МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ ПСИХОЛОГІЇ
КАФЕДРА ЕКСПЕРИМЕНТАЛЬНОЇ ТА ПРИКЛАДНОЇ ПСИХОЛОГІЇ

ДИПЛОМНА РОБОТА

«КОГНІТИВНІ УПЕРЕДЖЕННЯ МОЛОДІ ТА ДОРОСЛИХ ЩОДО ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ»

на здобуття освітньо-кваліфікаційного рівня «магістр»

з напряму 053 «Психологія»

Студентки 2 курсу ОС «Магістр»
ОП «Нейропсихологія»
Коробєйнікової Надії Олексіївни

Науковий керівник:
Кандидат психологічних наук,
доцент кафедри експериментальної
та прикладної психології
Ягіяєв Ілля Ігорович

Допустити до захисту в ЕК
кафедра експериментальної та прикладної психології
протокол № _____ від _____
завідувач кафедри:
кандидат психологічних наук, доцент
Малишева Каріне Олегівна

_____
(підпис)

КИЇВ – 2026

ЗМІСТ

АНОТАЦІЯ

Актуальність дослідження зумовлена стрімкою інтеграцією штучного інтелекту в усі сфери життєдіяльності, що докорінно трансформує архітектоніку людського пізнання та провокує виникнення специфічних когнітивних викривлень, зокрема упередженості автоматизації та алгоритмічної відрази. Метою магістерської роботи є дослідження психологічних, соціально-демографічних та поведінкових предикторів формування конспірологічних переконань і когнітивних упереджень щодо штучного інтелекту, а також розробка та наукове обґрунтування програми їх психологічної корекції.

Об'єктом дослідження є процеси соціальної когніції в умовах цифровізації, а предметом – психологічні детермінанти та механізми формування когнітивних упереджень щодо систем штучного інтелекту. Емпіричний збір даних здійснювався на вибірці обсягом 74 особи з використанням адаптованого психодіагностичного інструментарію, що включав шкали HEXACO, GAAIS та AICBS.

Наукова новизна та результати дослідження полягають в уточненні ролі хронологічного віку та стабільних рис особистості у формуванні ставлення до алгоритмічних систем. Емпірична перевірка на даній вибірці не виявила їхнього визначального впливу, що дає підстави висунути припущення про наявність феномену універсальної цифрової соціалізації, за якої конспірологічні страхи мають виражений монологічний характер: недовіра в одному специфічному домені генералізується на інші аспекти функціонування штучного інтелекту.

Практичне значення роботи полягає у розробці структурно-функціональної тренінгової програми розвитку цифрової грамотності та критичного мислення. Програма безпосередньо спрямована на подолання дезадаптивного феномену когнітивного розвантаження через імплементацію когнітивних примусових функцій, застосування

запитів на порівняння та психологічно обґрунтованого розроблення підказок. Впровадження розробленого комплексу забезпечує трансформацію пасивної довіри у раціональний науковий сумнів, сприяючи формуванню відкаліброваної довіри та збереженню когнітивної автономії особистості у сучасному соціотехнічному середовищі.

Ключові слова: когнітивні упередження, штучний інтелект, упередженість автоматизації, алгоритмічна відраза, конспірологічні переконання, цифрова грамотність, когнітивне розвантаження, соціальна когніція.

ВСТУП

Швидка інтеграція систем штучного інтелекту (ШІ) у ключові сфери життєдіяльності суспільства стимулює глибокі соціально-психологічні трансформації, фундаментально змінюючи механізми обробки інформації та прийняття рішень (Соколова & Мошик, 2023; Castelo, 2019). З точки зору когнітивної психології, ШІ являє собою нову онтологічну категорію, до якої людська психіка не має еволюційно сформованих механізмів адаптації, що змушує індивідів застосовувати до алгоритмів традиційні евристики та концепти «народної психології» (folk psychology) (Laakasuo та ін., 2021). Як зазначає В. Герхеш (Gherheş, 2018, С. 42), ШІ можна визначити як «здатність машин або програм імітувати процеси людського мислення», проте його стрімкий розвиток часто супроводжується ірраціональними страхами. Сьогодні алгоритмічні системи перебирають на себе функції, які традиційно вважалися виключно людськими, що викликає складний спектр поведінкових реакцій: від надмірної довіри (автоматизаційного упередження) до конспірологічного мислення (Kieslich та ін., 2021; Rastogi та ін., 2022).

Особливої актуальності набуває проблема вікових відмінностей у сприйнятті ШІ, оскільки процеси цифровізації по-різному впливають на когнітивний розвиток молоді та психологічну адаптацію дорослих. Сучасні дослідження доводять, що сприйняття автоматизованих систем супроводжується антропоцентричним упередженням, через яке люди знецінюють продукти штучного інтелекту, намагаючись захистити власну унікальність (Millet та ін., 2023). Водночас виникає феномен алгоритмічної відрази (algorithm aversion), зокрема в ситуаціях, що вимагають морального вибору, оскільки «люди відчувають відразу до машин, які приймають моральні рішення» (Bigman & Gray, 2018, С. 21). Недостатня прозорість алгоритмів (ефект «чорної скриньки») породжує у частини населення ірраціональні страхи та переконання щодо намірів ШІ (Ferrer та ін., 2021; Gherheş, 2018).

Стан розроблення цієї проблеми в сучасній психологічній літературі характеризується активним накопиченням міждисциплінарних емпіричних даних. У світовій науці вивченням механізмів прийняття рішень, когнітивних упереджень та соціальної когніції займалися такі відомі вчені, як Д. Канеман, А. Тверскі, Г. Саймон, П. Словік, С. Фіске, М. Газзаніга, Н. Крігескорте, Дж. Тененбаум (Kriegeskorte & Douglas, 2018; Lake та ін., 2017). Попри їхній значний внесок у розуміння природи евристик та людської ірраціональності, недостатньо дослідженим залишається питання специфіки структурування когнітивних упереджень щодо ШІ залежно від віку користувачів (молоді порівняно з дорослими), а також роль конкретних особистісних рис як предикторів алгоритмічної відрази та конспірологічних переконань, що й зумовило вибір нашої теми.

Мета дослідження: з'ясувати психологічну структуру та вікові особливості когнітивних упереджень щодо використання штучного інтелекту в представників молоді та дорослих, а також емпірично виявити ключові особистісні предиктори упередженого ставлення та конспірологічних переконань.

Завдання дослідження:

1. Здійснити теоретико-методологічний аналіз проблеми виникнення, функціонування та видів когнітивних упереджень людини під час взаємодії з системами штучного інтелекту;
2. З'ясувати вікові особливості сприйняття технологій ШІ, рівень довіри до них та інтенсивність прояву феномену алгоритмічної відрази у групах молоді та дорослих;
3. Виявити ключові особистісні предиктори за моделлю HEXACO, які достовірно детермінують схильність індивіда до формування конспірологічних переконань та негативних установок;

4. Порівняти структуру когнітивних упереджень, установок та конспірологічних переконань у представників молоді та дорослих, визначивши статистичну значущість відмінностей;
5. Розробити структурно-психологічну модель впливу індивідуально-типологічних та вікових чинників на формування упередженого ставлення до ШІ.

Об'єкт дослідження – соціально-когнітивні установки та ставлення особистості до новітніх технологій.

Предмет дослідження – когнітивні упередження молоді та дорослих щодо використання штучного інтелекту.

Гіпотеза дослідження – існують виражені вікові відмінності у структурі когнітивних упереджень щодо штучного інтелекту: для дорослих більш характерним є високий рівень алгоритмічної відрази та конспірологічних переконань, тоді як молодь демонструє вищий ступінь некритичної довіри (автоматизаційного упередження). При цьому передбачається, що ключовими незалежними предикторами загального негативного ставлення та конспірологічних переконань виступають специфічні особистісні риси за моделлю HEXACO (зокрема, низькі показники відкритості до досвіду, висока емоційність та низька чесність-скромність), які у сукупності з віковим фактором формують рівень когнітивної упередженості особистості.

Методи дослідження. Для досягнення поставленої мети було застосовано такі методи: теоретичні (науковий аналіз, синтез, порівняння та узагальнення літератури з когнітивної психології), емпіричні (психодіагностичне тестування із застосуванням таких методик: Шестифакторний опитувальник особистості (HEXACO-60), Шкала загального ставлення до штучного інтелекту (General Attitudes towards Artificial Intelligence Scale – GAAIS), Шкала конспірологічних переконань щодо штучного інтелекту (Artificial Intelligence Conspiracy Beliefs Scale – AICBS) та методи математичної статистики (описова

статистика, кореляційний аналіз за Спірменом, t-критерій Стьюдента для незалежних вибірок, множинний регресійний аналіз).

База дослідження. У дослідженні взяли участь 74 респонденти.

Наукова новизна одержаних результатів:

1. Проаналізовано в українській вибірці структуру когнітивно-особистісних детермінант, що зумовлюють схильність до віри в конспірологічні теорії щодо штучного інтелекту (AICBS) та розвиток алгоритмічної аверсії. Зокрема, встановлено взаємозв'язок між нейропсихологічними конструктами соціальної когніції та суб'єктивною оцінкою загрози з боку автономних агентів;
2. Удосконалено концептуалізацію вікових розбіжностей у процесах сприйняття та прийняття ШІ-технологій. Отримано додаткові свідчення, що відмінності в цифровій соціалізації «цифрових аборигенів» та дорослих користувачів корелюють із різною інтенсивністю прояву антропоцентричного упередження та ступенем делегування когнітивного навантаження інтелектуальним системам;
3. Подальшого розвитку набуло дослідження механізмів цифрової дискримінації та алгоритмічного упередження через призму нейроеволюційної теорії когнітивних упереджень. Обґрунтовано, що автоматизаційне упередження та надмірна довіра до алгоритмів є наслідком специфічної роботи евристичних механізмів мозку в умовах інформаційної невизначеності сучасного цифрового простору.

Практичне значення одержаних результатів. Результати дослідження можуть бути використані організаційними психологами та HR-фахівцями під час розробки програм адаптації персоналу до впровадження новітніх технологій автоматизації на робочих місцях з метою зниження алгоритмічної відрази (Mariani & Baggio, 2022; Nye та ін., 2014). Виявлені

вікові та особистісні закономірності становлять інтерес для педагогів та психологів у сфері освіти для побудови безпечного цифрового середовища та розвитку критичного мислення (Єрмошкіна, 2025; Di Salvo, 2025). Крім того, розроблена структурно-психологічна модель може слугувати підґрунтям для розробників інтерфейсів систем штучного інтелекту (UI/UX) для створення більш транспарентних технологій, що знижують рівень конспірологічних побоювань у користувачів (Rastogi та ін., 2022).

Структура та обсяг роботи. Магістерська робота складається зі вступу, трьох розділів, висновків до кожного розділу, загальних висновків, списку використаних джерел, що налічує 62 найменувань. Повний обсяг роботи становить 87 сторінок.

РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ КОГНІТИВНИХ УПЕРЕДЖЕНЬ У КОНТЕКСТІ ВЗАЄМОДІЇ ЗІ ШТУЧНИМ ІНТЕЛЕКТОМ

1.1. Психологічна сутність та класифікація когнітивних упереджень особистості в умовах цифровізації

У сучасній когнітивній психології феномен когнітивного упередження (cognitive bias) розглядається як систематичне відхилення від нормативної раціональності або об'єктивної логіки під час сприйняття, обробки інформації та прийняття рішень. Фундаментальні засади вивчення цієї проблеми були закладені в межах евристичної концепції, згідно з якою людський мозок в умовах невизначеності покладається на спрощені когнітивні стратегії (евристики), що дозволяють швидко приймати рішення, проте часто призводять до систематичних похибок (Kliegr та ін., 2021). З точки зору теорії подвійних процесів, ці викривлення виникають переважно внаслідок домінування автоматичної, швидкої та інтуїтивної «Системи 1» над аналітичною, повільною та свідомою «Системою 2», що особливо яскраво проявляється в умовах інформаційного перевантаження, притаманного сучасному цифровому середовищу (Rastogi та ін., 2022).

З еволюційної перспективи когнітивні упередження розглядаються не просто як дисфункції мислення, а як адаптивні механізми, що максимізували шанси на виживання в умовах середовища епохи плейстоцену, однак виявилися частково дезадаптивними в сучасному високотехнологічному світі. Індивід взаємодіє з новітніми цифровими алгоритмами та системами штучного інтелекту (ШІ) за допомогою нейрокогнітивного апарату, який еволюційно не пристосований до обробки складної імовірнісної або алгоритмічної інформації. Як наслідок, виникає безпрецедентний когнітивний дисонанс: людська психіка стикається з об'єктами, які діють цілеспрямовано і раціонально, але позбавлені свідомості, що змушує мозок застосовувати до них застарілі концепції «народної психології» (folk psychology) (Laakasuo та ін., 2021). Таким чином, процеси цифровізації не

лише актуалізують старі когнітивні викривлення, але й створюють підґрунтя для формування нових специфічних патернів ірраціональності.

В умовах тотальної цифровізації та інтеграції ШІ у повсякденне життя класифікація когнітивних упереджень потребує перегляду та розширення. Базуючись на аналізі сучасної психологічної та міждисциплінарної літератури (Echterhoff та ін., 2024; Kliegr та ін., 2021; Korteling & Toet, 2022; Rastogi та ін., 2022; Wang & Redelmeier, 2024), доцільно виокремити три ключові групи когнітивних упереджень особистості, що активізуються в цифровому середовищі: 1) упередження обробки інформації; 2) специфічні упередження взаємодії в соціотехнічній системі «людина – алгоритм»; 3) соціокогнітивні та антропоцентричні упередження.

Перша група – упередження обробки інформації – охоплює класичні когнітивні викривлення, дія яких багаторазово посилюється алгоритмами цифрового середовища. До цієї категорії належить упередження підтвердження (confirmation bias), що полягає у схильності індивіда шукати, інтерпретувати та запам'ятовувати інформацію таким чином, щоб вона підтверджувала його попередні переконання (Kliegr та ін., 2021). У цифровому просторі це упередження каталізується алгоритмами персоналізації та рекомендаційними системами, які створюють так звані «інформаційні бульбашки» (filter bubbles). До цієї ж групи входить упередження доступності (availability bias), коли ймовірність чи частота події оцінюється на основі легкості, з якою приклади спадають на думку (Kliegr та ін., 2021). В умовах інформаційного перенасичення цифрові платформи маніпулюють увагою користувача, роблячи певні стимули (наприклад, негативні новини) гіпердоступними, що спотворює об’єктивну оцінку реальності.

Друга група – специфічні упередження взаємодії в системі «людина – алгоритм» – є безпосереднім наслідком делегування когнітивних функцій комп'ютерним системам. Центральне місце тут посідає автоматизаційне упередження (automation bias), яке

проявляється у некритичній довірі до рішень, згенерованих машиною, навіть за наявності очевидних емпіричних доказів їхньої хибності (Rastogi та ін., 2022). Користувачі часто розглядають алгоритми як бездоганні, об'єктивні сутності, що призводить до зниження пильності та перекладання власної когнітивної відповідальності на машину. Тісно пов'язаним із цим є ефект алгоритмічної прив'язки (anchoring bias), за якого первинна рекомендація ШІ стає потужним когнітивним якорем; під впливом ліміту часу чи втоми користувач виявляється нездатним достатньою мірою скоригувати своє остаточне рішення відносно цієї алгоритмічної підказки (Rastogi та ін., 2022).

Протилежним феноменом у цій же групі є алгоритмічна відраза (algorithm aversion) – стійке небажання людини покладатися на алгоритми, особливо після того, як вона стає свідком їхньої, навіть незначної, помилки (Castelo, 2019). Дослідження засвідчують, що алгоритмічна відраза різко зростає у ситуаціях, які стосуються морального вибору, медицини чи творчості, оскільки суб'єкт вважає такі завдання глибоко суб'єктивними і такими, що вимагають виключно людської емпатії та інтуїції (Bigman & Gray, 2018).

Третя група – соціокогнітивні та антропоцентричні упередження – відображає глибинні психологічні захисти та проєкції людської психіки у відповідь на появу інтелектуальних машин. З одного боку, спостерігається схильність до антропоморфізації, коли людина підсвідомо приписує алгоритмам інтенції, емоції, переконання та особистісні риси (Laakasuo та ін., 2021). Таке перенесення власних психічних станів на неживі об'єкти може породжувати невиправдану довіру або, навпаки, ірраціональний страх перед «намірами» ШІ (Gherheş, 2018; Kieslich та ін., 2021). З іншого боку, як захисна реакція на онтологічну загрозу з боку машин, активізується антропоцентричне упередження. Це явище полягає в упередженому знеціненні продуктів, створених алгоритмами, задля збереження віри в людську винятковість та унікальність. Наприклад, в емпіричних дослідженнях доведено, що люди систематично занижують оцінку креативності та відчувають менше

благоговіння (awe) перед творами мистецтва, якщо дізнаються, що вони згенеровані штучним інтелектом, а не людиною (Millet та ін., 2023).

Слід також зазначити, що в умовах цифровізації традиційні моделі поведінки, які спиралися на концепцію обмеженої раціональності, зазнають глибокої трансформації. Сьогодні ШІ виступає не лише пасивним інструментом, а й активним середовищем, що формує нові цифрові стимули. Взаємодія людини з алгоритмами провокує «ефект снігової кулі», коли когнітивні упередження розробників та користувачів інтегруються в набори даних, алгоритми навчаються на цих викривленнях (алгоритмічне упередження), а згодом ці ж системи закріплюють і посилюють ірраціональні патерни у нових користувачів, створюючи замкнену петлю зворотного зв'язку (Єрмошкіна, 2025; Ntoutsi та ін., 2020). Відповідно, глибоке теоретичне осмислення та класифікація цих упереджень є необхідним етапом для розробки психологічних інтервенцій, здатних гармонізувати взаємодію у системі «людина – штучний інтелект».

Поглиблюючи класифікацію когнітивних упереджень в умовах цифровізації, варто виокремити додаткову низку специфічних епістемічних викривлень, які виникають безпосередньо під час обробки згенерованих машиною масивів даних. Одним із таких феноменів є інформаційне упередження (information bias), що проявляється як стійка тенденція суб'єкта шукати додаткову інформацію від систем штучного інтелекту, навіть якщо вона є нерелевантною для прийняття поточного рішення (Kliegr та ін., 2021). В умовах взаємодії з аналітичними та експертними системами користувачі часто хибно вважають, що збільшення кількості алгоритмічних змінних чи розширення обсягу вхідних даних автоматично підвищує точність і надійність результату, що на практиці призводить до когнітивного перевантаження та зниження якості кінцевого судження.

Не менш вагомим чинником, що спотворює сприйняття алгоритмічної інформації, є ефект рейтерації (reiteration effect), відомий також як ілюзія правди. Сутність цього

упередження полягає у тому, що багаторазове повторення певного висновку системою машинного навчання поступово підвищує суб'єктивну оцінку його достовірності (Kliegr та ін., 2021). У ситуаціях, коли штучний інтелект генерує велику кількість подібних або структурно надлишкових правил чи прогнозів, користувач починає сприймати їх як більш переконливі, навіть якщо первинно ставився до них критично. Цей ефект тісно переплітається з упередженням первинності (primacy effect), за якого перша інформація, надана алгоритмом, справляє непропорційно сильний вплив на подальшу інтерпретацію всіх наступних даних, змушуючи мозок адаптувати нові факти під первинно сформований штучним інтелектом конструкт (Kliegr та ін., 2021).

