

ЕВОЛЮЦІЯ МЕТОДІВ АНАЛІЗУ ОСОБИСТІСНИХ РИС З ТЕКСТОВИХ ДАНИХ СОЦІАЛЬНИХ МЕРЕЖ У КОГНІТИВНІЙ ЛІНГВІСТИЦІ: ОГЛЯД ЛІТЕРАТУРИ ЗА ОСТАННІ ДВА ДЕСЯТИЛІТТЯ

THE EVOLUTION OF METHODS FOR ANALYZING PERSONALITY TRAITS FROM SOCIAL MEDIA TEXTS IN COGNITIVE LINGUISTICS: A LITERATURE REVIEW OF THE LAST TWO DECADES

Проаналізовано еволюцію підходів до аналізу особистісних рис із текстових даних соціальних мереж у когнітивній лінгвістиці за останні два десятиліття. Мета дослідження – визначити ключові зміни в теоретичних і методологічних підходах, виявити патерни й розробити рекомендації для майбутніх досліджень. Використано методи літературного огляду, зокрема якісний синтез, тематичний аналіз і хронологічне картування, із залученням даних із провідних академічних баз. Досліджено перехід від лексичних і словникових моделей до когнітивно-лінгвістичних підходів. Встановлено, що сучасні методи, такі як трансформерні моделі та глибоке навчання, підвищили точність прогнозування рис особистості. З'ясовано вплив культурних і платформних факторів на мовні патерни. Показано важливість контекстуальних і лонгітюдних підходів у вивченні змін мовних сигналів із часом. Особливу увагу приділено адаптації моделей для багатомовних та культурно специфічних середовищ, що дозволяє зменшити похибки, пов'язані з мовними та культурними відмінностями. Доведено, що інтеграція когнітивно-лінгвістичних підходів та сучасних обчислювальних технологій сприяє створенню точніших і етично обґрунтованих моделей аналізу особистісних рис у текстових даних. Показано, що застосування explainable AI дозволяє підвищити зрозумілість і прийнятність моделей для практичного використання. Розглянуто перспективи інтеграції міждисциплінарних підходів для підвищення ефективності аналізу текстових даних. Зазначено, що майбутні дослідження мають приділяти увагу впровадженню моделей, здатних адаптуватися до швидких змін у мовних патернах. Наголошено на важливості використання лонгітюдних даних для вивчення тривалих ефектів культурних та соціальних факторів на прояви особистісних рис. Підкреслено необхідність створення методологій, які поєднують високоточні алгоритми з теоретично обґрунтованими лінгвістичними.

Ключові слова: когнітивна лінгвістика, особистісні риси, соціальні мережі, аналіз тексту, машинне навчання, глибоке навчання, моделі трансформерів, лексичний аналіз, метафори та концептуальні схеми, психолінгвістичні ресурси.

The evolution of approaches to analyzing personality traits from social media textual data within the field of cognitive linguistics over the past two decades has been examined. The study aimed to identify key changes in theoretical and methodological approaches, uncover patterns, and develop recommendations for future research. Literature review methods were applied using data from leading academic databases, including qualitative synthesis, thematic analysis, and chronological mapping. The transition from lexical and dictionary-based models to cognitive-linguistic approaches was analyzed. It was established that modern methods, such as transformer models and deep learning, have significantly improved the accuracy of personality trait prediction. The influence of cultural and platform-specific factors on linguistic patterns was identified. The importance of contextual and longitudinal approaches in studying changes in linguistic signals over time was highlighted. Special attention was given to the adaptation of models for multilingual and culturally specific environments, which helps reduce errors associated with linguistic and cultural differences. It was demonstrated that integrating cognitive-linguistic approaches and modern computational technologies facilitates the development of more accurate and ethically sound models for analyzing personality traits in textual data. Applying explainable AI enhanced the interpretability and practical acceptability of such models. The potential for integrating interdisciplinary approaches to improve the effectiveness of text data analysis was discussed. Future research directions were proposed, including developing models capable of adapting to rapid changes in linguistic patterns. The importance of using longitudinal data to study the long-term effects of cultural and social factors on personality expression was emphasized. The necessity of creating methodologies combining precision algorithms with theoretically grounded approaches underlined.

Key words: cognitive linguistics, personality traits, social media, text analysis, machine learning, deep learning, transformer models, lexical analysis, metaphors and conceptual schemas, psycholinguistic resources.

