультимодальних даних для прогнозування намірів та побудови проактивних інтерфейсів, здатних адаптуватися до поведінки користувача, залишаються недостатньо дослі-



дженими. Існує потреба в розробці узагальнених архітектурних моделей та методів обробки мультимодальних даних, які б дозволили поєднувати інформацію з різних каналів для прогнозування дій користувача.

Таким чином, виникає суперечність між зростаючими вимогами до ефективності, інтуїтивності та адаптивності людино-машинної взаємодії в AR-середовищах, з одного боку, та обмеженістю реактивних AR-систем, які не здатні адаптуватися до динамічних намірів користувача на основі мультимодального аналізу, з іншого. Це зумовлює актуальність наукового завдання з дослідження та розробки підходів до створення прогностичних систем доповненої реальності, що базуються на мультимодальному аналізі поведінки користувача. Розв'язання цієї проблеми дозволить підвищити ефективність НМІ у середовищах доповненої реальності та мінімізувати когнітивне навантаження на користувача під час виконання завдань.

Аналіз останніх досліджень і публікацій. Ключовою особливістю AR-систем є поєднання фізичного та віртуального простору, що дозволяє користувачеві сприймати цифрову інформацію безпосередньо в контексті навколишнього середовища [12]. У межах цього дослідження під системою доповненої реальності розумітимемо апаратно-програмний комплекс, який забезпечує суміщення в реальному часі віртуальних об'єктів з фізичним оточенням користувача з метою розширення інформаційного контексту та підтримки виконання прикладних завдань.

Однією з визначальних характеристик сучасних AR-систем є орієнтація на природні способи взаємодії [3-5]. Традиційні графічні інтерфейси на основі клавіатури та маніпулятора типу «миша» виявляються неефективними в тривимірних інтерактивних середовищах, що стимулювало розвиток мультимодальних підходів до організації НМІ [10]. Мультимодальна взаємодія визначається як використання двох або більше каналів (модальностей) введення інформації для передачі команд та даних від користувача до системи. До основних модальностей, що застосовуються в AR, належать: жести, що розпізнавані за допомогою комп'ютерного зору або сенсорів, голосові команди, відстеження руху очей, тактильні сигнали, а також дані з різноманітних сенсорів [3].

Мультимодальний аналіз ґрунтується на інтеграції даних, отриманих з цих різноманітних джерел, з метою інтерпретації дій, контексту та намірів користувача. Цей процес передбачає розв'язання низки складних завдань:

1. синхронізація потоків даних, тобто узгодження в часі інформації, що надходить з різною частотою від різних сенсорів [2];

2. розробка методів комбінування різноманітних сигналів для формування цілісної інформаційної моделі поточного стану взаємодії [4];

3. обробка неповних, суперечливих або зашумлених даних, що є неминучим у реальних умовах;

4. врахування семантики середовища та поточного завдання для коректного розуміння значення мультимодальних сигналів [12].

Подальший розвиток AR-технологій нерозривно пов'язаний із використанням методів штучного інтелекту. Сучасні алгоритми машинного навчання, зокрема глибинні нейронні мережі, відкривають можливості для ефективної обробки великих обсягів мультимодальних даних та виявлення прихованих закономірностей у поведінці користувачів [6-9]. Це створює підґрунтя для переходу від реактивних систем до інтелектуальних адаптивних інтерфейсів. У таких системах вирішальну роль відіграють моделі класифікації та прогнозування, які дозволяють не лише ідентифікувати поточні дії, але й передбачати подальші наміри користувача [13], реалізуючи концепцію прогностичної взаємодії.

Прогностична взаємодія передбачає створення інформаційних систем, здатних на основі аналізу поведінки користувача та контексту середовища проактивно формувати рекомендації, адаптувати інтерфейс або виконувати підготовчі дії ще до моменту явного формулювання запиту [7]. Реалізація такого підходу вимагає подолання зазначених вище викликів мультимодального аналізу та інтеграції його результатів у моделі прийняття рішень, що функціонують у масштабі реального часу [8-9].