Окрім кількісних та структурних аспектів подачі інформації, критичну роль у взаємодії зі штучним інтелектом відіграє емоційна валентність згенерованих даних, що актуалізує упередження негативності (negativity bias). Еволюційно закріплена властивість людської психіки надавати більшої ваги негативним стимулам порівняно з позитивними призводить до того, що алгоритмічні прогнози, сформульовані з використанням термінів із негативним забарвленням, привертають більше уваги та здаються користувачам більш значущими (Kliegr та ін., 2021). Водночас під час оцінки імовірнісних прогнозів штучного інтелекту регулярно активізується нехтування базовою ймовірністю (base-rate neglect). Користувачі схильні надмірно фокусуватися на конкретних, репрезентативних ознаках, які видає класифікаційний алгоритм, повністю ігноруючи апріорну статистичну ймовірність події, що є прямим наслідком нездатності інтуїтивної системи мислення ефективно оперувати складною байєсівською статистикою (Kliegr та ін., 2021).

З позицій нейропсихології та теорії подвійних процесів, усі ці викривлення не слід розглядати виключно як ситуативні дисфункції мислення. Вони є результатом функціонування «жорстко закодованих» (hard-wired) нейронних механізмів, які формувалися протягом мільйонів років еволюції для швидкого реагування в умовах первісного середовища, проте виявляються неадаптивними для вирішення складних

когнітивних завдань сучасності (Korteling & Toet, 2022). Отже, інтелектуальна вразливість людини перед машинними алгоритмами значною мірою зумовлена тим, що система свідомого аналітичного контролю (Система 2) часто не здатна вчасно розпізнати та загальмувати автоматичні евристичні реакції під час взаємодії з високотехнологічним та інформаційно насиченим середовищем (Korteling & Toet, 2022; Rastogi та ін., 2022). Таке глибоке розуміння біологічних обмежень людського мозку формує концептуальне підґрунтя для подальшої розробки спеціалізованих нейрокогнітивних інтервенцій, здатних мінімізувати вплив автоматичних упереджень на процеси прийняття рішень за участю штучного інтелекту.

Важливим аспектом класифікації когнітивних упереджень в умовах цифровізації є група ймовірнісних та логіко-оціночних викривлень, що виникають під час інтерпретації результатів машинного навчання. Зокрема, під час аналізу згенерованих штучним інтелектом правил або висновків у користувачів часто активізується помилка кон'юнкції (conjunction fallacy) та супутнє нерозуміння логічних конструктів. Це проявляється у схильності індивіда вважати складні, багатокомпонентні алгоритмічні пояснення більш правдоподібними, ніж прості, що є прямим порушенням законів теорії ймовірностей, оскільки додавання кожної нової умови об'єктивно знижує статистичну ймовірність події (Kliegr та ін., 2021). Тісно пов'язаним із цим феноменом є ефект плутанини оберненості (confusion of the inverse), за якого користувачі систем штучного інтелекту хибно ототожнюють ймовірність причини та наслідку, плутаючи прогностичну впевненість алгоритмічної моделі з фактичною (апріорною) поширеністю явища у вибірці (Kliegr та ін., 2021).

Окрему увагу в контексті взаємодії з алгоритмами слід приділити упередженням надмірної та недостатньої впевненості (overconfidence and underconfidence), які динамічно змінюються залежно від складності інтелектуального завдання. Дослідники зазначають, що оператори схильні переоцінювати надійність рекомендацій штучного інтелекту, якщо вони

супроводжуються високими числовими показниками впевненості системи, ігноруючи при цьому недостатній обсяг чи низьку якість вихідних даних (нечутливість до розміру вибірки) (Berthet, 2022; Kliegr та ін., 2021). Водночас за наявності незрозумілих або прихованих змінних у моделі машинного навчання (особливо в системах типу «чорна скринька») стрімко зростає упередження уникнення невизначеності (ambiguity aversion). Це змушує індивіда ірраціонально відхиляти об'єктивно точні прогнози лише через психологічний дискомфорт та відсутність толерантності до невідомих алгоритмічних параметрів (Kliegr та ін., 2021).

На етапі верифікації та ретроспективної оцінки результатів, згенерованих штучним інтелектом, визначальну роль відіграє упередження селективної доступності (selective accessibility). Цей когнітивний механізм полягає у тому, що особа під час прийняття рішення несвідомо виокремлює з усього запропонованого алгоритмом масиву даних лише ті ознаки, які резонують з її первинною гіпотезою, повністю ігноруючи інший інформаційний простір та альтернативні предиктори (Rastogi та ін., 2022). Зрештою, після того як певний прогноз експертної системи підтверджується фактичними подіями, у користувача часто формується упередження ретроспективного погляду (hindsight bias), яке характеризується хибним відчуттям того, що результат був очевидним від самого початку (Berthet, 2022). Такий феномен призводить до того, що індивід ретроспективно завищує оцінку власних аналітичних здібностей і знецінює внесок машини, що в подальшому провокує необґрунтовану самовпевненість та відмову від алгоритмічної підтримки у ситуаціях високого ризику. Розгляд цих додаткових категорій викривлень дозволяє сформувати цілісну картину нейрокогнітивних вразливостей людини в цифровому середовищі та доводить необхідність розробки людиноцентричних інтерфейсів (Explainable AI), адаптованих до особливостей людського мислення.

1.2. Психологічні особливості сприйняття та ставлення до технологічних інновацій

Фундаментальну роль у формуванні ставлення до ШІ відіграють процеси емоційної та когнітивної оцінки загрози. Згідно з розширеною моделлю паралельних процесів (EPPM), страх перед інноваціями виникає тоді, коли суб'єкт сприймає високий рівень загрози від об'єкта, але водночас оцінює свій потенціал подолання (coping potential) цієї загрози як низький (Kieslich та ін., 2021). У цьому контексті ШІ виступає не як безпосередній об'єкт страху, а як тригерний чинник середовища, що запускає процес когнітивного оцінювання (Kieslich та ін., 2021). Рівень сприйняття такої загрози є глибоко контекстуальним і залежить від домену застосування технології. Дослідження показують, що використання ШІ у медичній діагностиці викликає значно менше тривоги порівняно з делегуванням йому рішень у сферах відбору персоналу або схвалення кредитів, де людина гостро відчуває втрату суб'єктності та контролю над власним життям (Kieslich та ін., 2021),.

Крім доменної специфіки, сприйняття ШІ опосередковується балансом між очікуваною корисністю технології (perceived usefulness) та психологічним комфортом під час взаємодії з нею (Castelo, 2019),. Надання алгоритмам когнітивних ознак людяності (наприклад, здатності до аналітичного мислення) підвищує оцінку їхньої корисності, однак наділення їх афективними ознаками (здатністю відчувати або розпізнавати емоції) часто провокує психологічний дискомфорт, пов'язаний з ефектом «моторошної долини» (uncanny valley) та порушенням меж людської унікальності (Castelo, 2019). Цим пояснюється феномен алгоритмічної відрази (algorithm aversion) при виконанні завдань, що вимагають морального вибору або інтуїції. Люди систематично відмовляються покладатися на алгоритми у питаннях життя і смерті чи етичних дилемах, аргументуючи це тим, що машинам бракує автентичного емоційного досвіду та здатності до емпатії (Bigman & Gray, 2018).

Вторгнення алгоритмів у сфери, що традиційно вважалися виключною прерогативою людини, становить серйозний виклик для онтологічної безпеки особистості, активізуючи

антропоцентричне упередження (anthropocentric bias). Коли ШІ починає генерувати продукти мистецтва або демонструвати креативність, люди відчувають екзистенційну загрозу своєму статусу як біологічного виду (Millet та ін., 2023). Як психологічний захист вмикається механізм вмотивованого міркування (motivated reasoning): суб'єкти починають систематично знецінювати якість і креативність творів, якщо дізнаються, що вони створені штучним інтелектом, і відчувають перед ними значно менше благоговіння (awe) порівняно з роботами авторів-людей (Millet та ін., 2023).

Не менш важливим аспектом сприйняття інновацій є процеси соціального порівняння (social comparison), які яскраво проявляються в умовах загрози автоматизації робочих місць. Парадоксально, але ставлення до заміни людини машиною змінюється залежно від перспективи спостерігача. Загалом люди вважають більш етичним, коли одного працівника замінює інша людина, а не робот, керуючись просоціальними мотивами; проте, коли йдеться про загрозу власному працевлаштуванню, цей патерн радикально змінюється (Granulo та ін., 2019),. Звільнення на користь іншої людини завдає потужного удару по самооцінці, тоді як заміна алгоритмом сприймається як менш принизлива для ідентичності, навіть попри те, що роботизація викликає більший страх щодо майбутніх економічних перспектив індивіда (Granulo та ін., 2019). Таким чином, ставлення до інновацій модулюється несвідомою потребою збереження позитивного Я-концепту.

Нарешті, на рівні буденної свідомості сприйняття ШІ часто викривлюється упередженням егоцентричної телеології (egocentric teleology bias). Сутність цього феномену полягає в ілюзорній впевненості, що будь-яка складна цілеспрямована технологія, створена людством, є апріорі морально нейтральною або такою, що служить прогресу (Laakasuo та ін., 2021). Ця наївна оптимістичність часто зіштовхується з проблемою «чорної скриньки», коли нездатність зрозуміти логіку алгоритму або породжує беззаперечну (автоматизаційну) довіру до машини, перекладаючи на неї тягар прийняття рішень, або, навпаки, активізує упередження уникнення невизначеності, змушуючи ірраціонально відхиляти навіть

об'єктивно точні машинні прогнози (Kliegr та ін., 2021; Laakasuo та ін., 2021). Отже, психологічні особливості сприйняття ШІ не є результатом суто раціонального розрахунку користі, а визначаються складним переплетінням еволюційних механізмів обробки інформації, захисних реакцій ідентичності та соціокогнітивних упереджень.

Розширюючи теоретико-методологічний концепт класифікації когнітивних упереджень, доцільно звернутися до просторової моделі спільного прийняття рішень, яка дозволяє систематизувати джерела виникнення ірраціональних патернів. Згідно з цією моделлю, взаємодія між людиною та алгоритмом розгортається у трьох функціональних вимірах: спостережуваний простір (набір вхідних ознак та інформації), простір прогнозів (результати роботи алгоритму) та сприйманий простір (когнітивна сфера людини-оператора) (Rastogi та ін., 2022). Когнітивні викривлення формуються безпосередньо на перетині цих просторів. Наприклад, упередження доступності та евристика репрезентативності виникають під час взаємодії сприйманого простору зі спостережуваним, тоді як автоматизаційне упередження, ефект алгоритмічної прив'язки та ефект слабких доказів (weak evidence effect) є прямим наслідком інтеграції сприйманого простору з машинними прогнозами (Kliegr та ін., 2021; Rastogi та ін., 2022). Така просторова диференціація доводить, що викривлення не є однорідними, а залежать від того, на якому саме етапі обробки даних фокусується увага суб'єкта.

Окремий клас викривлень формується під впливом лінгвістичних та семантичних особливостей комунікації з генеративними моделями та системами обробки природної мови. Сучасні дослідження виявляють у таких системах наявність вираженого ефекту фреймування (framing bias) та упередження групової атрибуції (group attribution bias), які здатні дзеркально відтворювати людські стереотипи (Jones & Steinhardt, 2022; Tversky & Kahneman, 1981). Людина-користувач, взаємодіючи з алгоритмом, виявляється когнітивно вразливою до того, як саме штучний інтелект структурує та формулює свої висновки (через позитивний чи негативний семантичний фрейм), що безпосередньо модулює фінальне

рішення оператора. Крім того, в процесі тривалої експлуатації технологій виникає так зване емерджентне упередження (emergent bias), яке формується не на етапі математичної розробки моделі, а в процесі безпосереднього користувацького досвіду через специфіку графічних інтерфейсів та поступову зміну поведінкових звичок людини у взаємодії з уже розгорнутим цифровим продуктом (Ferrer та ін., 2021).

У спеціалізованих високоризикових доменах, таких як медична діагностика чи правосуддя, делегування когнітивних функцій штучному інтелекту актуалізує низку специфічних операційних помилок мислення. Серед них особливо виокремлюється помилка єдиного діагнозу (fallacy of a single diagnosis) – психологічна тенденція передчасно припиняти пошук альтернативних рішень після того, як експертна система генерує першу найбільш вірогідну гіпотезу (Redelmeier & Shafir, 2023). Супутнім феноменом виступає упередження бездіяльності (omission bias), за якого фахівці схильні надавати перевагу пасивному прийняттю алгоритмічної рекомендації, уникаючи активного втручання або перевірки. Це пояснюється тим, що психологічно шкода від власної безпосередньої помилкової дії сприймається суб'єктом набагато важче, ніж гіпотетична шкода від бездіяльності у разі суворого дотримання машинної підказки (Aberegg та ін., 2005).

Нарешті, на макрокогнітивному та організаційному рівнях впровадження складних систем штучного інтелекту часто супроводжується активізацією системного виправдання (system justification) та помилки незворотних витрат (sunk-cost fallacy). Системне виправдання проявляється у підсвідомій схильності індивідів захищати та легітимізувати існуючий статус-кво, включно з новими інтегрованими технологічними інфраструктурами, сприймаючи алгоритмічні правила як апріорі справедливі, раціональні та об'єктивні (Korteling & Toet, 2022). Зі свого боку, помилка незворотних витрат змушує організації та окремих користувачів продовжувати покладатися на неефективні або відверто упереджені моделі штучного інтелекту лише через значні фінансові, часові чи когнітивні ресурси, які вже були інвестовані в їхнє розгортання та адаптацію (Korteling & Toet, 2022). Глибоке

усвідомлення цих додаткових психологічних бар'єрів дозволяє перейти від простої констатації алгоритмічних похибок до комплексного розуміння еволюційних та організаційних детермінант людської ірраціональності в умовах безперервної цифрової трансформації.

Поглиблюючи аналіз психологічних особливостей сприйняття штучного інтелекту, неможливо оминути вплив індивідуально-типологічних характеристик користувачів, зокрема базових рис особистості. У сучасній психологічній науці найчастіше використовується п'ятифакторна модель особистості («Велика п'ятірка»), яка включає екстраверсію, доброзичливість, сумлінність, нейротизм та відкритість до досвіду. Ці риси виступають вагомими предикторами того, як індивід взаємодіятиме з новітніми технологіями, наскільки він буде схильним до довіри до алгоритмів або ж, навпаки, до алгоритмічної відрази. Емпіричні дослідження засвідчують, що відкритість до досвіду є одним із ключових чинників формування позитивного ставлення до штучного інтелекту. Особи з високим рівнем відкритості виявляють більшу інтелектуальну цікавість та готовність до експериментів з інноваціями, частіше відчувають позитивні емоції від взаємодії з алгоритмами та загалом вище оцінюють продукти, згенеровані нейромережами, зокрема у сфері художньої творчості (Grassini & Koivisto, 2024).

Сумлінність та екстраверсія також демонструють значущі, хоча й специфічні, зв'язки з готовністю використовувати технології штучного інтелекту. Дослідники зазначають, що екстраверти частіше фіксують вищу частоту застосування алгоритмічних рішень та чат-ботів, оскільки соціальна природа їхньої особистості знаходить відгук в інтерактивній та динамічній комунікації з діалоговими системами (Kiński та ін., 2025). Зі свого боку, сумлінні особи, яким притаманні самодисципліна та прагнення до порядку, використовують штучний інтелект переважно як інструмент для підвищення продуктивності та структурованості робочих процесів, що безпосередньо призводить до покращення результативності виконання завдань в умовах алгоритмічного асистування (Kovbasiuk та ін., 2025).

Надзвичайно цікавим та неоднозначним є вплив доброзичливості (agreeableness) на взаємодію з алгоритмами. З одного боку, існують дані про негативну кореляцію між доброзичливістю та інтенсивністю використання штучного інтелекту, що пояснюється гострою потребою таких осіб в емпатійній, автентичній людській комунікації, якої позбавлені бездушні машинні системи (Kiński та ін., 2025). З іншого боку, в умовах вирішення складних організаційних завдань особи з низьким рівнем доброзичливості (ті, що характеризуються більшою незалежністю, асертивністю та скептицизмом) демонструють значно вищу якість виконання роботи саме за допомогою штучного інтелекту. Алгоритмічна об’єктивність та автономність дозволяють таким індивідам уникати труднощів традиційної міжособистісної взаємодії та неефективності людської командної роботи (Kovbasiuk та ін., 2025).

Щодо нейротизму (емоційної нестабільності), то він традиційно асоціюється зі скептицизмом та тривожністю щодо впровадження новітніх технологій, що може знижувати загальний рівень адаптивності особистості до цифрового середовища (Kiński та ін., 2025). Однак сучасні результати показують, що помірний рівень нейротизму в окремих випадках сприяє кращому виконанню когнітивних завдань у взаємодії з чат-ботами, оскільки підвищена обережність, ретельність та тривожність змушують таких користувачів уважніше ставитися до машинних підказок (Kovbasiuk та ін., 2025). Водночас висока емоційна стабільність позитивно корелює зі схильністю віддавати перевагу автентичній людській діяльності та вище оцінювати продукти, створені людиною, порівняно з алгоритмічними аналогами (Grassini & Koivisto, 2024).

Слід також зазначити, що відносини між рисами «Великої п'ятірки» та штучним інтелектом сьогодні набувають двостороннього характеру. Сучасні алгоритми обробки природної мови (NLP) демонструють здатність самостійно діагностувати особистісні риси користувача на основі глибинного аналізу його текстової взаємодії з чат-ботами, показуючи прийнятні психометричні показники конвергентної та факторної валідності (Ouyang та ін.,

2023). Це створює рекурсивний зв'язок: індивідуальні риси людини визначають її ставлення до штучного інтелекту, а сам штучний інтелект здатний зчитувати ці риси, що відкриває нові можливості для розробки персоналізованих адаптивних інтерфейсів, здатних підлаштовуватися під психологічний профіль оператора для мінімізації його когнітивних упереджень.

Осмислення психологічних особливостей сприйняття штучного інтелекту неможливе без залучення концептуального апарату когнітивної обчислювальної нейронауки (cognitive computational neuroscience), яка формує методологічний міст між нейробіологією, психологією та комп'ютерними науками (Kriegeskorte & Douglas, 2018). Традиційні парадигми дослідження, що спиралися виключно на поведінкові реакції або концепцію «чорної скриньки», сьогодні піддаються гострій науковій критиці через їхню редукціоністську природу. Сучасні дослідники наголошують на необхідності створення повноцінних моделей мозкових обчислень (brain-computational models, BCMs), які не просто імітують когнітивні функції на алгоритмічному рівні, а й відображають реальну динаміку біологічних нейронних мереж (Kriegeskorte & Douglas, 2018). У цьому контексті з'явився новий міждисциплінарний напрям – «синтетична нейрофізіологія», яка використовує аналіз репрезентативної подібності (representational similarity analysis) для порівняння просторової та часової активації у штучних глибинних нейромережах із реакціями приматів та людей, що дозволяє виявити фундаментальні відмінності в обробці зорової та семантичної інформації (Barrett та ін., 2019).

З погляду нейропсихології, сприйняття інновацій значною мірою детермінується архітектонікою пам'яті та механізмами забування, які принципово різняться у біологічних та штучних системах. Як зазначають Дж. Зао та ін. (2022), людське забування є пасивним, адаптивним нелінійним процесом: часто що більше індивід намагається витіснити інформацію, то сильніше вона закріплюється у пам'яті через емоційне забарвлення. Натомість машинне забування є актом активного видалення даних. Цей нейрокогнітивний

дисонанс між очікуваннями людини та реакціями машини формує підсвідомий психологічний бар'єр у користувачів, оскільки алгоритмічні системи не здатні симулювати суб'єктивні емоційні стани, які виникають під час людського когнітивного процесу (Zhao та ін., 2022). Водночас спроби розробників наділити алгоритми людськими рисами призвели до зародження так званої «машинної теорії розуму» (machine theory of mind). Наприклад, емпірично доведено, що штучні нейронні мережі здатні спонтанно формувати «преференцію форми» (shape preference) – схильність класифікувати об'єкти переважно за їхніми контурами, а не за кольором чи текстурою, що до цього вважалося виключно людською когнітивною особливістю (Zhao та ін., 2022). Усвідомлення того, що машина може копіювати глибинні патерни людського сприйняття, викликає у користувачів складний спектр реакцій: від захоплення до екзистенційної тривоги.