УДК 81'27:004.77]:159.923
DOI <https://doi.org/10.32782/2663-5208.2025.69.1.45>

Чайка Р.М.

доктор філософії у психології,
доцент кафедри теоретичної
та практичної психології
Національний університет
«Львівська політехніка»

Вступ. Вивчення текстових даних соціальних мереж стає все більш важливим у сучасному науковому дискурсі, особливо в контексті когнітивної лінгвістики, яка досліджує взаємозв'язок між мовою, мисленням і соціальними процесами. Розуміння особистісних рис через аналіз текстів відкриває нові можливості для прогнозування поведінки, персоналізації цифрових послуг та створення інноваційних

підходів у психології й лінгвістиці. Соціальні мережі як середовище самовираження забезпечують унікальний доступ до автентичних мовних даних, які дозволяють виявляти глибші когнітивні механізми особистості, що не доступні через традиційні методи, такі як анкетування чи опитування [1; 2].

Науковий інтерес до цієї тематики зумовлений потребою не лише вивчити когнітивні процеси,

які лежать в основі вираження особистісних рис, а й вирішити практичні завдання, такі як удосконалення систем рекомендацій чи аналіз великих масивів тексту для психологічних чи соціальних досліджень. Такі напрями знаходять застосування у сфері цифрових технологій, маркетингу, психологічного консультування й навіть державного управління [3].

Однак, незважаючи на численні огляди літератури, багато робіт зосереджуються на інших темах, не приділяючи уваги комплексному баченню еволюції методів визначення особистісних рис з текстових даних за останні два десятиліття.

Ця прогалина у дослідженнях стає очевидною на тлі зростаючого попиту на інтегративні підходи, що поєднують когнітивну лінгвістику, психологію та комп'ютерні науки.

Мета дослідження – здійснення огляду та аналізу еволюції методів визначення особистісних рис із текстових даних соціальних мереж у контексті когнітивної лінгвістики за останні два десятиліття.

Методи та організація дослідження. У цьому дослідженні здійснено систематичний літературний огляд для аналізу еволюції методів визначення особистісних рис із текстових даних соціальних мереж у контексті когнітивної лінгвістики.

Стратегія пошуку. Для збору наукових джерел використовувалися провідні наукові бази даних: Scopus, Web of Science та Google Scholar. Пошукова стратегія включала комбінації таких ключових слів, як «personality traits analysis», «social media text», «cognitive linguistics», «Big Five personality model», «text analysis methods», «open-vocabulary approach», «language models».

Для забезпечення релевантності було використано такі критерії включення: публікації за період з 2005 до 2024 року; дослідження, які зосереджені на когнітивній лінгвістиці або аналізі тексту для визначення особистісних рис; рецензовані статті, монографії, глави у книгах, а також конференційні матеріали з високим індексом цитування.

Критерії виключення охоплювали роботи, що мали слабкий зв'язок із темою дослідження, оглядові статті без зазначених методологій та публікації, доступні лише в неповному вигляді.

Метод аналізу. Для обробки відібраної літератури застосовано якісний синтез із використанням елементів тематичного аналізу для виявлення ключових концептів і тенденцій. Хронологічне картування дозволило відстежити поступову еволюцію методів – від психолінгвістичних підходів до сучасних комп'ютерних технік.

Вибірка дослідження. Для проведення літературного огляду було відібрано 73 наукові роботи, які відповідають визначеним критеріям включення та виключення.

Типи публікацій. У вибірку увійшли такі типи публікацій: статті з рецензованих журналів: 51 робота, що охоплює провідні видання з когнітивної лінгвістики, комп'ютерної лінгвістики та психології; матеріали міжнародних конференцій:

14 публікацій, присвячених впровадженню нових методів аналізу тексту, зокрема використанню мовних моделей; глави у книгах та монографіях: 8 робіт, які узагальнюють теоретичні основи та міждисциплінарні підходи.

Результати дослідження та їх обговорення. **Зміни в основних теоретичних підходах.** На

початку 2000-х років дослідження визначення особистісних рис із текстових даних у соціальних мережах переважно спиралися на лексичні та словникові підходи, що базувалися на традиційних лінгвістичних рамках. Наприклад, у ранніх дослідженнях широко застосовувалися методи частотного аналізу слів або базові аналізи емоційності текстів [4; 5]. Ці дослідження фокусувалися на підрахунку функціональних слів (наприклад, займенників, артиклів, сполучників) чи емоційно забарвлених слів (наприклад, негативних/позитивних прикметників), припускаючи, що такі поверхневі мовні маркери відображають когнітивні й психологічні стани [6].