Постановка завдання. Метою дослідження є проведення літературного огляду наявних наукових праць, присвячених мультимодальним AR-системам, включно з методами інтерактивної взаємодії користувача через голос, жести та інші сенсорні канали, а також підходами до прогнозування поведінки користувача. На основі аналізу досліджень здійснюється формулювання наукової проблеми дисертаційного дослідження з обґрунтуванням його актуальності, визначенням прогалів у сучасній науковій літературі та постановкою компонентів подальшого дослідження у сфері інтеграції мультимодальних інтерфейсів в AR-системи.

Виклад основного матеріалу. Мультимодальні техніки AR-систем. Мультимодальна

взаємодія є одним із ключових напрямів еволюції сучасних AR-систем, оскільки вона дозволяє користувачеві одночасно використовувати кілька природних каналів комунікації. Фундаментальне значення цього підходу підтверджується дослідженням Сіті С. М. Нізам та співавт. [1], які демонструють, що мультимодальні інтерфейси суттєво перевершують одномодальні за точністю виконання завдань та суб'єктивним задоволенням користувачів. Автори пропонують класифікацію основних модальностей та наголошують, що їх синергійне поєднання дозволяє створювати більш інтуїтивні та природні способи маніпулювання віртуальними об'єктами.

У роботі Джу Чан Кім та співавт. [2] досліджуються архітектурні аспекти побудови мультимодальних AR-систем, інтегрованих з екосистемою Інтернету речей (IoT). Автори виділяють типовий набір модальностей, що включає відстеження погляду, розпізнавання жестів та голосу, а також дані з IoT-сенсорів. Для ефективного функціонування таких гетерогенних систем вкрай необхідні стандартизовані методи синхронізації та інтеграції даних, які наразі перебувають на стадії активних досліджень.

Емпіричні підтвердження ефективності мультимодальності наведено в роботі Чжимін Ван та співавт. [3]. Результати їхніх експериментів свідчать, що комбінування модальностей не лише підвищує загальну продуктивність користувача, але й дозволяє скоротити час виконання типових завдань у віртуальному середовищі. Сяоян Чжоу та співавт. [4] дослідили специфіку взаємодії у складних багатопредметних AR-сценах та показали, що інтеграція жестів і мовленнєвих команд є ефективною в умовах високої когнітивної навантаженості, забезпечуючи точніше та швидше виконання завдань.

Важливу роль у становленні цього наукового напрямку відіграли класичні роботи, зокрема дослідження Сільвії Іраваті та співавторів [5]. Їхні експерименти підтвердили фундаментальну перевагу мультимодальних підходів: поєднання жестів та голосових команд суттєво покращує ефективність взаємодії порівняно з традиційними методами управління.

Можна констатувати, що мультимодальний підхід є загально визнаним стандартом для побудови сучасних людино-машинних інтерфейсів у доповненій реальності. Водночас, ці роботи зосереджені переважно на реактивній парадигмі взаємодії, тобто на інтерпретації вже виконаних користувачем дій. Питання використання багатих

мультимодальних потоків даних для прогнозування ще не виконаних дій та прихованих намірів користувача залишається відкритим і потребує окремого дослідження.

Прогностичні AR-помічники. Прогностична взаємодія в AR-системах спрямована на передбачення дій користувача до моменту їх явного виконання, що дозволяє підвищити ефективність, швидкість та інтуїтивність роботи з цифровим середовищем. У роботі Рагху Пара [6] досліджується інтеграція мультимодальних сигналів (голосу, жестів і відстеження погляду) для моделювання намірів користувача в AR-середовищах електронної комерції. Запропонована система здатна аналізувати поведінку покупця та на основі виявлених патернів надавати персоналізовані рекомендації, адаптуючи віртуальне оточення під індивідуальні потреби. Робота демонструє потенціал мультимодального аналізу для створення адаптивних інтерфейсів.

Більш узагальнений підхід запропоновано в дослідженні Юань Сю та співавт. [7], де представлено ієрархічну модель прогнозування майбутньої поведінки користувача. Модель базується на аналізі траєкторій погляду та семантичного контексту сцени, що дозволяє AR-системі передбачати дії користувача у віртуальних тренажерах та навчальних комплексах. Автори демонструють, що врахування контексту суттєво підвищує точність прогнозування порівняно з аналізом лише ізольованих поведінкових актів.