Критичний аналіз сприйняття цифрових порад та підказок також вимагає розуміння нейродинамічних процесів прийняття рішень. Біологічний мозок спирається на роботу атракторних мереж (attractor networks), які забезпечують перетворення безперервного потоку ймовірнісних стимулів у стабільні, дискретні стани свідомості через гомеостатичну пластичність та конкурентну динаміку за принципом «переможець отримує все» (winner-takes-all) (Khona & Fiete, 2022). Коли користувач стикається зі складною багатокритеріальною рекомендацією ШІ, його атракторні нейромережі змушені редукувати цю складність до однозначного вибору. Саме на цьому етапі нейронної редукції виникають специфічні помилки, зокрема «плутанина оберненості» (confusion of the inverse), коли людина-оператор нейробіологічно не здатна відрізнити прогностичну впевненість алгоритму від реальної апріорної ймовірності події (Kliegr та ін., 2021). Таким чином, сліпа довіра до математичної точності машини стикається з фізіологічними обмеженнями підтримки стійкої нейронної активності (persistent activity) в префронтальній корі людини (Khona & Fiete, 2022).

Ще одним об'єктом гострої критики в нейропедагогіці та когнітивній психології є вплив постійної взаємодії з ШІ на мережу пасивного режиму роботи мозку (default mode network, DMN). Ця нейронна мережа є критично важливою для саморефлексії, автобіографічної пам'яті, соціального пізнання та побудови моделі психічного стану іншої людини (Theory of Mind). М. Ді Сальво (2025) зауважує, що біологічний інтелект послуговується соціоемоційним «кермом» (rudder), яке скеровує процеси мислення шляхом афективної оцінки кожного когнітивного кроку. Надмірне покладання на ШІ та делегування йому аналітичних функцій позбавляє мозок необхідності активувати власні механізми виконавчого контролю (executive control) та орієнтування, що призводить до гіпоактивації DMN під час відпочинку. Внаслідок цього розмиваються межі суб'єктивності індивіда, а безперервне інформаційне перевантаження алгоритмічними стимулами порушує природні процеси нейропластичності (Di Salvo, 2025). Отже, сприйняття штучного інтелекту має розглядатися не лише як проблема психологічної довіри чи ергономіки інтерфейсів, але й як фундаментальний нейробіологічний виклик, що змінює архітектуру синаптичних зв'язків та еволюційні траєкторії розвитку людського мозку.

1.3. Вікові особливості сприйняття та взаємодії зі штучним інтелектом у молодіжному та дорослому віці

Впровадження технологій штучного інтелекту відбувається на тлі вираженої міжгенераційної дихотомії, де вік виступає одним із ключових предикторів сприйняття, довіри та патернів взаємодії з алгоритмічними системами. Перехід від аналогового до цифрового суспільства зумовлює формування принципово різних когнітивних екологій для молоді, яка соціалізується в умовах тотальної алгоритмізації, та дорослих, чиї нейронні мережі та виконавчі функції формувалися у доцифрову епоху (Di Salvo, 2025). Згідно з дослідженнями у галузі психології та комп'ютерних наук, базові демографічні змінні, зокрема віковий фактор, не лише модерують загальну готовність індивіда до прийняття інновацій, але й визначають специфіку виникнення когнітивних упереджень, рівня психологічного комфорту та оцінки ризиків під час делегування завдань штучному інтелекту (Castelo, 2019). Відповідно, глибоке розуміння нейропсихологічних та соціокогнітивних особливостей цих двох вікових груп є необхідною умовою для створення комплексних та безпечних моделей взаємодії людини й машини.

Для молодіжного віку характерним є феномен «цифрової аборигенності», який супроводжується більш високим рівнем базової довіри до технологій та меншим ступенем алгоритмічної відрази порівняно зі старшим поколінням. Нейропсихологічний аналіз показує, що у молоді префронтальна кора головного мозку, зокрема дорсолатеральна ділянка, що безпосередньо відповідає за когнітивний контроль, гальмування імпульсів та виконавчі функції, продовжує складний процес мієлінізації орієнтовно до двадцятип'ятирічного віку (Di Salvo, 2025). Оскільки цей період активної нейропластичності та формування нейронних ансамблів сьогодні протікає в середовищі, безпрецедентно насиченому алгоритмами штучного інтелекту та соціальними медіа, молодь виявляється більш схильною до автоматизаційного упередження та надмірного покладання на машинні підказки. Як зазначає М. Ді Сальво (2025), інтенсивне використання цифрових

комунікацій змушує молодих людей фокусуватися на конкретних, негайних стимулах, об'єктивно знижуючи їхню здатність до абстрактного морального міркування, емоційної рефлексії та довгострокового планування. Це робить молодь більш когнітивно вразливою до алгоритмічних маніпуляцій, попри те, що загалом молодші користувачі демонструють вищий рівень психологічного комфорту під час взаємодії з роботизованими агентами та адаптуються до їхніх функцій без гострого відчуття екзистенційної загрози (Castelo, 2019).

Натомість дорослі користувачі, чий когнітивний світогляд базується на тривалому аналоговому досвіді, значно частіше сприймають штучний інтелект як онтологічну та соціально-економічну загрозу, що активізує механізми психологічного захисту та конспірологічне мислення. У контексті професійної діяльності дорослі особи демонструють вищий рівень тривожності щодо можливої заміни алгоритмами на робочих місцях; причому усвідомлення такої технологічної заміни здатне критично знижувати їхню мотивацію до пошуку нової роботи та посилювати страх перед економічним майбутнім, знецінюючи їхній накопичений досвід (Granulo та ін., 2019). Окрім прагматичних економічних побоювань, старші вікові групи виявляють виражене антропоцентричне упередження під час оцінки результатів роботи штучного інтелекту. Емпіричні дослідження доводять, що чим старшим є респондент, тим нижче він оцінює привабливість візуальних творів та відчуває менше позитивних емоцій, якщо дізнається, що ці стимули згенеровані комп'ютерним алгоритмом (Grassini & Koivisto, 2024). Парадоксально, але дорослі та літні користувачі також частіше помиляються в об'єктивній атрибуції авторства, приписуючи згенеровані штучним інтелектом зображення людині, що свідчить про недостатню гнучкість їхнього перцептивного апарату у розпізнаванні новітніх цифрових паттернів (Grassini & Koivisto, 2024).

Відмінності між молоддю та дорослими також яскраво проявляються у сприйнятті моральної суб'єктності алгоритмів та делегуванні їм відповідальності за прийняття рішень. Дорослі особи переважно дотримуються жорстких онтологічних меж між біологічним та

штучним, відчуваючи глибоку алгоритмічну відразу до машин, які намагаються втручатися у сфери моралі, медицини чи правосуддя, оскільки обґрунтовано вважають такі системи позбавленими автентичного досвіду та здатності до справжньої емпатії (Bigman & Gray, 2018). З іншого боку, звикання молоді до постійної та безперебійної алгоритмічної підтримки несе приховані нейрокогнітивні ризики іншого характеру. Систематичне делегування аналітичних та соціально-когнітивних функцій штучному інтелекту штучно знижує навантаження на систему виконавчого контролю та може призводити до довгострокової гіпоактивації мережі пасивного режиму роботи мозку (Default Mode Network), яка є критично важливою для процесів саморефлексії, автобіографічної пам'яті та побудови моделі психічного стану інших людей (Di Salvo, 2025). Отже, якщо для дорослих ключовою психологічною проблемою є подолання бар'єру тривожності, недовіри та антропоцентричного опору, то для молоді головний виклик полягає у збереженні власної когнітивної автономії та критичного мислення в умовах безперервної симбіотичної взаємодії з інтелектуальними алгоритмами. Описані відмінності диктують необхідність вікової диференціації під час розробки стратегій впровадження систем штучного інтелекту та проєктування адаптивних цифрових інтерфейсів (Murthy та ін., 2025).

1.4. Моделі та психологічні механізми подолання когнітивних викривлень у системі «людина – штучний інтелект» та природа упереджень, страхів та конспірологічних переконань щодо ШІ

Стрімке впровадження технологій штучного інтелекту (ШІ) провокує широкий спектр психологічних реакцій, які варіюються від ентузіазму та некритичного прийняття до глибокої алгоритмічної відрази, страху та формування конспірологічних переконань (Stein та ін., 2024). З точки зору еволюційної психології, така реакція є закономірною, оскільки людська когнітивна система стикається з абсолютно новою онтологічною категорією (new ontological category) – штучними агентами, які демонструють ознаки цілеспрямованості, але позбавлені біологічної свідомості (Laakasuo та ін., 2021). Нездатність інтуїтивно осягнути алгоритмічні принципи роботи ШІ змушує індивідів застосовувати до нього механізми «народної психології», підсвідомо приписуючи машинам людські мотиви та наміри, що створює сприятливе підґрунтя для виникнення ірраціональних страхів (Laakasuo та ін., 2021).

Страх перед штучним інтелектом варто розглядати не як ізольовану фобію чи неприйняття інновацій, а як результат складного когнітивного процесу оцінки загрози (Kieslich та ін., 2021). Згідно з розширеною моделлю паралельних процесів (Extended Parallel Process Model, EPPM), страх виникає у тих ситуаціях, коли суб'єкт сприймає високий рівень загрози для власного благополуччя, але водночас оцінює свій потенціал подолання цієї загрози (coping potential) як низький (Kieslich та ін., 2021). У цьому контексті ШІ виступає потужним тригером, оскільки його функціонування (ефект «чорної скриньки») є непрозорим для пересічного користувача, що радикально знижує відчуття контролю над ситуацією та породжує тривогу (Laakasuo та ін., 2021). Дослідники зазначають, що емоційна реакція на ШІ є глибоко контекстуальною і залежить від конкретного домену його застосування, що безпосередньо впливає на рівень переживаної небезпеки (Kieslich та ін., 2021).

Фундаментальні страхи щодо ШІ можна умовно поділити на соціально-економічні та екзистенційно-моральні. В економічній площині ключовою є загроза технологічного безробіття та втрати соціального статусу; емпіричні дослідження демонструють, що перспектива втрати робочого місця через заміну роботизованим алгоритмом викликає у працівників значно сильніші побоювання щодо власного економічного майбутнього, ніж заміна іншою людиною, навіть якщо технологічна заміна завдає дещо меншого удару по самооцінці (Granulo та ін., 2019). З моральної точки зору, люди відчувають глибоку алгоритмічну відразу до передачі штучному інтелекту повноважень у ситуаціях, що вимагають етичного вибору, наприклад, у медицині, юриспруденції чи військовій справі (Bigman & Gray, 2018). Ця відраза опосередковується нейропсихологічним феноменом сприйняття розуму (mind perception): хоча машинам користувачі здатні частково приписувати здатність до аналітичного мислення (агентність), вони сприймаються як сутності, що повністю позбавлені здатності до афективного переживання та болю (досвіду), що робить їх, в очах людини, нездатними до емпатії та несення справжньої моральної відповідальності (Bigman & Gray, 2018).

Відчуття неконтрольованості та онтологічної невизначеності є живильним середовищем для формування конспірологічних переконань щодо штучного інтелекту, які виступають своєрідним захисним когнітивним механізмом, покликаним структурувати незрозумілу та лякаючу технологічну реальність (Stein та ін., 2024). Науковці виявили, що конспірологічна ментальність є одним із найбільш значущих предикторів загального негативного ставлення до ШІ та тісно корелює з глибокою недовірою до новітніх технологій (Kiński та ін., 2025). Ці переконання часто ґрунтуються на упередженні егоцентричної телеології (egocentric teleology bias) – схильності людини вважати, що складні цілеспрямовані технологічні системи приховують глобальні (часто зловмисні) мотиви, які насправді проєктуються на них самими ж користувачами через механізм антропоморфізації (Laakasuo та ін., 2021). Такі уявлення щедро підживлюються апокаліптичними наративами

масової культури, формуючи у свідомості стійкий стереотип про неминуче «повстання машин» та екзистенційну загрозу виживанню людства (Bigman & Gray, 2018).

Додатковим чинником виникнення конспірологічних страхів є сприйняття ШІ як безпосередньої загрози людській ідентичності та унікальності. Відповідно до теорії соціальної ідентичності, поява алгоритмів, здатних виконувати складні креативні чи аналітичні завдання на рівні експертів, розглядається як символічне вторгнення «чужинця» (outgroup), що розмиває еволюційні межі винятковості людського виду (Castelo, 2019). Люди схильні захищати свій антропоцентричний статус, підсвідомо знецінюючи досягнення ШІ або демонізуючи його наміри, що є прямим проявом групового упередження «ми проти них», перенесеного на міжвидовий рівень взаємодії (Laakasuo та ін., 2021). Отже, страхи та конспірологічні переконання щодо штучного інтелекту не є випадковими дисфункціями індивідуального мислення, а виступають закономірними продуктами еволюційно сформованої нейрокогнітивної архітектури, яка намагається адаптуватися до безпрецедентних викликів цифрової епохи, гіперактивуючи традиційні нейронні механізми захисту, уникнення невизначеності та соціального поділу (Korteling & Toet, 2022).

Проблема мінімізації та подолання когнітивних викривлень у процесі взаємодії людини з системами штучного інтелекту є центральною для забезпечення об'єктивності прийняття рішень. З позицій когнітивної психології, більшість упереджень виникають внаслідок домінування інтуїтивної, автоматичної системи обробки інформації (Системи 1) над аналітичною та свідомою (Системою 2), що особливо яскраво проявляється в умовах інформаційного перевантаження (Kliegr та ін., 2021). Відповідно, фундаментальний психологічний механізм подолання цих викривлень полягає у цілеспрямованому застосуванні стратегій, які змушують користувача активізувати Систему 2, сповільнюючи процес обробки даних та стимулюючи критичне переосмислення алгоритмічних рекомендацій (Kliegr та ін., 2021).

Однією з найбільш ефективних моделей викривлень є стратегія часового розподілу, безпосередньо спрямована на подолання ефекту алгоритмічної прив'язки та автоматизаційного упередження. Як доводять К. Растогі та колеги (2022), надання людині-оператору додаткового часу на обробку рекомендацій штучного інтелекту, особливо в ситуаціях, коли алгоритм демонструє низький рівень впевненості, достовірно знижує інтенсивність прояву когнітивної прив'язки. Цей підхід дозволяє капіталізувати людську експертизу, стимулюючи оператора відхилятися («де-якоритися») від хибних машинних прогнозів (Rastogi та ін., 2022). Крім того, поєднання такого часового розширення з наданням експліцитних пояснень щодо рівня впевненості моделі діє як когнітивна примусова функція, що зупиняє автоматичне слідування за підказкою та суттєво підвищує загальну точність спільної роботи в системі «людина – штучний інтелект» (Rastogi та ін., 2022).

Вагому роль у подоланні когнітивних викривлень відіграють інтерфейсні та репрезентативні моделі, які визначають формат подачі алгоритмічної інформації. Т. Кліегр та колеги (2021) наголошують, що заміна складних імовірнісних чи логічних висновків на так звані «частотні формати» здатна нівелювати помилки суджень, пов'язані з нехтуванням базовою ймовірністю та нечутливістю до розміру вибірки. З метою протидії ефекту рейтерації (ілюзії правдивості через багаторазове повторення алгоритмом одного й того ж висновку) розробникам пропонується кластеризувати надлишкові машинні правила, а також надавати користувачеві збалансовані інструкції, які вимагають пошуку доказів як «за», так і «проти» згенерованої системою гіпотези (Kliegr та ін., 2021). Доповнюючи цю тезу, Т. Міллер (2019) стверджує, що для ефективного подолання упереджень системи пояснюваного штучного інтелекту повинні спиратися на соціально-психологічні моделі людської комунікації, надаючи контрастивні пояснення, які відповідають когнітивним очікуванням суб'єкта,.

На організаційному та поведінковому рівнях пом'якшення автоматизаційного упередження та уникнення невизначеності вимагає впровадження жорстких механізмів підзвітності. О. В. Єрмошкіна (2025) зазначає, що ефективний контроль за алгоритмами потребує впровадження людського нагляду за системою за принципом «stop-loss» та обов'язкового культивування раціонального сумніву серед користувачів. Вимога до аналітиків експліцитно та структурно обґрунтовувати свою згоду або незгоду з рекомендацією штучного інтелекту штучно підвищує їхні когнітивні зусилля та радикально зменшує схильність до пасивного прийняття алгоритмічного вибору за замовчуванням (Kliegr та ін., 2021). Це повністю узгоджується з необхідністю встановлення чітких особистісних меж використання алгоритмів та усвідомленого визначення пріоритетності цілей під час вирішення завдань із високим ступенем ризику (Єрмошкіна, 2025).

Зрештою, стійкість до когнітивних викривлень забезпечується через динамічне калібрування довіри та безперервне оновлення ментальних моделей користувача. Згідно з концептуальною моделлю Дж. Лі та К. Сі (2004), довіра до автоматизації має бути «відповідною», тобто не призводити ані до надмірного покладання на машину, ані до дезадаптивної алгоритмічної відрази. У динамічному цифровому середовищі періодичне оновлення моделей штучного інтелекту може руйнувати ефективну співпрацю, якщо нова поведінка алгоритму починає непомітно суперечити вже сформованій ментальній моделі користувача (Bansal та ін., 2019). Тому системне забезпечення сумісності під час оновлення алгоритмів та прозоре інформування суб'єкта про зміну меж алгоритмічних помилок є критично важливими психологічними механізмами для збереження відкаліброваної довіри та підтримки високої результативності спільної діяльності (Bansal та ін., 2019).

Висновки до розділу 1

Узагальнення теоретико-методологічних засад дослідження дозволяє стверджувати, що стрімка інтеграція штучного інтелекту формує новітню соціотехнічну парадигму, в якій людська психіка стикається з об'єктами онтологічної невизначеності, що демонструють ознаки агентності без біологічної свідомості (Laakasuo та ін., 2021). Еволюційна невідповідність між принципами функціонування людського мозку та цифровими алгоритмами призводить до виникнення специфічних когнітивних упереджень, оскільки користувачі переносять застарілі соціальні евристики на автоматизовані системи. З позицій теорії подвійних процесів, ці викривлення виникають через домінування автоматичної обробки інформації в умовах перевантаження, що генерує такі феномени, як автоматизаційне упередження, ефект алгоритмічної прив'язки, ілюзію правдивості та упередження селективної доступності (Kliegr та ін., 2021; Rastogi та ін., 2022). При цьому системи штучного інтелекту не лише виступають тригерами людської ірраціональності, а й здатні акумулювати та масштабувати ці викривлення, створюючи замкнений цикл алгоритмічної упередженості.

Теоретичний аналіз підтверджує наявність глибокої міжгенераційної дихотомії у сприйнятті штучного інтелекту, яка детермінує розбіжності у формуванні когнітивних упереджень. Молодь, соціалізація якої відбувається в умовах цифрової насиченості на тлі тривалої мієлінізації префронтальної кори, демонструє вищий базовий рівень довіри до технологій та схильність до некритичного делегування когнітивних функцій машинам. Попри більший психологічний комфорт, це значно підвищує ризик гіпоактивації мережі пасивного режиму роботи мозку та знижує автономність аналітичного мислення (Di Salvo, 2025). Натомість для дорослих користувачів, чий когнітивний досвід спирається на аналогові патерни, більш характерними є алгоритмічна відраза та антропоцентричне упередження. Дорослі особи схильні сприймати ШІ як екзистенційну та соціально-економічну загрозу, що призводить до ірраціонального знецінення машинних рішень,

особливо у ситуаціях морального вибору, та стимулює формування конспірологічних переконань через відчуття втрати суб'єктності (Bigman & Gray, 2018; Granulo та ін., 2019).