Проте з розвитком галузі дослідники почали впроваджувати більш складні когнітивно-лінгвістичні теорії, які визнавали взаємодію між концептуальними знаннями, дискурсивним контекстом і культурним фоном. У фокусі опинилося не лише те, які слова використовуються, але й як вони використовуються та в якому комунікативному контексті [7]. Ця зміна була частково зумовлена появою більш потужних обчислювальних інструментів і доступом до великих корпусів текстів, що дозволяло виявляти складніші лінгвістичні феномени, такі як метафорична мова [8] чи структура наративу. У результаті сучасні моделі враховують прагматику дискурсу та контекстуальне використання, визнаючи, що значення є динамічним і виникає у взаємодії [9].

Одночасно змінилася і концептуалізація самої особистості. Тоді як ранні дослідження в цій галузі майже виключно використовували модель Великої п'ятірки, сучасні роботи дедалі частіше інтегрують додаткові моделі рис або багатовимірні підходи. Наприклад, дослідники почали звертатися до таких альтернативних рамок, як модель HEXACO [10] або спеціалізовані конструкти, такі як Темна тріада (нарцисизм, макіавеллізм, психопатія) [11]. Такі розширення підкреслюють визнання багатогранності особистості й те, що підходи, засновані на одній моделі, можуть пропускати важливі аспекти індивідуальних відмінностей. Інтегруючи ці моделі з когнітивно-лінгвістичними підходами, сучасні дослідження формують більш цілісне уявлення про те, як особистість проявляється у текстовій комунікації, особливо в динамічному та мультимодальному контексті соціальних мереж.

Загалом, основні теоретичні рамки галузі еволюціонували від словниково-орієнтованих і поверхневих лексичних теорій до комплексних, когнітивно обґрунтованих моделей. Ці моделі враховують не лише те, які слова використовуються, а й чому вони використовуються в певних контек-

стах, а також як це використання взаємодії з ширшими концепціями особистості. Така теоретична еволюція створила умови для більш глибокого розуміння особистості через мову, відкриваючи шлях до складніших методів і міждисциплінарних досліджень у майбутньому.

Методологічна еволюція. Від ручного кодування до автоматизованих технік. На ранніх етапах дослідження особистісних рис у текстових даних часто використовували ручне або напівавтоматизоване кодування.

З розширенням обсягів даних соціальних мереж і зростанням доступності обчислювальних ресурсів дослідники перейшли до автоматизованих підходів із використанням обробки природної мови (NLP). Це дозволило швидко аналізувати великі корпуси текстів. Класифікатори машинного навчання, наприклад, логістична регресія, метод опорних векторів (SVM) та випадковий ліс (Random Forest), почали замінювати або доповнювати ручне кодування, значно підвищуючи швидкість та масштабованість виявлення особистісних рис.

Зміни джерел даних. Поряд із методологічним прогресом змінилися й джерела збору даних. У перших дослідженнях (2000-ті роки) використовувалися невеликі вибірки зручного доступу, отримані з онлайн-форумів, блогів або обмежених груп користувачів [12]. Такі невеликі вибірки створювали проблеми з репрезентативністю та варіативністю, що ускладнювало узагальнення результатів. Протягом 2010-их років спостерігається перехід до масштабних, різноманітних наборів даних, отриманих із платформ, таких як Twitter, Instagram та Reddit [1; 13]. Різноманітність цих джерел не лише підвищує статистичну потужність аналізу, але й дозволяє проводити детальніші дослідження, наприклад, крослінгвістичні порівняння або аналіз контекстуальних дискурсів, що раніше були недоступними. У 2020-ті роки розвиток Reddit, TikTok і нових платформ, орієнтованих на текст, таких як Threads, відкрив додаткові можливості для збору та аналізу даних. Крім того, інтеграція мультимедійних даних (тексту, зображень, відео) стала новим вектором досліджень, що забезпечує більш повне розуміння особистісних проявів.

Впровадження глибинного навчання. Значний прорив у методології стався з появою глибинного навчання, що дозволило моделювати більш складні мовні ознаки, пов'язані з особистістю.