Перспективний напрям становить використання когнітивних моделей типу Belief-Desire-Intention (BDI) для створення прогностичних AR-помічників. BDI-модель, що запозичена з теорії агентів та філософії дії, описує раціонального агента через три компоненти: переконання, бажання та наміри. Внесок у розвиток цього підходу зробили Чень Лі та співавт. [8], представивши систему Satori – проактивного AR-асистента, здатного аналізувати наміри користувача та пропонувати релевантні дії або контент на основі моделювання його переконань, бажань і намірів у реальному часі. Автори демонструють, що BDI-підхід дозволяє системі не просто реагувати на команди, а проактивно передбачати потреби користувача.

Розвиток цієї ідеї представлено в роботі Салми Нурунніса [9], де BDI-модель застосовується для покращення співпраці між людиною та інтелектуальними агентами в AR-середовищах. Авторка підкреслює, що моделювання намірів та бажань користувача дозволяє агентам передбачати його

поведінку і діяти проактивно, суттєво зменшуючи затримки у взаємодії та підвищуючи загальну інтуїтивність системи. Увага приділяється механізмам адаптації моделі до змін у поведінці та цілях користувача.

Розглянуті дослідження закладають теоретичні та практичні основи для створення прогностичних AR-систем. Вони демонструють, що мультимодальні дані є інформативним джерелом для виявлення намірів користувача. Когнітивні моделі пропонують концептуально потужний апарат для формалізації та прогнозування поведінки.

Однак, аналіз цих робіт виявляє низку невирішених проблем, які стримують практичне впровадження прогностичних AR-систем, та визначають сутність наукової проблеми, зокрема відсутність узагальнених методів інтеграції мультимодальних потоків даних із високорівневими когнітивними моделями типу BDI.

Архітектурні підходи до організації AR-систем. Ефективність користувацького досвіду в системах доповненої реальності критично залежить від архітектурної організації, яка забезпечує узгоджену обробку сенсорних даних, синхронізацію віртуального та реального простору, а також реалізацію інтерактивних сценаріїв у реальному часі. Як зазначають Теофілос Пападопулос та співавт. [10], AR-системи мають забезпечувати безшовну інтеграцію різноманітних методів взаємодії, що висуває підвищені вимоги до модульності, масштабованості та продуктивності програмно-апаратних рішень. Автори підкреслюють, що ключовим викликом залишається узгодження різноманітних сенсорних потоків та створення адаптивних інтерфейсів, здатних підлаштовуватися під індивідуальні особливості користувача та контекст середовища.

Специфіку архітектурних рішень для окремих прикладних областей розкрито в роботі Рьо Судзукі та співавт. [11], які досліджують застосування AR у робототехніці. Автори пропонують класифікацію архітектурних підходів до інтеграції мультимодальних інтерфейсів у системи людино-роботизованої взаємодії. Дослідження демонструє, що адаптивні архітектури, здатні враховувати як дії оператора, так і стан роботизованих систем, підвищують ефективність управління складними технічними комплексами. Особлива увага приділяється питанням синхронізації даних між різноманітними сенсорами та забезпеченню детермінованих затримок, критичних для безпеки в людино-машинних системах.

Ще одним з напрямів архітектурних досліджень є контекстно-орієнтовані AR-системи.

Томі Тахара та співавт. [12] пропонують підхід Retargetable AR, який використовує тривимірні графі для моделювання семантичних зв'язків між об'єктами в реальному середовищі. Така архітектура дозволяє системі не лише відстежувати просторове розташування об'єктів, але й розуміти їх функціональне призначення та взаємозв'язки, що є критичним для коректної інтерпретації дій користувача.

У своєму дослідженні Євгеній Канарський та співавт. [13] розглядають застосування марківських моделей для моделювання поведінки оператора в AR-середовищі, призначеному для моніторингу небезпечних зон за допомогою безпілотних літальних та наземних апаратів. Запропонований підхід демонструє, як статистичні моделі можуть бути інтегровані в архітектуру AR-системи для прогнозування наступних дій оператора на основі аналізу послідовностей його попередніх дій.