Виникнення ірраціональних страхів та алгоритмічної недовіри опосередковується не лише віком, але й індивідуально-типологічними особливостями особистості. Риси характеру, зокрема відкритість до досвіду, рівень емоційної стабільності, сумлінність та екстраверсія, виступають ключовими предикторами здатності до адаптації в цифровому середовищі та визначають інтенсивність прояву алгоритмічної відрази (Grassini & Koivisto, 2024). Непрозорість глибинних нейромереж (ефект «чорної скриньки») у поєднанні зі схильністю психіки до антропоморфізації запускає процес когнітивного оцінювання загрози, за якого низький потенціал подолання невизначеності трансформується у захисні механізми, такі як упередження уникнення невизначеності та егоцентрична телеологія (Castelo, 2019; Kieslich та ін., 2021). Таким чином, негативне ставлення та конспірологічні побоювання є системною нейропсихологічною реакцією на порушення меж людської винятковості.

Зрештою, мінімізація когнітивних викривлень у багаторівневій системі спільного прийняття рішень вимагає впровадження науково обґрунтованих стратегій усунення упередженостей, спрямованих на активацію свідомого аналітичного контролю (Системи 2). Ефективна протидія автоматизаційному упередженню та ефекту прив'язки передбачає застосування когнітивних примусових функцій і стратегій часового розподілу, що стимулюють користувача формувати раціональний сумнів та об'єктивно оцінювати машинну впевненість (Rastogi та ін., 2022). Формування відкаліброваної довіри до автоматизованих агентів неможливе без врахування соціально-психологічних аспектів комунікації та створення прозорих пояснювальних систем, які б зважали на нейродинамічні та епістемічні обмеження людського мозку, забезпечуючи тим самим стійку і безпечну взаємодію людини з новітніми технологіями (Lee & See, 2004; Miller, 2019).

РОЗДІЛ 2. ЕМПІРИЧНЕ ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ КОГНІТИВНИХ УПЕРЕДЖЕНЬ ЩОДО ШТУЧНОГО ІНТЕЛЕКТУ В ОСІБ РІЗНОГО ВІКУ

2.1. Методологія, етапи та методи дослідження

Методологічним підґрунтям емпіричного дослідження виступили фундаментальні положення когнітивної психології та психології прийняття рішень, які розглядають взаємодію людини з новітніми технологіями через призму теорії подвійних процесів та концепції обмеженої раціональності (Kliegr та ін., 2021). З огляду на те, що ставлення до систем штучного інтелекту формується в умовах онтологічної невизначеності, коли людська психіка намагається адаптувати еволюційно застарілі евристики до оцінки штучних агентів, дослідницька стратегія була спрямована на виявлення вікових відмінностей та індивідуально-типологічних предикторів, які детермінують виникнення когнітивних викривлень (Laakasuo та ін., 2021).

Відповідно до мети дослідження було сформульовано емпіричну гіпотезу про те, що існують виражені вікові відмінності у структурі когнітивних упереджень щодо штучного інтелекту: молодь (віком 18-25 років) демонструє статистично вищий рівень некритичної довіри та позитивного ставлення до новітніх алгоритмічних систем, тоді як для дорослих (віком від 26 років) є характерним вищий рівень алгоритмічної відрази та віри в конспірологічні загрози (Di Salvo, 2025; Granulo та ін., 2019). Дослідження має на меті конкретизувати вплив індивідуально-психологічних диспозицій на ці процеси та передбачає, що ключовими предикторами загального негативного ставлення та конспірологічних переконань виступають специфічні особистісні риси за моделлю HEXACO, які безпосередньо опосередковують рівень когнітивної упередженості. Зокрема, передбачалося, що відкритість до досвіду є головним негативним предиктором упереджень (знижує їхню інтенсивність), емоційність позитивно корелює з тривожністю та загальним негативним ставленням до алгоритмів, а низькі показники чесності-скромності пов'язані зі специфічною недовірою до корпоративних і владних аспектів розробки штучного інтелекту

(Grassini & Koivisto, 2024). Окремо передбачалося, що інтенсивність використання цифрових інструментів здатна виступати модератором цих взаємозв'язків.

Для досягнення поставленої мети та забезпечення високого рівня надійності результатів емпіричне дослідження було структуровано через чотири послідовні взаємопов'язані етапи: 1) підготовчо-теоретичний етап, під час якого здійснювався аналіз наукової літератури, формулювалась гіпотеза та здійснювався підбір релевантного психодіагностичного інструментарію, 2) організаційно-емпіричний етап, що передбачав формування вибірки респондентів та безпосередній збір первинних емпіричних даних шляхом стандартизованого психологічного тестування, 3) аналітико-статистичний етап, присвячений математичній обробці отриманих масивів даних із застосуванням методів описової та індуктивної статистики, 4) інтерпретаційно-узагальнювальний етап, на якому здійснювалося психологічне тлумачення результатів та формулювання загальних висновків роботи.

Емпіричну базу дослідження склали 74 респонденти, які були розподілені на дві вікові когорти: група молоді (n = 18) та група дорослих (n = 56). Зважаючи на виражену диспропорцію у кількісному наповненні груп, об'єктивна оцінка результатів вимагає ретельної деталізації демографічних характеристик когорти дорослих користувачів. Віковий діапазон у цій групі становить від 26 до 76 років, що свідчить про її значну гетерогенність. У групі дорослих мода дорівнює 37, медіана – 42, а середнє значення – 44,59. Такий розподіл відображає широку варіативність технологічного досвіду респондентів, оскільки до однієї категорії увійшли як особи періоду ранньої дорослості, так і представники пізньої зрілості, чия цифрова соціалізація відбувалася за принципово різних соціотехнічних умов.

Важливо також зазначити обмеження дослідження, яке заключається у структурній асиметрії та обсягу сформованої вибірки, у якій група молоді (n = 18) виявилася кількісно

непропорційною відносно групи дорослих (n = 56). Водночас когорта дорослих характеризується надмірно широким і неоднорідним віковим діапазоном, об'єднуючи індивідів із потенційно різними когнітивними патернами взаємодії з цифровим середовищем. Зазначені демографічні диспропорції та мала репрезентативність молодіжної підвибірки суттєво обмежують можливості широкої статистичної генералізації висновків щодо вікової детермінації когнітивних упереджень та конспірологічного мислення. Відповідно, подолання цього концептуального обмеження вимагає проведення подальших наукових розвідок із залученням більш збалансованих, репрезентативних та жорстко стратифікованих за віком вибірок.

Процедура збору даних здійснювалася дистанційно за допомогою електронних платформ (Google Forms), що дозволило охопити широку аудиторію за одночасної мінімізації впливу фактора соціальної бажаності та тиску з боку дослідника. Участь в опитуванні була виключно добровільною та анонімною, кожен респондент надавав інформовану згоду перед початком тестування та мав право припинити участь на будь-якому етапі.

Комплекс методів дослідження включав структуровану анкету для збору соціально-демографічних даних (вік, стать) та показників частоти і глибини використання технологій штучного інтелекту (зокрема платформ ChatGPT, Midjourney тощо). Для психометричного оцінювання застосовувалася батарея з трьох стандартизованих методик: 1) коротка версія Шестифакторного опитувальника особистості HEXACO-60, яка дозволяє об'єктивно виміряти шість фундаментальних рис (чесність-скромність, емоційність, екстраверсію, доброзичливість, сумлінність та відкритість до досвіду), що розглядаються як ключові предиктори схильності до технологічної довіри (Ashton & Lee, 2009), 2) Шкала загального ставлення до штучного інтелекту (General Attitudes towards Artificial Intelligence Scale, GAAIS), яка вимірює загальний вектор сприйняття технологій через дві субшкали: позитивне ставлення (оцінка корисності та ентузіазм) і негативне ставлення (дискомфорт та

побоювання) (Schepman & Rodway, 2023), 3) Шкала конспірологічних переконань щодо штучного інтелекту (Artificial Intelligence Conspiracy Beliefs Scale, AICBS), яка операціоналізує специфічні ірраціональні упередження через субшкали глобального контролю, дезінформації, загрози людській праці, суперництва озброєнь та впливу на міжособистісні стосунки (Lin та ін., 2025). Для потреб регресійного аналізу показники субшкал конспірології також агрегувалися в єдиний інтегральний показник загальної конспірологічної упередженості.

Математико-статистична обробка даних здійснювалася за допомогою програмного комплексу IBM SPSS Statistics. На першому етапі використовувалася описова статистика для розрахунку середніх арифметичних значень (M) та стандартних відхилень (SD), а також критерій Шапіро-Вілка разом із показниками асиметрії та ексцесу для перевірки відповідності розподілу закону нормальності. Перевірка гіпотези про наявність вікових відмінностей здійснювалася шляхом порівняльного аналізу із застосуванням t-критерію Стьюдента для незалежних вибірок (для змінних із нормальним розподілом), а також його непараметричного аналога, U-критерію Манна-Уітні, для шкал, що продемонстрували статистично значуще відхилення від нормальності. Для з'ясування характеру та сили взаємозв'язків між рисами особистості за моделлю HEXACO та специфічними показниками упередженості застосовувався кореляційний аналіз за допомогою непараметричного коефіцієнта Спірмена. На фінальному етапі застосовувався метод множинного лінійного регресійного аналізу (методом Enter), мета якого полягала у визначенні прогностичної цінності особистісних факторів та віку щодо схильності суб'єкта до віри в конспірологічні загрози та негативного ставлення до систем штучного інтелекту.

2.2. Емпіричний аналіз проявів упередженості автоматизації та алгоритмічної відрази серед молоді та дорослих

Першим етапом реалізації емпіричного дослідження став розрахунок первинних статистичних показників (описової статистики) для перевірки якості зібраних даних, формування загального психологічного профілю вибірки (N = 74) та визначення можливості застосування параметричних методів математичного аналізу. Для обробки масивів «сирих» даних усі реверсивні твердження у методиках HEXACO-60 та Шкалі загального ставлення до штучного інтелекту (GAAIS) були попередньо перекодовані відповідно до стандартизованих ключів, після чого було обчислено середні арифметичні значення для кожної шкали та субшкали (Ashton & Lee, 2009; Schepman & Rodway, 2023). У контексті методології дослідження важливим є однозначне трактування процедури обробки даних за шкалою загального ставлення до штучного інтелекту (GAAIS). Згідно з оригінальним алгоритмом розробників методики, усі пункти, що формують субшкалу негативного ставлення, підлягають реверсивному перекодуванню: відповідь «цілком згоден» оцінюється в 1 бал, а «цілком не згоден» — у 5 балів. Відповідно, вищий підсумковий бал за кожною з субшкал (як позитивною, так і негативною) концептуально відображає більш сприятливе, позитивне ставлення до систем штучного інтелекту. Іншими словами, високий бал за субшкалою негативного ставлення (Negative GAAIS) математично свідчить про те, що респондент систематично відкидає тривожні твердження, демонструючи відсутність алгоритмічної відрази та низький рівень занепокоєння. Для уникнення термінологічної суперечливості та дотримання жорсткої логіки інтерпретації впродовж усього дослідження, результати за обома субшкалами аналізуються за принципом прямої пропорційності: зростання числового показника завжди еквівалентне зростанню рівня оптимізму, технологічного прийняття та толерантності до штучного інтелекту.

Аналіз показників центральної тенденції та варіативності за шестифакторним опитувальником особистості HEXACO-60 засвідчив, що найвищий середній бал у

досліджуваній вибірці респонденти продемонстрували за шкалою відкритості до досвіду (M = 3,56; SD = 0,58), що вказує на загальну готовність вибірки до сприйняття інновацій та інтелектуальну гнучкість. Інші особистісні риси розподілилися таким чином: сумлінність (M = 3,45; SD = 0,59), чесність-скромність (M = 3,36; SD = 0,63), екстраверсія (M = 3,24; SD = 0,59), доброзичливість (M = 3,00; SD = 0,47) та емоційність, яка отримала найнижчий середній показник (M = 2,98; SD = 0,70). Найменший розкид відповідей, що свідчить про високу однорідність вибірки за цією ознакою, спостерігається за шкалою доброзичливості (Таблиця 1).

Таблиця 1

Описова статистика основних шкал опитувальників

Змінна	**N**	**Мін.**	**Макс.**	**Середнє**	**Станд.відхил.**	**Асиметрія (Стат)**	**Асиметрія (Пом.)**	**Ексцес (Стат.)**	**Ексцес (Пом.)**
Чесність та Скромність	74	1,60	4,60	3,3608	0,63436	-0,807	0,279	1,089	0,552
Емоційність	74	1,60	4,50	2,9770	0,69648	0,130	0,279	-0,617	0,552
Екстраверсія	74	1,90	4,50	3,2392	0,59468	-0,220	0,279	-0,205	0,552
Доброзичливість	74	1,30	4,20	3,0027	0,47368	-0,311	0,279	1,412	0,552
Сумлінність	74	2,00	4,60	3,4541	0,59177	-0,371	0,279	-0,035	0,552
Відкр. досвіду	74	2,10	4,70	3,5557	0,58441	-0,209	0,279	-0,412	0,552
GAAIS (позит.)	74	1,00	4,83	3,3818	0,78985	-0,697	0,279	0,513	,552
GAAIS (негат.)	74	1,25	4,63	3,0270	0,84280	-,0364	0,279	0,692	0,552
Глобальний контроль	74	1,17	5,00	2,7218	1,08533	0,559	0,279	0,574	0,552
Дезінформація	74	1,00	5,00	3,1486	0,99481	-0,237	0,279	0,295	0,552
Загроза людській праці	74	1,00	4,92	2,9977	0,92364	0,188	0,279	0,489	0,552
Суперництво озброєнь та мілітаризація	74	1,00	5,00	3,0656	0,99655	0,149	0,279	0,364	0,552
Міжособистісні стосунки	74	1,00	5,00	3,1959	0,98231	-0,148	0,279	-0,432	0,552
Дійсне N	74								

Оцінка загального вектора ставлення до систем штучного інтелекту за методикою GAAIS виявила переважання позитивних очікувань над страхами. Так, середній бал за шкалою позитивного ставлення становить 3,38 (SD = 0,79), тоді як показник негативного ставлення та побоювань дорівнює 3,03 (SD = 0,84). Отриманий результат свідчить про помірний рівень технологічного дискомфорту у респондентів.

Щодо специфічних конспірологічних переконань (за методикою AICBS), загальний профіль вибірки демонструє середній або нижче середнього рівень упередженості. Найбільше занепокоєння у респондентів викликають загрози втручання штучного інтелекту в міжособистісні стосунки (M = 3,20; SD = 0,98) та ризики генерування дезінформації (M = 3,15; SD = 0,99). Дещо нижчі показники зафіксовано щодо побоювань мілітаризації та загрози миру (M = 3,07; SD = 1,00) і витіснення людської праці (M = 3,00; SD = 0,92). Найнижчий рівень віри спостерігається щодо теорій глобального контролю та поневолення людства алгоритмами (M = 2,72; SD = 1,09), що узгоджується з даними розробників шкали про те, що екзистенційні страхи є менш поширеними порівняно з прагматичними соціально-економічними побоюваннями (Lin та ін., 2025).

Ключовим кроком для вибору подальших методів індуктивної статистики стала перевірка відповідності емпіричного розподілу даних закону нормальності. З огляду на обсяг вибірки (N = 74), основним інструментом було обрано критерій Шапіро-Вілка (Таблиця 2). Результати засвідчили, що для більшості змінних ми не можемо відхилити нульову гіпотезу про нормальність розподілу ($p > 0,05$). Зокрема, нормальний розподіл підтверджено для шкал емоційності ($p = 0,272$), екстраверсії ($p = 0,323$), доброзичливості ($p = 0,138$), сумлінності ($p = 0,199$), відкритості до досвіду ($p = 0,287$), а також для конспірологічних субшкал дезінформації ($p = 0,179$), людської праці ($p = 0,343$), загрози миру ($p = 0,068$) та міжособистісних стосунків ($p = 0,325$). Водночас критерій Шапіро-Вілка виявив статистично значущі відхилення від нормальності ($p < 0,05$) для шкал чесності-

скромності (p = 0,002), глобального контролю (p = 0,002), а також позитивного (p = 0,039) та негативного (p = 0,035) ставлення до штучного інтелекту.

Таблиця 2

Результати перевірки на нормальність розподілу за критерієм Шапіро-Вілка

Змінна	**Шапіро-Вілка**		
	Статистика	**df**	**Знач. (Sig.)**
Чесність та Скромність	0,943	74	0,002
Емоційність	0,979	74	0,272
Екстраверсія	0,981	74	0,323
Доброзичливість	0,974	74	0,138
Сумлінність	0,977	74	0,199
Відкритість досвіду	0,980	74	0,287
GAAIS Позитивне	0,965	74	0,039
GAAIS Негативне	0,964	74	0,035
Глобальний контроль	0,943	74	0,002
Дезінформація	0,976	74	0,179
Загроза людській праці	0,981	74	0,343
Суперництво озброєнь та мілітаризація	0,969	74	0,068
Міжособистісні стосунки	0,981	74	0,325

Зважаючи на виявлені відхилення, було додатково проаналізовано показники асиметрії та ексцесу (Таблиця 1). Аналіз засвідчив, що для всіх змінних показники асиметрії знаходяться в безпечних межах від -0,807 (для чесності-скромності) до 0,559 (для глобального контролю), а показники ексцесу коливаються від -0,692 до 1,412 (для доброзичливості). Відповідно до психометричних стандартів, оскільки значення асиметрії та ексцесу не перевищують критичних порогових значень ±2, правомірно вважати розподіл усіх даних наближеним до нормального. Це методологічно обґрунтовує можливість застосування параметричних методів аналізу, зокрема t-критерію Стьюдента та регресійного аналізу, проте для шкал із виявленими відхиленнями за Шапіро-Вілком

результати обов'язково повинні верифікуватися їхніми непараметричними аналогами (U-критерієм Манна-Уітні та коефіцієнтом рангової кореляції Спірмена).

На фінальному етапі первинної обробки даних було підтверджено високу Внутрішню узгодженість та надійність використаного психодіагностичного інструментарію на українській вибірці. Для забезпечення наукової достовірності отриманих емпіричних даних було проведено аналіз надійності та внутрішньої узгодженості використаних психометричних шкал за допомогою розрахунку коефіцієнта альфа Кронбаха (α). Варто зазначити, що для Шестифакторного опитувальника особистості (HEXACO-60) додаткова перевірка внутрішньої узгодженості не проводилася, оскільки у дослідженні було використано стандартизовану україномовну версію методики, високі психометричні характеристики якої вже були підтверджені в межах попередніх адаптаційних процедур (Карпенко & Климпуш, 2025). Натомість фокус перевірки надійності був зосереджений на інструментах, що безпосередньо вимірюють когнітивні упередження та ставлення до новітніх технологій.

Перевірка Шкали загального ставлення до штучного інтелекту (GAAIS) здійснювалася окремо для її позитивного та негативного компонентів, що повністю відповідає двофакторній структурі, запропонованій розробниками (Schepman & Rodway, 2023). Позитивна субшкала, яка містить 12 пунктів, виявила високу внутрішню узгодженість із показником 0,883. Зі свого боку, негативна субшкала, що складається з 8 пунктів і вимірює рівень занепокоєння та побоювань користувачів щодо штучного інтелекту, продемонструвала надійність на рівні 0,844. Отримані результати доводять стійкість вимірюваних конструктів загального ставлення у межах досліджуваної вибірки (Таблиця 3-4).