Дослідники почали впроваджувати архітектури нейронних мереж, такі як рекурентні нейронні мережі та згорткові нейронні мережі, для врахування часових та ієрархічних залежностей у текстах [14]. Останнім часом трансформерні моделі, такі як BERT, здобули популярність завдяки здатності моделювати складні контекстуальні залежності у значних масштабах [15]. Ці техніки глибинного навчання демонструють підвищену чутливість до тонких мовних сигналів, пов'язаних із рисами особистості, часто перевершуючи традиційні методи машинного навчання за точністю прогнозів [16].

Використання психолінгвістичних ресурсів. Протягом усіх етапів методологічного розвитку психолінгвістичні інструменти відігравали важливу роль у визначенні особистісних рис і їхньої інтерпретації. Зокрема, Linguistic Inquiry and Word Count (LIWC) став основою для аналізу психологічних процесів і емоційних станів через текст [6]. Хоча LIWC та подібні інструменти залишаються широко використовуваними завдяки своїй валідації та інтерпретації, дослідники дедалі частіше інтегрують спеціалізовані словники та автоматизовані інструменти анотації для ідентифікації специфічних мовних маркерів, наприклад, сарказму, жаргону або скорочень, характерних для певних платформ [17].

Таким чином, методологічна еволюція кардинально змінила підходи до вивчення проявів особистості в текстах соціальних мереж. Від трудомістких практик ручного кодування та невеликих вибірок галузь перейшла до високоавтоматизованих, керованих даними підходів, що використовують архітектури глибинного навчання та різноманітні цифрові корпуси. Ця трансформація забезпечує дедалі складніші інсайти про те, як мова відображає, а іноді й формує глибинні риси особистості в онлайн-середовищі.

Валідація та надійність. Еволюція методів валідації. На початкових етапах досліджень у сфері визначення особистісних рис із тексту основним засобом валідації були самозвітні анкети, особливо Big Five Inventory. У таких дослідженнях учасники заповнювали стандартизовані інструменти оцінки особистості, а дослідники корелювали результати самозвітних анкет із особистісними рисами, вилученими з текстів [18]. Хоча цей підхід заклав основу для розуміння зв'язку між мовними маркерами та особистістю, він мав низку обмежень. Зокрема, самозвітні дані можуть бути спотворені через упередження соціальної бажаності або неточне самосприйняття [19].

Протягом останнього десятиліття галузь перейшла до більш складних і багатовимірних стратегій валідації, які враховують різні джерела даних. Замість того, щоб покладатися виключно на самозвітні анкети як на ультимативну правду, дослідження почали включати поведінкові дані, такі як частота публікацій, цифрові сліди та метрики взаємодії, а також характеристики соціальних мереж [20; 1]. Такий підхід до перехресної валідації дозволяє оцінювати конвергентну валідність текстових прогнозів особистості, зіставляючи їх із кількома індексами поведінки й контексту.

Крім того, деякі дослідження інтегрують оцінки інших людей, наприклад, анкетування знайомих учасників для оцінки їхніх рис особистості. Це дозволяє враховувати ті перспективи, які самозвітні тести можуть не відображати повністю [21]. Використовуючи різноманітні джерела даних, сучасні дослідження досягають вищої валідності та знижують ризик упереджень, властивих будь-якому окремому типу вимірювання.

Лонгітюдні та крос-секційні дослідження. Ще одним важливим досягненням у валідації та забезпеченні надійності стало зростання уваги до лонгітюдних досліджень, які дозволяють відстежувати зміни в мовному вираженні особистості з часом. Ранні дослідження здебільшого використовували крос-секційний підхід, збираючи текстові вибірки в одному часовому зрізі та намагаючись вивести стабільні особистісні риси з таких статичних знімків мовлення [4]. Хоча ці дослідження заклали основи для аналізу, вони не враховували часову динаміку, зокрема те, як життєві події, зміни у соціальних платформах чи еволюція онлайн-норм можуть впливати на мовлення.

Натомість дедалі більше сучасних досліджень використовують лонгітюдні дані, збираючи текстові вибірки з інтервалами в кілька тижнів, місяців або навіть років [22; 23]. Спостерігаючи за змінами у мовних профілях протягом тривалого часу, дослідники можуть оцінювати як стабільність мовних ознак, пов'язаних із особистістю, так і ситуаційні фактори, які можуть спричинити короткострокові коливання (наприклад, під час стресових періодів або значущих життєвих переходів).