Аналіз архітектурних підходів свідчить про поступову еволюцію AR-систем від простих візуалізаторів до інтелектуальних середовищ, здатних враховувати контекст та семантику взаємодії. Водночас, наявні рішення є вузькоспеціалізованими та використовують обмежений набір вхідних даних.

Мультимодальні AR-системи в промисловості. Ольга Павлова та співавт. [14] досліджують інформаційні технології візуалізації 3D-моделей об'єктів на основі доповненої реальності, демонструючи, що накладання віртуальних інструкцій та схем безпосередньо на реальне обладнання сприяє більш точному та швидкому виконанню технологічних операцій, а також суттєво зменшує кількість помилок при взаємодії з технікою.

Валерій Капустник та співавт. [15] аналізують вплив технологій доповненої реальності на ефективність праці та психологічний стан працівників. Автори підкреслюють, що застосування AR може підвищувати продуктивність виконання завдань та покращувати сприйняття інформації, однак це потребує ретельного врахування ергономічних аспектів та індивідуальних особливостей користувачів при проектуванні інтерфейсів.

Дослідження демонструють ефективність застосування AR-технологій у промисловості, підтверджуючи їх здатність покращувати візуалізацію, підвищувати інтерактивність та персоналізувати виробничий досвід. Водночас, в промислових AR-системах проактивна підтримка користувачів залишається нереалізованою.

Висновки. На основі аналізу публікацій виконана постановка наукової проблеми дисертацій-

ного дослідження, яка полягає в необхідності розробки теоретичних засад, методів та засобів побудови прогностичних AR-систем, здатних на основі комплексного аналізу мультимодальних даних про поведінку користувача передбачати його наміри та проактивно адаптувати віртуальне середовище в реальному часі. Розв'язання цієї проблеми потребує подолання зазначених вище суперечностей між наявним теоретичним та практичним доробком та відсутністю цілісних методологічних та архітектурних рішень.

Відповідно до сформульованої проблеми визначено основні компоненти дисертаційного дослідження.

Об'єктом дослідження є процес людино-машинної взаємодії в середовищах доповненої реальності на основі мультимодальних даних. Предметом дисертаційного дослідження є моделі, методи та засоби прогнозування намірів користувача в системах доповненої реальності на основі мультимодального аналізу поведінкових даних.

Метою дисертаційного дослідження є підвищення ефективності людино-машинної взаємодії в середовищах доповненої реальності шляхом розробки моделей та методів прогностичного аналізу мультимодальних даних для проактивної адаптації інтерфейсу.

Список літератури:

1. Nizam S. M., Abidin R. Z., Hashim N. C., Lam M. C., Arshad H., Majid N. A. A. A review of multimodal interaction technique in augmented reality environment. *International Journal of Advanced Science, Engineering and Information Technology*. 2018. Vol. 8, No. 4-2. P. 1460. DOI: 10.18517/ijaseit.8.4-2.6824.
2. Kim J. C., Laine T. H., Åhlund C. Multimodal interaction systems based on internet of things and augmented reality: a systematic literature review. *Applied Sciences*. 2021. Vol. 11, No. 4. Art. 1738. DOI: 10.3390/app11041738.
3. Wang Z., et al. Comparing single-modal and multimodal interaction in an augmented reality system. *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2020. DOI: 10.1109/ISMAR-Adjunct51615.2020.00052.
4. Zhou X., Williams A. S., Ortega F. R. Eliciting multimodal gesture + speech interactions in a multi-object augmented reality environment. *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. 2022. P. 1-10. DOI: 10.1145/3562939.3565637.
5. Irawati S., et al. An evaluation of an augmented reality multimodal interface using speech and paddle gestures. *Proceedings of the International Conference on Artificial Reality and Telexistence*. Springer Berlin Heidelberg, 2006. P. 272-283. DOI: 10.1007/11941354_28.
6. Para R. K. Intent prediction in AR shopping experiences using multimodal interactions of voice, gesture, and eye tracking: a machine learning perspective. *Journal of Artificial Intelligence General Science (JAIGS)*. 2024. Vol. 7, No. 1. P. 52-62. DOI: 10.60087.
7. Xu Y., et al. Seeing my future: predicting situated interaction behavior in virtual reality. *arXiv preprint*. 2025. DOI: 10.48550/arXiv.2510.10742.
8. Li C., et al. Satori: towards proactive AR assistant with belief-desire-intention user modelling. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 2025. P. 1-24. DOI: 10.48550/arXiv.2410.16668.
9. Noorunnisa S. An extended belief-desire-intention model for human-agent collaboration : PhD thesis. CQUniversity, 2025.
10. Papadopoulos T., et al. Interactions in augmented and mixed reality: an overview. *Applied Sciences*. 2021. Vol. 11, No. 18. Art. 8752. DOI: 10.3390/app11188752.
11. Suzuki R., et al. Augmented reality and robotics: a survey and taxonomy for AR-enhanced human-robot interaction and robotic interfaces. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022. P. 1-33. DOI: 10.1145/3491102.3517719.
12. Tahara T., et al. Retargetable AR: context-aware augmented reality in indoor scenes based on 3D scene graph. *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2020. P. 249-255. DOI: 10.48550/arXiv.2008.07817.
13. Kanarskyi Y., et al. Markov modelling of human-machine interaction in an augmented reality environment for UAV/UGV-based hazardous area monitoring systems. *Radioelectronic and Computer Systems*. 2025. No. 4. P. 35-54. DOI: 10.32620/reks.2025.4.03.
14. Pavlova O., Bashta A., Kovtoniuk M. Augmented reality based information technology for objects 3D models visualization. *Computer systems and information technologies*. 2023. № 1. P. 68-74. DOI: <https://doi.org/10.31891/csit-2023-1-9>.
15. Капустник В. А., Стукалкіна Д. С., Завгородній І. В. Аналіз впливу доповненої реальності як прикладу цифровізації на ефективність праці та стан працівників. *Актуальні проблеми профілактичної медицини : зб. наук. пр.* Харків : ХНМУ, 2020. С. 61-63. URL: <https://repo.knmu.edu.ua/handle/123456789/26395>

Puchkov V.O. PROBLEM STATEMENT OF RESEARCH ON PREDICTIVE AR SYSTEMS BASED ON MULTIMODAL ANALYSIS

The article addresses the development of predictive augmented reality (AR) systems based on multimodal analysis of user behavioral data. The relevance of this research is driven by the growing role of AR in enabling effective human–machine interaction and the need to shift from reactive to proactive interfaces. The use of multimodal interfaces that integrate gesture recognition, voice commands, eye movements, and haptic signals allows for increased accuracy, speed, and naturalness of user interaction with the system, while also reducing cognitive load by leveraging natural perception and control channels. This work proposes a concept of proactive interaction, wherein the system can anticipate user actions and adapt the interface based on the analysis of multistream data obtained from various sensory sources. The application of machine learning methods and cognitive modeling for predicting user intentions is explored, creating the foundation for adaptive and intelligent AR systems. A review of approaches demonstrating the effectiveness of multimodal interfaces in applied domains, such as education, industry, and robotics, is also presented. The limitations of current AR systems are analyzed, highlighting insufficient integration of data from different modalities, computational constraints when processing large volumes of real-time data, and the narrow specialization of existing algorithms. It is concluded that overcoming these limitations requires a comprehensive approach combining multimodal signals, cognitive models, and intelligent data processing methods. The results of the study confirm the feasibility of further developing adaptive predictive AR systems aimed at personalizing user interaction and proactively supporting user actions in real time.

Keywords: *augmented reality, multimodal interaction, predictive systems, prediction, human–machine interaction, cognitive models, machine learning, interactive interfaces.*

Дата першого надходження статті до видання: 25.03.2026

Дата прийняття статті до друку після рецензування: 21.04.2026

Дата публікації (оприлюднення) статті: 19.05.2026