Таблиця 3

Результати розрахунку альфа Кронбаха для позитивної субшкали GAAIS

Пункт	**Середнє шкали при вилученні пункту**	**Дисперсія шкали при вилученні пункту**	**Скоригована кореляція пункту із загальним результатом**	**Альфа Кронбаха при вилученні пункту**
GAAIS_1	37,89	76,235	0,513	0,878
GAAIS_2	37,12	75,999	0,658	0,870
GAAIS_4	37,72	78,644	0,423	0,883
GAAIS_5	36,86	79,023	0,429	0,883
GAAIS_7	37,20	74,739	0,532	0,878
GAAIS_11	37,09	75,046	0,693	0,868
GAAIS_12	36,95	77,641	0,571	0,875
GAAIS_13	37,15	75,115	0,650	0,870
GAAIS_14	36,57	76,852	0,653	0,871
GAAIS_16	37,20	76,246	0,615	0,872
GAAIS_17	37,54	76,087	0,610	0,872
GAAIS_18	37,09	73,840	0,717	0,866

Таблиця 4

Результати розрахунку альфа Кронбаха для негативної субшкали GAAIS

Пункт	**Середнє шкали при вилученні пункту**	**Дисперсія шкали при вилученні пункту**	**Скоригована кореляція пункту із загальним результатом**	**Альфа Кронбаха при вилученні пункту**
GAAIS_3	21,35	39,354	0,424	0,842
GAAIS_6	21,92	42,350	0,141	0,870
GAAIS_8	21,24	32,406	0,808	0,794
GAAIS_9	21,00	32,055	0,745	0,802
GAAIS_10	21,04	31,902	0,819	0,792
GAAIS_15	21,03	35,260	0,614	0,821
GAAIS_19	20,66	34,747	0,606	0,822
GAAIS_20	21,27	36,583	0,461	0,840

Оцінка Шкали конспірологічних переконань щодо штучного інтелекту (AICBS) продемонструвала виняткову надійність інструментарію на даній вибірці. Для повної шкали, що складається з 30 пунктів, загальний коефіцієнт альфа Кронбаха склав 0,957, що є практично ідентичним до результатів оригінальної валідизації авторів методики (Lin та ін., 2025). Детальний аналіз окремих субфакторів також підтвердив їхню високу внутрішню узгодженість (Таблиця 5-9): 1) субшкала «Глобальний контроль» (12 пунктів) продемонструвала показник α=0,923, 2) «Дезінформація» (5 пунктів) – α=0,864, 3) «Людська праця та інтелект» (6 пунктів) – α=0,835, 4) «Суперництво озброєнь» (5 пунктів) – α=0,852, 5) «Міжособистісні стосунки» (2 пункти) – α=0,817. Настільки високі значення навіть для коротких субшкал свідчать про концептуальну цілісність методики та її здатність прецизійно вимірювати різні специфічні грані когнітивних упереджень індивідів.

Таблиця 5

Результати розрахунку альфа Кронбаха для AICBS, субшкала «Глобальний контроль»

Пункт	**Середнє шкали при вилученні пункту**	**Дисперсія шкали при вилученні пункту**	**Скоригована кореляція пункту із загальним результатом**	**Альфа Кронбаха при вилученні пункту**
AICBS_блок1_1	29,68	147,839	0,643	0,957
AICBS_блок1_2	29,91	144,114	0,777	0,953
AICBS_блок1_3	29,62	143,389	0,783	0,953
AICBS_блок1_4	29,65	145,656	0,668	0,957
AICBS_блок1_5	30,22	140,966	0,855	0,951
AICBS_блок1_6	30,18	140,202	0,896	0,950
AICBS_блок1_7	30,28	141,795	0,852	0,951
AICBS_блок1_8	29,51	144,363	0,755	0,954
AICBS_блок1_9	30,39	144,214	0,825	0,952
AICBS_блок1_10	30,07	139,571	0,854	0,951
AICBS_блок1_11	29,84	143,179	0,726	0,955
AICBS_блок1_12	29,95	141,805	0,823	0,952

Таблиця 6

Результати розрахунку альфа Кронбаха для AICBS, субшкала «Дезінформація»

Пункт	**Середнє шкали при вилученні пункту**	**Дисперсія шкали при вилученні пункту**	**Скоригована кореляція пункту із загальним результатом**	**Альфа Кронбаха при вилученні пункту**
AICBS_блок2_1	12,61	16,214	0,716	0,855
AICBS_блок2_2	13,24	17,556	0,599	0,881
AICBS_блок2_3	12,35	16,505	0,724	0,854
AICBS_блок2_4	12,19	15,635	0,762	0,844
AICBS_блок2_5	12,58	15,617	0,775	0,841

Таблиця 7

Результати розрахунку альфа Кронбаха для AICBS, субшкала «Людська праця та інтелект»

Пункт	**Середнє шкали при вилученні пункту**	**Дисперсія шкали при вилученні пункту**	**Скоригована кореляція пункту із загальним результатом**	**Альфа Кронбаха при вилученні пункту**
AICBS_блок3_1	33,27	106,227	0,666	0,928
AICBS_блок3_2	33,03	101,917	0,829	0,922
AICBS_блок3_3	32,77	104,700	0,685	0,927
AICBS_блок3_4	33,42	104,165	0,692	0,927
AICBS_блок3_5	32,68	103,839	0,694	0,927
AICBS_блок3_6	33,50	104,418	0,572	0,932
AICBS_блок3_7	32,69	102,655	0,739	0,925
AICBS_блок3_8	32,96	102,067	0,728	0,925
AICBS_блок3_9	33,12	105,752	0,680	0,927
AICBS_блок3_10	33,16	104,603	0,743	0,925
AICBS_блок3_11	32,78	100,720	0,783	0,923
AICBS_блок3_12	32,32	105,318	0,662	,928

Таблиця 8

Результати розрахунку альфа Кронбаха для AICBS, субшкала «Суперництво озброєнь»

Пункт	Середнє шкали при вилученні пункту	Дисперсія шкали при вилученні пункту	Скоригована кореляція пункту із загальним результатом	Альфа Кронбаха при вилученні пункту
AICBS_блок4_1	18,30	38,130	0,659	0,934
AICBS_блок4_2	18,20	36,191	0,778	0,923
AICBS_блок4_3	18,47	36,335	0,823	0,919
AICBS_блок4_4	18,38	35,307	0,854	0,916
AICBS_блок4_5	18,34	36,008	0,784	0,923
AICBS_блок4_6	18,84	35,973	0,790	0,922
AICBS_блок4_7	18,23	35,111	0,805	0,921

Таблиця 9

Результати розрахунку альфа Кронбаха для AICBS, субшкала «Міжособистісні стосунки»

Пункт	Середнє шкали при вилученні пункту	Дисперсія шкали при вилученні пункту	Скоригована кореляція пункту із загальним результатом	Альфа Кронбаха при вилученні пункту
AICBS_блок5_1	22,14	48,940	0,650	0,920
AICBS_блок5_2	22,38	47,690	0,727	0,913
AICBS_блок5_3	22,15	46,457	0,824	0,906
AICBS_блок5_4	22,27	46,666	0,822	0,906
AICBS_блок5_5	22,76	48,433	0,702	0,915
AICBS_блок5_6	22,47	47,157	0,784	0,909
AICBS_блок5_7	22,64	49,112	0,702	0,915
AICBS_блок5_8	22,18	47,982	0,707	0,915

Крім того, аналіз значень зміни надійності при гіпотетичному видаленні окремих пунктів засвідчив, що вилучення жодного з тверджень не призвело б до статистично значущого підвищення загальної надійності шкал. Це вказує на високу якість застосованих перекладних версій опитувальників та їхню здатність адекватно відображати психологічний

зміст досліджуваних конструктів. Таким чином, підтверджені високі психометричні характеристики методик AICBS та GAAIS, де показники перевищують прийнятний у поріг 0,70, сформували надійне статистичне підґрунтя для подальшого застосування методів індуктивної статистики та регресійного аналізу.

Проведений емпіричний аналіз дозволяє перевірити теоретичні конструкти алгоритмічної відрази та упередженості автоматизації на конкретній вибірці респондентів (N = 74), зіставляючи їхні вікові особливості та особистісні диспозиції. Відповідно до початкової гіпотези очікувалося, що дорослі учасники демонструватимуть вищий рівень алгоритмічної відрази та конспірологічних переконань як захисну реакцію власної ідентичності, тоді як молодь, соціалізована в цифровому середовищі, буде більш схильною до некритичної довіри та автоматизаційного упередження. Однак результати порівняльного аналізу із застосуванням непараметричного U-критерію Манна-Уітні (Таблиця 10) не отримали емпіричного підтвердження наявності вираженої міжгенераційної дихотомії. Статистичні показники загального позитивного ставлення ($p = 0,319$) та негативного ставлення ($p = 0,561$) виявилися статистично подібними в обох вікових когортах. Попри теоретичні припущення про те, що дорослі гостріше реагують на загрозу втрати контролю або девальвації робочих місць (Granulo та ін., 2019), емпіричні дані нашої вибірки, наведені у Таблиці 11, засвідчили відсутність достовірних розбіжностей щодо страху глобального контролю з боку штучного інтелекту ($p = 0,159$) чи ризиків поширення дезінформації ($p = 0,288$). Таке нівелювання вікових меж може свідчити про глибоку гомогенізацію цифрового досвіду у сучасному суспільстві, де онтологічна невизначеність алгоритмічних систем викликає універсальні когнітивні реакції незалежно від етапу розвитку виконавчих функцій мозку індивіда.

Таблиця 10

Результати порівняльного аналізу із застосуванням непараметричного U-критерію Манна-Уітні

№	Нульова гіпотеза	Критерій (Test)	Знач. (Sig.)	Рішення (Decision)
1	Розподіл GAAIS_Positive однаковий для всіх категорій age_group (вікова група).	U-критерій Манна-Вітні для незалежних вибірок	0,319	Прийняти нульову гіпотезу.
2	Розподіл GAAIS_Negative однаковий для всіх категорій age_group (вікова група).	U-критерій Манна-Вітні для незалежних вибірок	0,561	Прийняти нульову гіпотезу.

Таблиця 11

Результати порівняльного аналізу показників GAAIS та AICBS за допомогою t-критерію Стьюдента для незалежних вибірок

Змінна		t-критерій для рівності середніх						
		t	df	Знач. (2-бічна)	Різниця середніх	Станд. похибка різниці	95% довірчий інтервал різниці	
							Нижня	Верхня
GAAIS_Positive	Передбачається рівність дисперсій	-1,130	72	0,262	-0,24140	0,21360	0,667	0,18441
	Не передбачається рівність дисперсій	-1,008	24,418	0,323	-0,24140	0,23945	0,735	0,25234
GAAIS_Negative	Передбач.	0,364	72	0,717	0,08358	0,22972	0,374	0,54153
	Не передбач.	0,345	26,559	0,732	0,08358	0,24196	0,4132	0,58043
Глобальний контроль	Передбач.	-1,423	72	0,159	-0,41551	0,29203	0,9976	0,16663
	Не передбач.	-1,404	28,150	0,171	-0,41551	0,29599	1,021	0,19066
Дезінформація	Передбач.	1,070	72	0,288	0,28810	0,26928	0,2487	0,82489
	Не передбач.	0,999	25,933	0,327	0,28810	0,28843	0,3048	0,88104
Загроза людській праці	Передбач.	0,182	72	0,856	0,04580	0,25193	0,4564	0,54802
	Не передбач.	0,175	27,104	0,862	0,04580	0,26171	0,4911	0,58270
Суперництво озброєнь та мілітаризація	Передбач.	0,453	72	0,652	0,12302	0,27150	0,4182	0,66423
	Не передбач.	0,472	30,805	0,641	0,12302	0,26082	-0,409	0,65510
Міжособистісні стосунки	Передбач.	-0,213	72	0,832	-0,0570	0,26791	-0,5911	0,47703
	Не передбач.	-0,222	30,773	0,826	-0,05704	0,25753	-0,5824	0,46834

Спроба пояснити рівень алгоритмічної відрази через призму стабільних особистісних рис за моделлю HEXACO також продемонструвала несподівані результати, які змушують переглянути традиційні погляди на природу когнітивних викривлень. Множинний лінійний регресійний аналіз засвідчив, що сукупність шести особистісних диспозицій та віку здатна пояснити лише 8,6% дисперсії негативного ставлення до штучного інтелекту (R-квадрат = 0,086; p = 0,522) та 4,9% варіативності загального конспірологічного мислення (R-квадрат = 0,049; p = 0,845), Таблиця 12 (модель 1 і 2 відповідно). Жоден із предикторів, включаючи очікувано впливові показники емоційності (яка асоціюється з тривожністю) чи відкритості до досвіду, не досяг порогу статистичної значущості. Цей емпіричний факт суперечить окремим попереднім даним щодо лінійного впливу індивідуальної тривожності на сприйняття інновацій (Grassini & Koivisto, 2024), проте він дає можливість припускати, що автоматизаційне упередження та алгоритмічна відраза є високо ситуативними феноменами. На їх формування в реальних умовах значно сильніше впливають контекст когнітивного завдання, загальний рівень інформаційної грамотності та вплив медійного дискурсу, аніж стабільні вроджені характеристики темпераменту чи характеру користувача.

Таблиця 12

Результати множинного лінійного регресійного аналізу

Модель	R	R-квадрат	Скоригований R-квадрат	Стандартна похибка оцінки
1	0,293	0,086	-0,011	0,84740
2	0,220	0,049	-0,052	0,86738

Водночас, попри відсутність прямого впливу віку та особистісних рис, емпіричний аналіз розкрив глибинну внутрішню структуру когнітивних викривлень, підтвердивши їхню комплексну системність. За допомогою кореляційного аналізу за критерієм Спірмена було виявлено потужні внутрішні зв'язки між різними аспектами упередженості. Зафіксовано сильний від'ємний зв'язок між реверсивною шкалою загального негативного ставлення (де вищий бал означає менше побоювань) та всіма конспірологічними субшкалами, який

досягає максимуму у вимірі страху глобального контролю ($\rho = -0,697$; $p < 0,001$) та загрози витіснення людської праці ($\rho = -0,545$; $p < 0,001$). Водночас усі складові конспірологічного мислення, включаючи дезінформацію, загрозу миру та вплив на міжособистісні стосунки, високо позитивно корелюють між собою ($p < 0,001$). З психологічної точки зору наявність тісних внутрішніх зв'язків між субшкалами статистично доводить, що когнітивні упередження щодо цифрових алгоритмів мають виражений монологічний характер: індивід, схильний вірити в одну теорію змови, з високою ймовірністю генералізує цю недовіру на всі інші аспекти функціонування технології. Якщо індивід формує упередження або відчуває загрозу в одному специфічному домені (наприклад, у сфері економічної безпеки), він несвідомо генералізує цю недовіру на всі інші аспекти функціонування штучного інтелекту, утворюючи щільну систему конспірологічного мислення, яка стає резистентною до раціональних доказів чи об'єктивних статистичних даних (Lin та ін., 2025).

2.3. Порівняльна характеристика динаміки когнітивних упереджень та особистісних корелятів молоді й дорослих

З метою емпіричної верифікації гіпотези щодо наявності виражених вікових відмінностей у структурі когнітивних упереджень стосовно штучного інтелекту було здійснено порівняльний аналіз груп молоді (n = 18) та дорослих (n = 56). Зважаючи на відхилення від нормального розподілу за окремими шкалами, розрахунки здійснювалися із застосуванням непараметричного U-критерію Манна-Уітні та параметричного t-критерію Стьюдента. Отримані результати засвідчили, що показники загального позитивного (p = 0,319) та негативного (p = 0,561) ставлення до штучного інтелекту є статистично подібними в обох вікових когортах, Таблиця 10. Незважаючи на теоретичні припущення про те, що дорослі гостріше реагують на загрозу втрати контролю, емпіричні дані не виявили достовірних розбіжностей щодо страху глобального контролю з боку штучного інтелекту (p = 0,159) чи ризиків поширення дезінформації (p = 0,278). Відповідно, гіпотеза дослідження не отримала емпіричного підтвердження на досліджуваній вибірці, що може пояснюватися явищем універсальної цифрової соціалізації, за якої інтенсивна інтеграція алгоритмічних систем у всі сфери життєдіяльності нівелює вікові межі у сприйнятті технологій (Di Salvo, 2025). Онтологічна невизначеність штучних агентів провокує універсальні еволюційно детерміновані реакції, які базуються на концептах «народної психології» незалежно від віку суб'єкта, змушуючи як молодь, так і дорослих екстраполювати людські інтенції на алгоритми (Laakasuo та ін., 2021).

Наступним етапом дослідження стала перевірка ролі специфічних особистісних рис за моделлю HEXACO як предикторів загального негативного ставлення та конспірологічних переконань. Результати кореляційного аналізу за критерієм Спірмена засвідчили відсутність статистично значущих лінійних зв'язків між базовими диспозиціями особистості та показниками ставлення до штучного інтелекту (усі показники p > 0,05). Зокрема, припущення про те, що низький рівень відкритості до досвіду детермінує

конспірологічне мислення, не підтвердилося емпірично ($p > 0,152$). Було виявлено лише слабкі тенденції, що не досягли порогу значущості: додатний зв'язок між емоційністю та вірою у глобальний контроль ($\rho = 0,219$; $p = 0,060$), а також від'ємний зв'язок між чесністю-скромністю та побоюваннями щодо глобального контролю ($\rho = -0,215$; $p = 0,066$). Відсутність сильних кореляцій частково суперечить результатам окремих попередніх досліджень, які фіксували вагомий вплив екстраверсії та сумлінності на частоту використання та позитивне ставлення до систем генеративного штучного інтелекту (Kiński та ін., 2025). Водночас отримані результати цілком узгоджуються з тезою про те, що в умовах високої контекстуальної невизначеності ситуативні фактори та специфіка когнітивного завдання домінують над стабільними рисами особистості під час оцінки алгоритмів (Castelo, 2019).

Попри відсутність вираженого зв'язку з особистісними диспозиціями, емпіричний аналіз виявив надзвичайно щільну внутрішню кореляцію між різними формами когнітивних упереджень щодо штучного інтелекту. Зафіксовано сильний від'ємний зв'язок між шкалою загального негативного ставлення та всіма субшкалами конспірологічних переконань. Найсильніше цей зв'язок виражено щодо страху глобального контролю ($\rho = -0,697$; $p < 0,001$) та загрози витіснення людської праці ($\rho = -0,545$; $p < 0,001$). Усі складові конспірологічного мислення, враховуючи дезінформацію, загрозу миру та вплив на міжособистісні стосунки, також високо корелюють між собою ($p < 0,001$), результати наведені у Таблиці 13. З психологічної точки зору це статистично доводить, що когнітивні упередження щодо цифрових алгоритмів мають виражений монологічний характер, оскільки індивід, схильний вірити в одну теорію змови, з високою ймовірністю генералізує цю недовіру на всі інші аспекти функціонування технології, формуючи цілісну конспірологічну ментальність (Lin та ін., 2025).

Таблиця 13

Кореляційний аналіз між негативною субшкалою GAAIS та субшкалами AICBS

			GAAIS Негативне	Глобальний контроль	Дезінформація	Загроза людській праці	Суперництво озброєнь та мілітаризація	Міжособистісні стосунки
ро Спірмена	**GAAIS Негативне**	Коефіцієнт кореляції	1,000	-0,697**	-0,540**	-0,545**	-0,402**	-0,550**
		Знач. (двобіч.)	-	0,000	0,000	0,000	0,000	0,000
		N	74	74	74	74	74	74
	Глобальний контроль	Коефіцієнт кореляції	-0,697**	1,000	0,604**	0,712**	0,609**	0,542**
		Знач. (двобіч.)	0,000	-	0,000	0,000	0,000	0,000
		N	74	74	74	74	74	74
	Дезінформація	Коефіцієнт кореляції	-0,540**	0,604**	1,000	0,644**	0,540**	0,469**
		Знач. (двобіч.)	0,000	0,000	-	0,000	0,000	0,000
		N	74	74	74	74	74	74
	Загроза людській праці	Коефіцієнт кореляції	-0,545**	0,712**	0,644**	1,000	0,756**	0,761**
		Знач. (двобіч.)	0,000	0,000	0,000	-	0,000	0,000
		N	74	74	74	74	74	74
	Суперництво озброєнь та мілітаризація	Коефіцієнт кореляції	-0,402**	0,609**	0,540**	0,756**	1,000	0,552**
		Знач. (двобіч.)	0,000	0,000	0,000	0,000	-	0,000
		N	74	74	74	74	74	74
	Міжособистісні стосунки	Коефіцієнт кореляції	-0,550**	0,542**	0,469**	0,761**	0,552**	1,000
		Знач. (двобіч.)	0,000	0,000	0,000	0,000	0,000	-
		N	74	74	74	74	74	74

**. Кореляція є значущою на рівні 0,01 (двобічна).

Для остаточної верифікації прогностичної здатності комплексу особистісних рис та віку було застосовано множинний лінійний регресійний аналіз методом примусового введення змінних. Як вже було зазначено у підрозділі 2.1, у першій моделі, де залежною змінною виступало негативне ставлення до штучного інтелекту, сукупність предикторів пояснила лише 8,6% загальної дисперсії (R-квадрат = 0,086), при цьому сама модель виявилася статистично незначущою за критерієм Фішера ($F = 0,887$; $p = 0,522$). Найбільший, проте незначущий, внесок продемонстрували відкритість до досвіду ($\beta = 0,201$; $p = 0,160$) та емоційність ($\beta = -0,182$; $p = 0,201$), тоді як віковий фактор також не набув прогностичної цінності ($\beta = 0,166$; $p = 0,232$). Аналогічні результати було отримано і в другій регресійній моделі для інтегрального показника конспірології, де коефіцієнт детермінації становив усього 4,9% (R-квадрат = 0,049; $F = 0,481$; $p = 0,845$). Жодна з незалежних змінних, включаючи вік ($\beta = 0,105$; $p = 0,458$) та доброзичливість ($\beta = -0,100$; $p = 0,432$), не продемонструвала достовірного впливу. Показники фактора інфляції дисперсії (VIF) для всіх змінних знаходилися в межах від 1,121 до 1,445, що виключає проблему мультиколінеарності та підтверджує математичну коректність побудованих регресійних моделей.

Узагальнюючи отримані емпіричні дані, правомірно констатувати, що когнітивні упередження щодо штучного інтелекту виступають як автономний психологічний конструкт не детермінується ані віковою приналежністю, ані стабільними диспозиціями за моделлю HEXACO. Оскільки регресійні моделі пояснюють менше 10% варіативності алгоритмічної відрази та конспірологічних переконань, ставлення до новітніх технологій формується під впливом інших екзогенних та індивідуальних чинників. На формування алгоритмічної відрази значно сильніше впливає ступінь суб'єктивності завдання та потреба в автентичному емоційному досвіді (Bigman & Gray, 2018), а також рівень сприйманої загрози для професійної ідентичності працівника (Granulo та ін., 2019). Зі іншого боку, автоматизаційне упередження ефективніше пояснюється механізмами когнітивної

прив'язки та ефектом слабких доказів під час безпосередньої взаємодії з інтерфейсом системи (Rastogi та ін., 2022).

Таким чином, проблема профілактики та подолання когнітивних викривлень виходить за межі суто індивідуально-типологічного профілювання і гостро потребує комплексного соціотехнічного підходу, орієнтованого на підвищення прозорості інтелектуальних систем та цілеспрямований розвиток інформаційної грамотності у користувачів.

2.4. Соціально-демографічні та технологічні детермінанти когнітивної упередженості щодо штучного інтелекту

Для комплексного розуміння природи когнітивних викривлень у системі «людина – штучний інтелект» емпіричний аналіз було розширено шляхом дослідження впливу соціально-демографічних чинників (зокрема, статевої ідентифікації) та інтенсивності (частоти) використання технологій штучного інтелекту. Загальний обсяг досліджуваної вибірки становив 74 особи. Аналіз здійснювався за допомогою методів параметричної та непараметричної статистики, що дозволило забезпечити високий рівень достовірності й обґрунтованості отриманих результатів.

Перевірка впливу статевої приналежності на формування когнітивних упереджень передбачала порівняння трьох груп респондентів: жінок (n = 45), чоловіків (n = 28) та небінарних осіб (n = 1). На попередньому етапі було здійснено перевірку гомогенності дисперсій за допомогою тесту Левена. Для всіх досліджуваних шкал цей показник продемонстрував рівень значущості вище нормативного порогу ($p > 0,05$). Це підтвердило дотримання базових передумов та дозволило коректно застосувати параметричний дисперсійний аналіз (ANOVA). Результати дисперсійного аналізу засвідчили відсутність статистично значущого впливу статі на жоден із досліджуваних показників (Таблиця 14). Зокрема, для шкали загального позитивного ставлення до штучного інтелекту показник значущості становив $p = 0,122$, а для показника загрози витіснення людської праці – $p = 0,556$. Ці результати були підтверджені непараметричним критерієм Краскела-Уолліса, за результатами якого для всіх пунктів було ухвалено рішення зберегти нульову гіпотезу (Таблиця 15). Відповідно, емпіричні дані доводять, що чоловіки та жінки демонструють гомогенні установки та рівні занепокоєння стосовно штучного інтелекту. Це свідчить про універсальність сприйняття технологічних загроз та відсутність виражених гендерних розбіжностей у контексті когнітивних упереджень. Водночас слід зауважити, що група небінарних осіб у даному дослідженні була представлена лише одним учасником, що

обмежує можливості широкого статистичного узагальнення для цієї специфічної демографічної категорії.

Таблиця 14

Результати параметричного дисперсійного аналізу ANOVA, статева приналежність

		Сума квадратів	df	Середній квадрат	F	Знач. (Sig.)
GAAIS_Positive	Між групами	2,618	2	1,309	2,165	0,122
	Всередині груп	42,924	71	0,605		
	Загалом	45,542	73			
GAAIS_Negative	Між групами	2,375	2	1,187	1,704	0,189
	Всередині груп	49,477	71	0,697		
	Загалом	51,852	73			
Глобальний контроль	Між групами	5,004	2	2,502	2,194	0,119
	Всередині груп	80,985	71	1,141		
	Загалом	85,990	73			
Дезінформація	Між групами	1,107	2	0,553	0,552	0,578
	Всередині груп	71,138	71	1,002		
	Загалом	72,245	73			
Загроза людській праці	Між групами	1,020	2	0,510	0,591	0,556
	Всередині груп	61,258	71	0,863		
	Загалом	62,277	73			
Суперництво озброєнь та мілітаризація	Між групами	1,153	2	0,577	0,574	0,566
	Всередині груп	71,344	71	1,005		
	Загалом	72,498	73			
Міжособистісні стосунки	Між групами	1,974	2	0,987	1,024	0,365
	Всередині груп	68,466	71	0,964		
	Загалом	70,440	73			

Таблиця 15

Результати перевірки статистичних гіпотез щодо відмінностей у показниках за статтю (критерій Краскела-Уолліса)

№	Нульова гіпотеза	Знач. (Sig.)
1	Розподіл GAAIS_Positive однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,233
2	Розподіл GAAIS_Negative однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,147
3	Розподіл Глобального контролю однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,095
4	Розподіл Дезінформації однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,392

5	Розподіл Загрози людській праці однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,449
6	Розподіл Суперництва озброєнь... однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,461
7	Розподіл Міжособистісних стосунків однаковий для всіх категорій 1 - жінка, 2 чоловік, 3 - небінарна особа.	0,368

Для оцінки впливу інтенсивності взаємодії з технологіями штучного інтелекту на когнітивні та соціальні установки респондентів було проведено порівняльний аналіз із використанням параметричного дисперсійного аналізу (ANOVA) та непараметричного критерію Краскела-Уолліса. На початковому етапі обробки даних вихідні вісім категорій частоти використання було перекодовано у 3 групи: 1) особи, які використовують ШІ рідко (n = 14); 2) особи, які використовують технології іноді (n = 22); 3) особи, які користуються ними часто (n = 38). Такий підхід дозволив забезпечити більш рівномірний розподіл респондентів (N = 74).

Фундаментальною передумовою застосування параметричного дисперсійного аналізу виступила перевірка гомогенності дисперсій у сформованих групах (Таблиця 16). Результати тесту Левена підтвердили правомірність використання параметричних методів, оскільки для ключових досліджуваних конструктів показник асимптотичної значущості суттєво перевищив критичний поріг 0,05. Зокрема, для конспірологічних субшкал глобального контролю (p = 0,994) та загрози світовому миру (p = 0,951). Дотримання цієї статистичної вимоги дозволяє розглядати подальші результати дисперсійного аналізу як достовірні та репрезентативні для цієї вибірки.

Таблиця 16

Результати перевірки гомогенності дисперсій за критерієм Левена для показників ставлення до систем штучного інтелекту

	Статистика Левена	**df1**	**df2**	**Знач. (Sig.)**
GAAIS_Positive	0,881	2	71	0,419
GAAIS_Negative	0,065	2	71	0,937
Глобальний контроль	0,006	2	71	0,994

Дезінформація	0,344	2	71	0,710
Загроза людській праці	0,279	2	71	0,757
Суперництво озброєнь та мілітаризація	0,050	2	71	0,951
Міжособистісні стосунки	0,270	2	71	0,764

Згідно з результатами однофакторного дисперсійного аналізу (ANOVA), поданими у Таблиці 17, зафіксовано наявність статистично значущих відмінностей між групами респондентів за ключовими показниками конспірологічної упередженості. Достовірний вплив частоти використання виявлено щодо сприйняття загрози витіснення людської праці ($F(2, 71) = 3,408$; $p = 0,039$) та побоювань стосовно мілітаризації технологій і загрози світовому миру ($F(2, 71) = 3,975$; $p = 0,023$). Аналіз показників центральної тенденції вказує на лінійну обернену залежність між досвідом взаємодії та рівнем технологічної тривожності. Респонденти, які використовують штучний інтелект рідко (група 1), демонструють найвищий рівень занепокоєння як щодо працевлаштування ($M = 3,32$), так і щодо глобальної безпеки ($M = 3,56$). Водночас у групі активних користувачів (група 3) ці показники є нижчими ($M = 2,73$ та $M = 2,77$ відповідно).

Таблиця 17

Результати порівняльного аналізу показників сприйняття ШІ залежно від частоти взаємодії з технологіями за критерієм ANOVA

		Сума квадратів	**df**	**Середній квадрат**	**F**	**Знач. (Sig.)**
GAAIS_Positive	Між групами	2,328	2	1,164	1,913	0,155
	Всередині груп	43,214	71	0,609		
	Загалом	45,542	73			
GAAIS_Negative	Між групами	2,120	2	1,060	1,514	0,227
	Всередині груп	49,732	71	0,700		
	Загалом	51,852	73			
Глобальний контроль	Між групами	6,662	2	3,331	2,981	0,057
	Всередині груп	79,328	71	1,117		
	Загалом	85,990	73			

Дезінформація	Між групами	4,079	2	2,040	2,125	0,127
	Всередині груп	68,165	71	0,960		
	Загалом	72,245	73			
Загроза людській праці	Між групами	5,455	2	2,728	3,408	0,039
	Всередині груп	56,822	71	0,800		
	Загалом	62,277	73			
Суперництво озброєнь та мілітаризація	Між групами	7,300	2	3,650	3,975	0,023
	Всередині груп	65,198	71	0,918		
	Загалом	72,498	73			
Міжособистісні стосунки	Між групами	2,488	2	1,244	1,300	0,279
	Всередині груп	67,952	71	0,957		
	Загалом	70,440	73			

Для підтвердження надійності виявлених ефектів було застосовано непараметричний критерій Краскела-Уолліса (Таблиця 18). Результати аналізу засвідчили наявність статистично значущих відмінностей у розподілі показників за чотирма шкалами опитувальника. Зокрема, інтенсивність взаємодії з ШІ впливає на загальне позитивне ставлення до технології ($p = 0,046$), сприйняття глобального контролю ($p = 0,031$), ставлення до заміни людської праці машинами ($p = 0,024$) та занепокоєння щодо ймовірного зменшення миру у світі ($p = 0,028$).

Таблиця 18

Результати порівняльного аналізу показників ставлення до ШІ за критерієм Краскела-Уолліса залежно від досвіду взаємодії з технологіями

№	**Нульова гіпотеза**	**Критерій**	**Знач. (Sig.)**	**Рішення**
1	Розподіл GAAIS_Positive однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,046	Відхилити нульову гіпотезу.
2	Розподіл GAAIS_Negative однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,150	Прийняти нульову гіпотезу.
3	Розподіл Глобального контролю однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,031	Відхилити нульову гіпотезу.

4	Розподіл Дезінформації однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,083	Прийняти нульову гіпотезу.
5	Розподіл Загрози людській праці однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,024	Відхилити нульову гіпотезу.
6	Розподіл Суперництва озброєнь... однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,028	Відхилити нульову гіпотезу.
7	Розподіл Міжособистісних стосунків однаковий для всіх категорій 1 - рідко, 2 - іноді, 3 - часто.	Критерій Крускала-Волліса для незалежних вибірок	0,384	Прийняти нульову гіпотезу.

Для подальшого уточнення характеру виявлених відмінностей та визначення того, між якими саме підгрупами існують значущі розбіжності, було застосовано процедуру апостеріорних (Post Hoc) попарних порівнянь із використанням консервативної поправки Бонферроні. Найбільш виражену емпіричну закономірність зафіксовано за шкалою «Загрози мілітаризації». Виявлено статистично достовірну різницю ($p = 0,033$) між респондентами, які звертаються до штучного інтелекту «рідко», та тими, хто використовує його «часто». Аналіз середніх рангів вказує на вищий рівень тривоги та занепокоєння глобальними загрозами саме серед осіб із мінімальним досвідом взаємодії з технологією. Це явище можна інтерпретувати через механізм демістифікації: практичний досвід застосування ШІ дозволяє користувачам більш реалістично оцінювати можливості алгоритмів, знижуючи рівень катастрофізації та ірраціональних страхів, часто нав'язаних масовою культурою.

Специфічним феноменом виявилися результати попарних порівнянь для шкал позитивного ставлення до ШІ, глобального контролю та заміни людської праці. Попри те, що загальний критерій Краскела-Уолліса підтвердив наявність відмінностей між групами ($p < 0,05$), застосування жорсткої поправки Бонферроні знівелювало ці результати на рівні безпосередніх попарних порівнянь (всі показники $p > 0,05$). З точки зору статистичного аналізу, це свідчить про те, що вплив частоти використання ШІ на ці аспекти має скоріше градієнтний, кумулятивний характер. Зміни у сприйнятті контролю чи позитивному

ставленні розподіляються між усіма трьома групами поступово, без різких стрибків між полярними категоріями.

Узагальнення результатів емпіричного аналізу соціально-демографічних та технологічних детермінант дозволяє зробити низку важливих наукових висновків щодо природи формування когнітивних упереджень у процесі взаємодії людини зі штучним інтелектом.

По-перше, встановлено, що базова архітектоніка сприйняття технологічної невизначеності є переважно гомогенною та не детермінується статевою приналежністю суб'єкта. Відсутність статистично значущих гендерних відмінностей за показниками загального позитивного чи негативного ставлення, а також щодо більшості конспірологічних страхів (зокрема, глобального контролю, дезінформації та впливу на ринок праці) доводить, що первинні когнітивні реакції на штучних агентів є універсальним продуктом загальної цифрової соціалізації.

По-друге, дослідження дозволяє припускати, що ключовим модератором алгоритмічної відрази та конспірологічного мислення виступає інтенсивність практичного технологічного досвіду. Частота використання обернено пов’язана з рівнем упередженості, знижуючи страхи перед глобальним поневоленням чи витісненням людської праці (Lin та ін., 2025). Психологічний механізм цього явища полягає у деконструкції ефекту «чорної скриньки»: завдяки регулярній взаємодії користувачі переходять від міфологізованого чи апокаліптичного сприйняття алгоритмів до прагматичного усвідомлення їхніх функціональних обмежень та інструментальної корисності. Цей процес когнітивної адаптації супроводжується закономірним зростанням рівня позитивного ставлення до інновацій та формуванням відкаліброваної довіри (Schepman & Rodway, 2023). Також важливо зазначити, що цілком ймовірно, що не частота використання знижує побоювання, а низькі побоювання дозволяють індивідам користуватись ШІ активніше.

По-третє, зниження загального фону тривожності під час оцінки цифрових систем може теоретично пояснюватися механізмом когнітивної адаптації, де здобуття регулярного практичного досвіду стимулює перехід від емоційно-інтуїтивного реагування до раціонально-аналітичного осмислення результатів алгоритмічної діяльності, однак цей аспект потребує подальшої емпіричної верифікації з фокусом на особистісну емоційність.

Зрештою, отримані емпіричні результати можуть мати важливе практичне значення для розробки сучасних соціотехнічних стратегій. Оскільки рівень ірраціональних побоювань та алгоритмічної відрази обернено пропорційний частоті використання штучного інтелекту, ефективна профілактика упереджень вимагає відмови від фокусування виключно на демографічному чи особистісному профілюванні. Натомість стратегічним пріоритетом психологічної та освітньої практики має стати створення інклюзивних умов для масового залучення індивідів до безпечної практичної взаємодії з алгоритмами. Саме безпосередній технологічний досвід здатен забезпечити швидку трансформацію абстрактної конспірологічної тривожності у науково обґрунтоване та критичне прийняття новітніх інтелектуальних систем (Murthy та ін., 2025).

Висновки до розділу 2

Узагальнення результатів емпіричного дослідження, спрямованого на вивчення психологічних предикторів та механізмів формування когнітивних упереджень щодо штучного інтелекту, дозволяє сформулювати низку концептуальних висновків. Проведений комплексний математико-статистичний аналіз дає можливість припускати, що традиційні уявлення про соціодемографічну та диспозиційну детермінацію алгоритмічної відрази та автоматизаційного упередження потребують суттєвого перегляду. Емпірична перевірка на досліджуваній вибірці не виявила статистично значущих розбіжностей у сприйнятті штучного інтелекту між групами молоді та дорослих, що, з огляду на обмеження репрезентативності вибірки, дає підстави для попереднього припущення про можливу гомогенізацію цифрового досвіду в сучасному суспільстві. (Di Salvo, 2025; Laakasuo та ін., 2021).

Аналогічним чином, регресійний аналіз засвідчив статистичну незначущість побудованих моделей, що свідчить про відсутність вираженого лінійного впливу стабільних рис особистості за шестифакторною моделлю HEXACO на досліджуваній вибірці. Оскільки такі диспозиції не продемонстрували достовірної прогностичної цінності щодо технологічної тривожності чи довіри, правомірно висунути гіпотезу про багатофакторну детермінацію цих феноменів. Можна припустити, що формування когнітивних викривлень у системі «людина – штучний інтелект» здатне опосередковуватися ситуативними характеристиками, де специфіка завдання та рівень його суб'єктивності потенційно чинять вагоміший вплив, ніж індивідуально-типологічні властивості користувача (Castelo, 2019; Grassini & Koivisto, 2024). Разом із тим, дослідження розкрило щільну внутрішню кореляцію між різними формами конспірологічних страхів, статистично довівши монологічну природу упередженості: суб'єкт, схильний відчувати загрозу в одному специфічному домені, автоматично генералізує цю недовіру на всі інші аспекти

функціонування технології, утворюючи комплексну та стійку до раціональних доказів систему переконань (Lin та ін., 2025).

Проте фундаментальним фактором, який безпосередньо моделює архітектоніку когнітивних викривлень, виявилася інтенсивність практичного технологічного досвіду. Частота взаємодії з генеративними моделями виявилася вагомим емпіричним чинником, що достовірно обернено корелює зі специфічними аспектами конспірологічного мислення, зокрема, сприяючи зниженню ірраціональних побоювань щодо загрози мілітаризації та порушення світового миру в учасників даної вибірки.

Зрештою, результати другого розділу доводять, що подолання упередженості автоматизації та алгоритмічної відрази неможливо забезпечити лише шляхом демографічного чи індивідуально-типологічного профілювання користувачів. Оскільки регулярна взаємодія деконструює ефект «чорної скриньки» та трансформує ірраціональну тривожність у прагматичне розуміння функціональних обмежень машини, стратегічним завданням є створення інклюзивного середовища для масового практичного залучення індивідів до роботи з інтелектуальними алгоритмами (Murthy та ін., 2025). Саме такий емпіричний досвід гарантує формування адекватної, відкаліброваної довіри та мінімізацію ризиків дезадаптивного прийняття рішень у новітньому цифровому середовищі.

РОЗДІЛ 3. ПСИХОЛОГІЧНЕ ОБҐРУНТУВАННЯ ПРОГРАМИ КОРЕКЦІЇ КОГНІТИВНИХ УПЕРЕДЖЕНЬ ЩОДО ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

3.1. Обґрунтування програми психологічного супроводу та розвитку цифрової грамотності й критичного мислення в контексті взаємодії з ШІ

Результати проведеного емпіричного аналізу на даній вибірці засвідчили, що формування когнітивних упереджень стосовно систем штучного інтелекту не детермінується виключно вродженими особистісними диспозиціями чи хронологічним віком користувачів, а має виражену контекстуальну та поведінкову природу. Зважаючи на те, що інтенсивність практичної взаємодії з алгоритмами виявилася найпотужнішим предиктором зниження конспірологічного мислення та алгоритмічної відрази (Lin et al., 2025; Schepman & Rodway, 2022), розробка спеціалізованої програми психологічного супроводу постає як нагальна науково-практична необхідність. Концептуальне обґрунтування такої програми спирається на синтез положень когнітивної психології, теорії подвійних процесів прийняття рішень та концепцій інформаційної грамотності, що дозволяє розглядати подолання упереджень як керований процес психологічної та епістемічної адаптації індивіда до новітнього соціотехнічного середовища (Kliegr et al., 2021).

Фундаментальною мішенню розробленої програми виступає феномен «когнітивного розвантаження» (cognitive offloading), який розглядається як ключовий механізм зниження критичного мислення в епоху штучного інтелекту. Сучасні дослідження припускають, що надмірне делегування когнітивних завдань, зокрема пошуку інформації, аналітики та прийняття рутинних рішень, інтелектуальним системам призводить до когнітивних лінощів, за яких користувачі втрачають навички глибокого аналітичного мислення (Gerlich, 2025). Оскільки системи штучного інтелекту здатні генерувати переконливі, хоча й не завжди об'єктивно достовірні відповіді, користувачі потрапляють у пастку ілюзії знання та автоматизаційного упередження, пасивно споживаючи алгоритмічний контент (Rastogi et

al., 2022). Відповідно, програма психологічного супроводу повинна бути спрямована на формування когнітивної стійкості, що передбачає цілеспрямовану стимуляцію самостійного аналітичного мислення та навчання стратегіям виявлення алгоритмічних помилок («галюцинацій»), аби запобігти потраплянню особистості в інформаційні «ехо-камери» (Gerlich, 2025; Єрмошкіна, 2025).

Нейропсихологічне підґрунтя програми спирається на інтеграцію так званих когнітивних примусових функцій (cognitive forcing functions), розроблених у межах теорії швидкого та повільного мислення. Згідно з Kahneman (2011), у ситуаціях інформаційного перевантаження людська психіка схильна покладатися на швидку, інтуїтивну Систему 1, що провокує виникнення ефекту алгоритмічної прив'язки (anchoring bias), за якого перша ж підказка машини стає некритичним базисом для прийняття рішення. Для протидії цій тенденції програма передбачає впровадження технік штучного уповільнення процесу прийняття рішень, які змушують суб'єкта активувати аналітичну Систему 2 (Rastogi et al., 2022). Це досягається через навчання алгоритмам рефлексивного запитування, структурування етапів перевірки фактів та свідомого генерування альтернативних гіпотез до того, як буде прийнято фінальний машинний висновок (Kliegr et al., 2021). Такий підхід дозволяє змістити фокус із пасивної довіри на раціональний сумнів та об'єктивну оцінку валідності алгоритмічних рекомендацій.

Важливим вектором психологічного супроводу є калібрування довіри до штучного інтелекту, що виступає центральним елементом розвитку сучасної цифрової грамотності (Lee & See, 2004). Екстремальні прояви як алгоритмічної відрази, так і упередженості автоматизації (сліпої довіри) є дезадаптивними патернами поведінки. Психологічна корекція цих крайнощів вимагає деміфологізації технологій та подолання ефекту «чорної скриньки», який безпосередньо стимулює виникнення конспірологічних страхів (Lin et al., 2025). Програма має обов'язково включати психоедукаційний компонент, орієнтований на роз'яснення принципів роботи систем машинного навчання, їхньої імовірнісної, а не

абсолютної природи, та наявності вбудованих обмежень. Набуття такого розуміння формує в індивіда епістемічну компетентність: усвідомлення того, що штучний інтелект є не автономним моральним агентом чи загрозою екзистенційного масштабу, а складним математичним інструментом, ефективність якого напряму залежить від якості вхідних даних (Bigman & Gray, 2018).

Крім того, програма психологічного супроводу повинна враховувати емоційні та екзистенційні аспекти взаємодії, оскільки побоювання щодо втрати професійної ідентичності, загрози мілітаризації технологій чи заміщення людського спілкування є вагомими тригерами алгоритмічної відрази, особливо у ситуаціях, які вимагають емпатії або суб'єктивного судження (Castelo, 2019; Granulo et al., 2019). Інтеграція принципів гуманістичної психології та соціальної когніції дозволяє ефективно працювати зі страхами втрати суб'єктності, допомагаючи учасникам переосмислити роль штучного інтелекту як інструменту аугментації (розширення) людських можливостей, а не їх повної заміни (Raisch & Krakowski, 2021). В умовах тотальної цифровізації вкрай важливо зберегти потребу в міжособистісній взаємодії та автентичному емоційному досвіді, що вимагає розвитку навичок рефлексії власних станів під час роботи з машинами.

Таким чином, структурно-функціональна модель програми корекції когнітивних упереджень щодо штучного інтелекту обґрунтовано включає три взаємопов'язані блоки: 1) когнітивно-просвітницький блок, орієнтований на підвищення рівня знань про архітектуру та обмеження нейромереж для нівелювання конспірологічних теорій; 2) аналітико-поведінковий блок, спрямований на розвиток навичок критичного мислення, протидію когнітивному розвантаженню та імплементацію когнітивних примусових функцій у процес прийняття рішень; 3) емоційно-рефлексивний блок, який забезпечує подолання екзистенційної тривоги, трансформацію сприйняття технологічної загрози та формування відкаліброваної, адаптивної довіри до інтелектуальних систем. Імплементація такої комплексної програми дозволить оптимізувати процеси спільної діяльності в системі

«людина – штучний інтелект», забезпечуючи високу продуктивність за умови збереження когнітивної автономії та психологічного благополуччя особистості.

3.2. Структура та зміст тренінгової програми розвитку критичного мислення та раціональної взаємодії з технологіями штучного інтелекту

Спираючись на результати емпіричного дослідження, які засвідчили, що інтенсивність практичного досвіду є головним предиктором зниження конспірологічної упередженості та формування адекватного ставлення до новітніх алгоритмів, було розроблено спеціалізовану тренінгову програму. Головною метою цієї програми є трансформація пасивного когнітивного розвантаження, за якого індивід некритично делегує прийняття рішень машинам, у процес свідомої, раціональної та критичної взаємодії. Структурно-тематична модель тренінгу побудована за модульним принципом і охоплює три ключові напрями психологічної інтервенції: 1) когнітивно-просвітницький, 2) аналітико-поведінковий та 3) емоційно-рефлексивний, кожен з яких спрямований на деконструкцію специфічних проявів алгоритмічної відрази або упередженості автоматизації.

Перший, когнітивно-просвітницький модуль, спрямований на підвищення рівня базової алгоритмічної грамотності (AI literacy) та подолання ефекту «чорної скриньки», який виступає фундаментальним підґрунтям для виникнення конспірологічних страхів. Як засвідчують дані розробників шкали AICBS, ірраціональна віра у глобальний алгоритмічний контроль або загрозу дезінформації виникає переважно через нерозуміння принципів машинного навчання (Lin та ін., 2025). Змістове наповнення цього модуля передбачає ознайомлення учасників із вірогіднісною, а не детерміністичною природою великих мовних моделей, механізмами збору тренувальних даних та причинами виникнення алгоритмічних галюцинацій. Психологічна мішень цього етапу полягає у деміфологізації технології: учасники повинні усвідомити, що штучний інтелект є не автономним суб'єктом із власними інтенціями чи мораллю, а складним математичним інструментом, який відображає статистичні закономірності людської мови та поведінки (Bigman & Gray, 2018). Таке розуміння формує первинну епістемічну стійкість та знижує схильність до антропоморфізації цифрових систем.

Другий, аналітико-поведінковий модуль, є центральним компонентом програми та фокусується на безпосередньому розвитку навичок критичного мислення в умовах інформаційного перевантаження. Зміст модуля базується на таксономії критичного мислення Д. Халперн та передбачає тренування навичок перевірки гіпотез, аналізу аргументації та оцінки ймовірностей (Gerlich, 2025). Оскільки взаємодія зі штучним інтелектом часто провокує ефект алгоритмічної прив'язки (anchoring bias), за якого людина схильна беззаперечно приймати першу згенеровану машиною відповідь, у програму інтегровано методику «когнітивних примусових функцій» (Rastogi та ін., 2022). Ця техніка передбачає штучне уповільнення процесу прийняття рішення: перед тим як делегувати завдання нейромережі, учасники тренуються самостійно формулювати очікуваний результат, а після отримання машинної відповіді – застосовувати чек-листи для верифікації фактів та виявлення прихованих упереджень у згенерованому тексті (Kliegr та ін., 2021). Такий підхід змушує психіку перемикатися з інтуїтивної евристичної обробки інформації (Система 1) на глибокий аналітичний контроль (Система 2), ефективно протидіючи когнітивним лінощам та пасивному споживанню контенту (Darwin та ін., 2024; Gerlich, 2025).

Третій, емоційно-рефлексивний модуль, присвячений калібруванню довіри до технологій та опрацюванню екзистенційної тривоги. Згідно з концепцією відповідної довіри (appropriate reliance), ефективна взаємодія з автоматизованими системами вимагає балансу між надмірною довірою та безпідставним відторгненням (Lee & See, 2004). Зміст цього модуля спрямований на подолання страхів, пов'язаних із загрозою втрати професійної ідентичності чи витісненням людської праці, які, як доведено, провокують найсильніший психологічний опір (Granulo та ін., 2019). За допомогою методів групової рефлексії та проблемно-орієнтованого навчання учасники опрацьовують власні емоційні реакції на взаємодію з алгоритмами. Фокус уваги зміщується з парадигми «заміни» на парадигму «аугментації» (розширення), де штучний інтелект розглядається як партнерський

інструмент, що звільняє когнітивний ресурс людини для вищих форм творчості та емпатійної комунікації.

Методологічною особливістю реалізації всієї тренінгової програми є принцип обов'язкового занурення (experiential learning). Оскільки емпіричний аналіз у другому розділі здійснив верифікацію на даній вибірці, що частота та глибина безпосереднього використання інструментів штучного інтелекту є найсильнішим фактором зниження конспірологічних переконань, програма не обмежується теоретичними лекціями. Кожне заняття супроводжується практичною взаємодією з генеративними нейромережами, де учасники цілеспрямовано провокують алгоритми на помилки, стикаються з їхніми функціональними обмеженнями та навчаються формулювати складні аналітичні запити. Саме такий керований практичний досвід дозволяє перетворити абстрактну технологічну тривожність на відкалібровану довіру, формуючи стійкий імунітет до когнітивних маніпуляцій та алгоритмічної дискримінації в сучасному цифровому середовищі (Schepman & Rodway, 2023).

Окрім трьох базових модулів, ефективність тренінгової програми суттєво підвищується за рахунок інтеграції специфічних когнітивних стратегій та технік взаємодії, які безпосередньо впливають на архітектоніку прийняття рішень і ще не розглядалися в основній структурі. Важливим інноваційним компонентом програми має стати навчання основам психологічно обґрунтованого написання тексу вводу або «prompt engineering». Цей напрям базується на розумінні того, що ефективність взаємодії з алгоритмами залежить від здатності користувача враховувати принципи салієнтності (виразності) інформації, механізми пам'яті та когнітивні схеми під час формулювання запитів (Chaudhuri & Bhattacharya, 2025). Учасники тренінгу повинні навчитися структурувати свої звернення до штучного інтелекту таким чином, щоб активувати релевантні інформаційні схеми машини, використовуючи чіткі пошукові підказки (retrieval cues) та оптимальний рівень когнітивної складності, що запобігає отриманню розмитих чи неточних алгоритмічних відповідей.

Невіддільною частиною розвитку навичок взаємодії є тренування здатності формулювати контрастивні запитання. Згідно з дослідженнями у галузі соціальної когніції та філософії штучного інтелекту, людське пояснення за своєю природою є контрастивним: індивіди шукають відповідь не просто на питання про те, чому сталася подія P, а на питання, чому сталася подія P, а не очікувана альтернативна подія Q (Miller, 2019). Навчання користувачів застосовувати контрастивні запити під час роботи зі штучним інтелектом, вимагаючи від системи пояснити, чому вона обрала певне рішення замість конкретної альтернативи, дозволяє ефективніше розкривати логіку «чорної скриньки» та знижує рівень сліпої довіри до першої згенерованої відповіді (Miller, 2019).

Додатковим інструментом подолання упередженості автоматизації в межах програми є впровадження практик соціально-когнітивної усвідомленості. На відміну від традиційних медитативних практик, цей підхід, зокрема техніка «помічання нових деталей», спрямований на цілеспрямовану стимуляцію аналітичного мислення та виведення суб'єкта зі стану когнітивного автоматизму під час оцінки згенерованого контенту (Thiedmann та ін., 2025). У поєднанні зі стратегією «розгляду протилежного», ця техніка ефективно протидіє упередженню підтвердження та стратегії позитивного тестування, за яких люди схильні шукати лише ту інформацію, що підтверджує їхні попередні гіпотези (Kliegr та ін., 2021). Учасники програми отримують завдання штучно провокувати нейромережі на генерування контраргументів до їхніх власних переконань, що стимулює інтелектуальну гнучкість та формує навичку жорсткого наукового скептицизму щодо будь-якої неперевіреної інформації (Darwin та ін., 2024; Kliegr та ін., 2021).

Зрештою, програма повинна озброїти користувачів розумінням комунікативної природи штучного інтелекту через призму правил кооперативної бесіди, відомих у лінгвопсихології як максими Грайса. Сучасні розмовні агенти часто імітують дотримання максим якості, кількості, релевантності та манери спілкування, створюючи хибну ілюзію усвідомленого та експертного діалогу (Miller, 2019). Розуміння того, що багатослівність або

надмірна деталізація з боку алгоритму може спричиняти когнітивний «ефект розбавлення», коли великий обсяг нерелевантної інформації притупляє пильність та знижує здатність людини до критичного аналізу, є критично важливим для інформаційної гігієни (Miller, 2019). Розвиток стійкості до таких лінгвістичних та структурних маніпуляцій допомагає користувачам відокремлювати форму подачі від фактичного змісту, перетворюючи їх із пасивних споживачів алгоритмічного продукту на свідомих архітекторів своєї безпечної цифрової діяльності.

Висновки до розділу 3

Узагальнення теоретичних та емпіричних результатів дослідження актуалізувало необхідність розробки спеціалізованої програми психологічної корекції когнітивних упереджень у процесі взаємодії людини зі штучним інтелектом. Теоретико-методологічним підґрунтям такої програми виступив синтез положень когнітивної психології, теорії подвійних процесів прийняття рішень та концепцій інформаційної грамотності. Фундаментальною мішенню психологічного впливу визначено феномен когнітивного розвантаження, який в умовах тотальної цифровізації провокує зниження аналітичного контролю та призводить до некритичного споживання алгоритмічного контенту, формуючи ілюзію знання (Gerlich, 2025; Rastogi та ін., 2022). З огляду на це, стратегічним завданням розробленої програми є деконструкція упередженості автоматизації та алгоритмічної відрази шляхом трансформації пасивної довіри у раціональний науковий сумнів та розвиток епістемічної компетентності користувачів (Kliegr та ін., 2021).

Структурно-функціональна модель запропонованої тренінгової програми побудована за модульним принципом і охоплює три взаємопов'язані напрями інтервенції. Перший, когнітивно-просвітницький модуль, спрямований на деміфологізацію технологій, подолання ефекту «чорної скриньки» та зниження загального рівня конспірологічного мислення через глибоке усвідомлення ймовірнісної природи машинного навчання (Lin та ін., 2025). Другий, аналітико-поведінковий модуль, забезпечує безпосередній розвиток навичок критичного мислення завдяки імплементації когнітивних примусових функцій, що штучно уповільнюють прийняття рішень та стимулюють перемикання психіки на аналітичну систему обробки інформації (Rastogi та ін., 2022). Важливим інноваційним компонентом цього етапу є навчання психологічно обґрунтованому інжинірингу підказок та використанню контрастивних запитів, що дозволяє користувачам ефективно активувати релевантні когнітивні схеми машини та долати алгоритмічну маніпуляцію (Miller, 2019). Третій, емоційно-рефлексивний модуль, орієнтований на опрацювання екзистенційної

тривоги, страху втрати суб'єктності та формування відкаліброваної, відповідної довіри до технологій як до інструментів когнітивної аугментації, а не повної заміни людських здібностей (Lee & See, 2004).

Центральним методологічним принципом реалізації програми є обов'язкове емпіричне занурення та безпосередня практична взаємодія учасників з генеративними алгоритмами. Враховуючи результати попереднього статистичного аналізу, які довели превалюючу роль частоти використання технологій у нівелюванні упереджень, саме керований практичний досвід, поєднаний із техніками соціально-когнітивної усвідомленості, є найефективнішим механізмом корекції викривлень (Thiedmann та ін., 2025). Впровадження такої програми забезпечує комплексний соціотехнічний підхід, що дозволяє індивідам абстрагуватися від ілюзії експертного діалогу, обходити ефект розбавлення інформації та перетворюватися з пасивних споживачів алгоритмічних продуктів на свідомих, критично мислячих архітекторів своєї інтелектуальної діяльності. У підсумку це сприяє оптимізації спільної діяльності в системі «людина – штучний інтелект», гарантуючи високу продуктивність за умови збереження психологічного благополуччя та когнітивної автономії особистості.

ВИСНОВКИ

Узагальнення результатів теоретико-емпіричного дослідження, присвяченого вивченню психологічних закономірностей та когнітивних упереджень у процесі взаємодії людини зі штучним інтелектом, дає підстави для формулювання низки концептуальних висновків. Стрімка інтеграція інтелектуальних алгоритмів у всі сфери суспільного та індивідуального буття трансформувала штучний інтелект із суто технологічного інструменту на складну соціотехнічну систему, що безпосередньо впливає на архітектоніку людського пізнання та прийняття рішень. Теоретичний аналіз проблеми засвідчив, що людська психіка, еволюційно пристосована до обробки інформації в природному та соціальному середовищах, виявляється когнітивно вразливою перед лицем високошвидкісних імовірнісних обчислень, що призводить до систематичних викривлень у сприйнятті алгоритмічних результатів (Kliegr та ін., 2021).

Фундаментальним теоретичним висновком роботи є обґрунтування дихотомії неадаптивних когнітивних реакцій на штучний інтелект, що розкривається через феномени упередженості автоматизації та алгоритмічної відрази. Упередженість автоматизації виникає як наслідок прагнення мозку до когнітивного розвантаження, за якого людина-оператор некритично делегує прийняття рішень машині, покладаючись на її підказку як на потужний когнітивний якір, що пригнічує аналітичну систему мислення та стимулює інтуїтивне, недостатньо обґрунтоване прийняття рішень (Rastogi та ін., 2022). На протилежному полюсі знаходиться алгоритмічна відраза, яка має не стільки раціональну, скільки афективну та екзистенційну природу. Теоретично доведено, що рівень недовіри до алгоритмів жорстко детермінується ступенем суб'єктивності завдання: якщо в об'єктивних обчисленнях люди демонструють алгоритмічну вдячність, то у завданнях, що вимагають емоційного інтелекту, емпатії або морального судження, довіра до машин статистично значущо падає (Bigman & Gray, 2018; Castelo, 2019).

Крім того, теоретичний аналіз виявив, що вагомим тригером алгоритмічної відрази виступає відчуття екзистенційної загрози та загрози ідентичності. Перспектива заміни людської праці програмним забезпеченням сприймається суб'єктом як глибока небезпека для власної Я-концепції, що генерує значно потужніший психологічний опір, ніж суто фінансові чи економічні ризики (Granulo та ін., 2019). Як захисний механізм психіка актуалізує антропоцентричне упередження, суть якого полягає у систематичному знеціненні результатів алгоритмічної діяльності (зокрема, творчих продуктів) задля збереження відчуття людської винятковості та унікальності (Millet та ін., 2023).

Теоретичне осмислення проблеми також доводить, що формування конспірологічних переконань щодо штучного інтелекту є наслідком хибного застосування механізмів соціальної атрибуції до неживих об'єктів. Через нерозуміння математичної природи машинного навчання та вплив ефекту «чорної скриньки» люди схильні антропоморфізувати технології, наділяючи алгоритми власними інтенціями, віруваннями та мораллю, що лежить в основі фольклорної етики штучного інтелекту та провокує ірраціональні страхи перед глобальним контролем чи світовою мілітаризацією (Laakasuo та ін., 2021; Lin та ін., 2025). З огляду на це, стратегічним завданням у процесі оптимізації спільної діяльності є формування відкаліброваної, відповідної довіри, яка вимагає подолання обох когнітивних крайнощів шляхом глибокого психологічного та епістемічного супроводу користувачів (Lee & See, 2004).

Емпірична перевірка висунутої гіпотези дозволила докорінно переглянути традиційні уявлення про соціодемографічну та індивідуально-типологічну детермінацію ставлення до новітніх технологій. Результати статистичного аналізу на опитуваній вибірці довели відсутність значущих розбіжностей у сприйнятті штучного інтелекту між групами молоді та дорослих. Цей факт свідчить про глибоку гомогенізацію цифрового досвіду в сучасному суспільстві та підтверджує феномен універсальної цифрової соціалізації, за якої онтологічна невизначеність алгоритмів провокує однакові еволюційно детерміновані

реакції незалежно від віку індивіда (Di Salvo, 2025). Аналогічним чином, емпірична перевірка не підтвердила статистичної значущості впливу стабільних рис особистості за моделлю HEXACO на формування алгоритмічної відрази та автоматизаційного упередження в межах досліджуваної вибірки. Відсутність достовірних лінійних зв'язків між базовими диспозиціями та когнітивними викривленнями не дозволяє вважати їх єдиними чи визначальними предикторами технологічної тривожності. Отримані результати актуалізують необхідність подальшого наукового пошуку і дають підстави для теоретичного припущення, що когнітивні упередження щодо штучного інтелекту можуть мати більш складну, ймовірно ситуативну та контекстуальну природу, яка потребує окремого експериментального вивчення (Grassini & Koivisto, 2024).

Водночас емпіричне дослідження розкрило глибинну внутрішню структуру когнітивних викривлень, статистично підтвердивши їхній монологічний характер. Зафіксовано щільну кореляцію між усіма субшкалами конспірологічних страхів: індивід, який відчуває загрозу в одному специфічному домені (наприклад, у сфері втрати робочих місць), несвідомо генералізує цю недовіру на всі інші аспекти функціонування технологій, формуючи стійку та резистентну до раціональних аргументів систему переконань (Lin та ін., 2025).

Фундаментальним емпіричним спостереженням роботи є виявлення тісного оберненого зв'язку між інтенсивністю використання інструментів штучного інтелекту та загальним рівнем когнітивної упередженості. Зафіксовано, що регулярна практична взаємодія супроводжується прагматичнішим розумінням функціональних обмежень машини, що дає підстави розглядати технологічне занурення як імовірний засіб усунення когнітивних упереджень. Крім того, активні користувачі демонструють достовірно нижчий рівень загальної емоційної лабільності під час взаємодії з цифровим середовищем, що актуалізує питання про потенційну стабілізуючу функцію практичного досвіду (Schepman & Rodway, 2023).

Спираючись на отримані емпіричні докази того, що когнітивні упередження ефективно долаються саме через поведінковий досвід, у практичній частині дослідження було психологічно обґрунтовано та розроблено спеціалізовану тренінгову програму. Її стратегічною мішенню визначено феномен когнітивного розвантаження, який в умовах інформаційного перенасичення провокує когнітивні лінощі, некритичне делегування рішень машині та виникнення автоматизаційного упередження (Gerlich, 2025). Для протидії цим дезадаптивним тенденціям запропоновано структурно-функціональну модель інтервенції, що базується на синтезі когнітивної психології, теорії подвійних процесів та концепцій медіаграмотності.

Програма реалізується через три взаємопов'язані модулі. Когнітивно-просвітницький модуль забезпечує деміфологізацію технологій та подолання конспірологічного мислення шляхом роз'яснення ймовірнісної природи алгоритмів. Аналітико-поведінковий модуль орієнтований на безпосередній розвиток критичного мислення через імплементацію когнітивних примусових функцій, що штучно уповільнюють прийняття рішень та активують аналітичну систему обробки інформації (Rastogi та ін., 2022). На цьому етапі користувачі також опановують техніки психологічно обґрунтованого інжинірингу підказок та використання контрастивних запитів для подолання алгоритмічних маніпуляцій (Miller, 2019). Емоційно-рефлексивний модуль спрямований на опрацювання страху втрати суб'єктності та формування відповідної, відкаліброваної довіри, яка розглядає штучний інтелект як допоміжний інструмент, а не екзистенційну загрозу (Lee & See, 2004).

Головним методологічним принципом розробленої програми є обов'язкове емпіричне занурення. Враховуючи результати математичної статистики, саме керована практична взаємодія з алгоритмами у поєднанні з техніками соціально-когнітивної усвідомленості (Thiedmann та ін., 2025) визнається найефективнішим інструментом трансформації користувача з пасивного споживача на свідомого архітектора власної інтелектуальної діяльності. Упровадження запропонованої програми сприятиме оптимізації спільної

діяльності в системі «людина – штучний інтелект», забезпечуючи високу продуктивність прийняття рішень за умови збереження психологічного благополуччя, епістемічної стійкості та когнітивної автономії особистості у сучасному цифровому світі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Глибовець, М., & Олецький, О. (2002). *Штучний інтелект* (Видавничий дім «КМ Академія»).
2. Єрмошкіна, О. В. (2025). Формування когнітивних упереджень на фінансовому ринку в умовах розвитку штучного інтелекту. *Економіка, управління та адміністрування*, (3(113)), 111—118. https://doi.org/10.26642/ema-2025-3(113)-111-118
3. Карпенко, З., & Климпуш, А. (2025). Досвід Україномовної Адаптації Особистісного Опитувальника Hexaco M. C. Ashton & K. Lee. *Psychological Prospects Journal*, (46), 39—62. https://doi.org/10.29038/2227-1376-2025-46-kar
4. Соколова, Н. О., & Мошик, М. С. (2023). ВИКЛИКИ ШТУЧНОГО ІНТЕЛЕКТУ. *Електротехнічні та інформаційні системи*, (104), 9—17. https://doi.org/10.32782/EIS/2023-104-2
5. Aberegg, S. K., Haponik, E. F., & Terry, P. B. (2005). Omission Bias and Decision Making in Pulmonary and Critical Care Medicine. *Chest*, *128*(3), 1497—1505. https://doi.org/10.1378/chest.128.3.1497
6. Ashton, M. C., & Lee, K. (2009). An Investigation of Personality Types within the HEXACO Personality Framework. *Journal of Individual Differences*, *30*(4), 181—187. https://doi.org/10.1027/1614-0001.30.4.181
7. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59—64. https://doi.org/10.1038/s41586-018-0637-6
8. Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., & Horvitz, E. (2019). Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 2429—2437. https://doi.org/10.1609/aaai.v33i01.33012429

9. Barrett, D. G., Morcos, A. S., & Macke, J. H. (2019). Analyzing biological and artificial neural networks: Challenges with opportunities for synergy? *Current Opinion in Neurobiology*, *55*, 55—64. https://doi.org/10.1016/j.conb.2019.01.007
10. Berthet, V. (2022). The Impact of Cognitive Biases on Professionals' Decision-Making: A Review of Four Occupational Areas. *Frontiers in Psychology*, *12*, 802439. https://doi.org/10.3389/fpsyg.2021.802439
11. Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21—34. https://doi.org/10.1016/j.cognition.2018.08.003
12. Boit, S., & Patil, R. (2025). A Prompt Engineering Framework for Large Language Model–Based Mental Health Chatbots: Conceptual Framework. *JMIR Mental Health*, *12*, e75078—e75078. https://doi.org/10.2196/75078
13. Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*.
14. Castelo, N. (2019). *Blurring the Line Between Human and Machine: Marketing Artificial Intelligence*.
15. Chandler, C., Foltz, P. W., Cohen, A. S., Holmlund, T. B., Cheng, J., Bernstein, J. C., Rosenfeld, E. P., & Elvevåg, B. (2020). Machine learning for ambulatory applications of neuropsychological testing. *Intelligence-Based Medicine*, *1—2*, 100006. https://doi.org/10.1016/j.ibmed.2020.100006
16. Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, *11*(1), 2290342. https://doi.org/10.1080/2331186X.2023.2290342
17. Dastres, R., & Soori, M. (2021). *Artificial Neural Network Systems*.
18. Di Salvo, M. (2025). Neuroscience and education in the transition from analog to AI. *Journal of Neuroeducation*, *6*(1). https://doi.org/10.1344/joned.v6i1.49388

19. Echterhoff, J. M., Liu, Y., Alessa, A., McAuley, J., & He, Z. (2024). Cognitive Bias in Decision-Making with LLMs. *Findings of the Association for Computational Linguistics: EMNLP 2024*, 12640—12653. https://doi.org/10.18653/v1/2024.findings-emnlp.739

20. Ferrer, X., Nuenen, T. V., Such, J. M., Cote, M., & Criado, N. (2021). Bias and Discrimination in AI: A Cross-Disciplinary Perspective. *IEEE Technology and Society Magazine*, *40*(2), 72—80. https://doi.org/10.1109/MTS.2021.3056293

21. Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

22. Gherheş, V. (2018). Why Are We Afraid of Artificial Intelligence (Ai)? *European Review Of Applied Sociology*, *11*(17), 6—15. https://doi.org/10.1515/eras-2018-0006

23. Granulo, A., Fuchs, C., & Puntoni, S. (2019). Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour*, *3*(10), 1062—1069. https://doi.org/10.1038/s41562-019-0670-y

24. Grassini, S., & Koivisto, M. (2024). Understanding how personality traits, experiences, and attitudes shape negative bias toward AI-generated artworks. *Scientific Reports*, *14*(1), 4113. https://doi.org/10.1038/s41598-024-54294-4

25. Hikmawati, A., & Mohammad, N. K. (2025). Enhancing Critical Thinking with Gen AI: A Literature Review. *Buletin Edukasi Indonesia*, *4*(01), 40—46. https://doi.org/10.56741/bei.v4i01.764

26. Hu, A. (2024). *Developing an AI-Based Psychometric System for Assessing Learning Difficulties and Adaptive System to Overcome: A Qualitative and Conceptual Framework* (Версія 1). arXiv. https://doi.org/10.48550/ARXIV.2403.06284

27. Hu, X., & Zhu, M. (2024). Were Persulfate-Based Advanced Oxidation Processes Really Understood? Basic Concepts, Cognitive Biases, and Experimental Details. *Environmental Science & Technology*, *58*(24), 10415—10444. https://doi.org/10.1021/acs.est.3c10898

28. Irshad, S., Azmi, S., & Begum, N. (2022). Uses of Artificial Intelligence in Psychology. *Journal of Health and Medical Sciences*, *5*(4). https://doi.org/10.31014/aior.1994.05.04.242

29. Jones, E., & Steinhardt, J. (2022). *Capturing Failures of Large Language Models via Human Cognitive Biases* (Версія 2). arXiv. https://doi.org/10.48550/ARXIV.2202.12299

30. Khona, M., & Fiete, I. R. (2022). Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, *23*(12), 744—766. https://doi.org/10.1038/s41583-022-00642-0

31. Kieslich, K., Lünich, M., & Marcinkowski, F. (2021). The Threats of Artificial Intelligence Scale (TAI): Development, Measurement and Test Over Three Application Domains. *International Journal of Social Robotics*, *13*(7), 1563—1577. https://doi.org/10.1007/s12369-020-00734-w

32. Kiński, C., Kiński, B., & Przyborowska, A. (2025). Big Five personality traits, attitudes towards Artificial Intelligence and the use of AI solutions in foreign language learners. *Neofilolog*, (65/1), 65—83. https://doi.org/10.14746/n.2025.65.1.4

33. Kliegr, T., Bahník, Š., & Fürnkranz, J. (2021). A review of possible effects of cognitive biases on the interpretation of rule-based machine learning models. *Artificial Intelligence*, *295*, 103458. https://doi.org/10.1016/j.artint.2021.103458

34. Korteling, J. E., & Toet, A. (2022). Cognitive Biases. У *Encyclopedia of Behavioral Neuroscience, 2nd edition* (С. 610—619). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.24105-9

35. Kovbasiuk, A., Triantoro, T., Przegalińska, A., Sowa, K., Ciechanowski, L., & Gloor, P. (2025). The personality profile of early generative AI adopters: A big five perspective. *Central European Management Journal*, *33*(2), 252—264. https://doi.org/10.1108/CEMJ-02-2024-0067

36. Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148—1160. https://doi.org/10.1038/s41593-018-0210-5

37. Laakasuo, M., Herzon, V., Perander, S., Drosinou, M., Sundvall, J., Palomäki, J., & Visala, A. (2021). Socio-cognitive biases in folk AI ethics and risk discourse. *AI and Ethics*, *1*(4), 593—610. https://doi.org/10.1007/s43681-021-00060-5

38. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. https://doi.org/10.1017/S0140525X16001837

39. Lee, J. D., & See, K. A. (2004). *Trust in Automation: Designing for Appropriate Reliance*.

40. Lin, C., Brailovskaia, J., Üztemur, S., Gökalp, A., Değirmenci, N., Huang, P., Chen, I., Griffiths, M. D., & Pakpour, A. H. (2025). Dark Future: Development and Initial Validation of Artificial Intelligence Conspiracy Beliefs Scale (AICBS). *Brain and Behavior*, *15*(7), e70648. https://doi.org/10.1002/brb3.70648

41. Lionenko, M., & Huzar, O. (2023). Development of critical thinking in the context of digital learning. *Economics & Education*, *8*(2), 29—35. https://doi.org/10.30525/2500-946X/2023-2-5

42. Maharjan, J., Jin, R., Zhu, J., & Kenne, D. (2025). Intersection of Big Five Personality Traits and Substance Use on Social Media Discourse: AI-Powered Observational Study. *Journal of Medical Internet Research*, *27*, e79454—e79454. https://doi.org/10.2196/79454

43. Mariani, M., & Baggio, R. (2022). Big data and analytics in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management*, *34*(1), 231—278. https://doi.org/10.1108/IJCHM-03-2021-0301

44. Mariani, M. M., Perez-Vega, R., & Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, *39*(4), 755—776. https://doi.org/10.1002/mar.21619

45. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1—38. https://doi.org/10.1016/j.artint.2018.07.007

46. Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior*, *143*, 107707. https://doi.org/10.1016/j.chb.2023.107707

47. Murthy, Y. S., Bansal, R., Chodisetty, R. S. Ch. M., Chakravorty, C., & Sai, K. P. (2025). Transforming Minds AI and Machine Learning Applications in Cognitive Neuroscience: У R. Bansal, T. Maqableh, G. Shuklaa, F. Rabby, & R. Lathabhavan (Ред.), *Advances in Psychology, Mental Health, and Behavioral Studies* (С. 107—128). IGI Global. https://doi.org/10.4018/979-8-3693-9341-3.ch005

48. Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

49. Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427—469. https://doi.org/10.1007/s40593-014-0029-5

50. Ouyang, F., Dinh, T. A., & Xu, W. (2023). A Systematic Review of AI-Driven Educational Assessment in STEM Education. *Journal for STEM Education Research*, *6*(3), 408—426. https://doi.org/10.1007/s41979-023-00112-x

51. Pum, M. (2024). *Cognitive Bias in AI Development: Understanding the Psychological and Societal Influences that Shape AI Model Design and Outcomes*.

52. Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022). Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1—22. https://doi.org/10.1145/3512930

53. Redelmeier, D. A., & Shafir, E. (2023). The Fallacy of a Single Diagnosis. *Medical Decision Making*, *43*(2), 183—190. https://doi.org/10.1177/0272989X221121343

54. Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human–Computer Interaction*, *39*(13), 2724—2741. https://doi.org/10.1080/10447318.2022.2085400

55. Soleimani, M., Intezari, A., & Pauleen, D. J. (2021). Mitigating Cognitive Biases in Developing AI-Assisted Recruitment Systems: A Knowledge-Sharing Approach. *International Journal of Knowledge Management*, *18*(1), 1—18. https://doi.org/10.4018/IJKM.290022

56. Stein, J.-P., Messingschlager, T., Gnambs, T., Hutmacher, F., & Appel, M. (2024). Attitudes towards AI: Measurement and associations with personality. *Scientific Reports*, *14*(1), 2909. https://doi.org/10.1038/s41598-024-53335-2

57. Thiedmann, P., Dejardin, F., Reiter, L., & Tran, U. S. (2025). Brief Online Socio-Cognitive Mindfulness Interventions Neither Improve Socio-Cognitive Mindfulness nor Cognitive Biases: A Two-Study Conceptual Replication and Reanalysis of a Randomized Controlled Trial. *Mindfulness*, *16*(6), 1555—1568. https://doi.org/10.1007/s12671-025-02575-y

58. Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453—458. https://doi.org/10.1126/science.7455683

59. Varona, D., & Suárez, J. L. (2022). Discrimination, Bias, Fairness, and Trustworthy AI. *Applied Sciences*, *12*(12), 5826. https://doi.org/10.3390/app12125826

60. Wang, J., & Redelmeier, D. A. (2024). Cognitive Biases and Artificial Intelligence. *NEJM AI*, *1*(12). https://doi.org/10.1056/AIcs2400639

61. Zhang, L. (2023). Impact of AI on Human Decision-Making: Analysis of Human, AI, and Environment of Interaction. *Lecture Notes in Education Psychology and Public Media*, *28*(1), 239—245. https://doi.org/10.54254/2753-7048/28/20231348

62. Zhao, J., Wu, M., Zhou, L., Wang, X., & Jia, J. (2022). Cognitive psychology-based artificial intelligence review. *Frontiers in Neuroscience*, *16*, 1024316. https://doi.org/10.3389/fnins.2022.1024316