Такий підхід не лише підвищує надійність, вивчаючи консистентність із часом, але й забезпечує глибші теоретичні інсайти щодо того, як особистість взаємодіє зі змінними соціальними й культурними контекстами. Наприклад, дослідження показують, що деякі мовні маркери можуть тимчасово змінюватися під час значущих особистих потрясінь, але згодом повертаються до базових патернів після відновлення [24].

Таким чином, використання багатосторонніх стратегій валідації та аналізу тимчасових даних дозволяє створити більш деталізовану та точну картину взаємозв'язку між мовленням і особистісними конструкціями.

Інтеграція когнітивно-лінгвістичних підходів. *Метафоричний та концептуальний аналіз.* Ранні дослідження зв'язку між особистісними рисами та мовними патернами здебільшого фокусувалися на поверхневих лексичних індикаторах, таких як емоційно забарвлені слова або використання займенників, і не враховували глибші когнітивно-лінгвістичні феномени, такі як використання метафор чи концептуальне змішування [8; 25]. Більшість підходів базувалася на аналізі частотності слів і простих словникових методах, припускаючи, що особистість можна вивести з прямо-лінійних лексичних сигналів [18].

Однак із розвитком галузі нові дослідження почали вивчати, як метафорична мова та концептуальні схеми формують не лише світогляд людини, а і спосіб вираження таких рис, як відкритість, доброзичливість чи невротизм [7].

Наприклад, аналіз метафор показав, що вибір метафор – чи то просторових («Я відчуваю себе пригніченим»), контейнерних («Я наче зачинений у пляшці»), чи орієнтованих на подорож («Я рухаюся вперед») – може слугувати маркерами

емоційних станів і тенденцій особистості [26]. Деякі дослідники також впровадили теорії концептуального змішування для виявлення того, як взаємодіють кілька ментальних просторів, коли людина виражає складні емоції чи ставлення в онлайн-контекстах [25]. Інтегруючи ці просунуті когнітивно-лінгвістичні рамки в алгоритми аналізу текстів, дослідники відкривають багатші, більш чутливі до контексту сигнали вираження особистості, підкреслюючи тонку взаємодію між концептуальними знаннями та використанням мови.

Контекстуальні підходи. Паралельно з метафоричними та концептуальними аналізами зростала увага до врахування контекстуальних факторів у інтерпретації мовних маркерів. Ранні дослідження тексту й особистості зазвичай використовували універсальну модель, аналізуючи агреговані частотності слів у великих корпусах без достатнього врахування ситуаційних, культурних чи дискурсивних контекстів [4].

З часом дослідники когнітивної лінгвістики довели, що мова не може бути повністю зрозумілою поза рамками, у які вона вписана [27]. Ці рамки – широкі концептуальні структури, які впливають на значення, визначають, як певні слова, метафори чи відсилання сприймаються в конкретному соціальному або культурному середовищі.

Концепт «конструювання» [28] підкреслює, що мовці (або автори текстів) можуть акцентувати або приглушувати різні аспекти ситуації залежно від того, як вони ментально її представляють. Сучасні обчислювальні підходи зараз інтегрують семантику рамок і теорії рівня конструювання для кращої інтерпретації того, як люди проєктують і реалізують особистісні риси в різноманітних онлайн-просторах [9].

Ключові результати та патерни. *Передбачуваність особистості за мовою.* Протягом двох десятиліть досліджень послідовно демонструється зв'язок між специфічними мовними ознаками та особистісними рисами. Ранні дослідження показали, що використання займенників, функціональних слів та загальна емоційна забарвленість тексту були одними з найбільш надійних предикторів особистісних рис, таких як екстраверсія, невротизм і добросовісність [12; 18].

Наприклад, часте використання займенників першої особи однини («я», «мене», «моє») часто асоціюється з невротизмом, а іноді й інтроверсією [24]. Водночас різноманітність позитивних емоційних слів або багатий словниковий запас корелює з відкритістю до досвіду [29]. Згодом ці патерни було уточнено та розширено за рахунок глибшого аналізу синтаксичної складності, когерентності дискурсу та семантичних мереж [1].

Сучасні моделі, які використовують контекстуальні векторні представлення, підкреслили важливість нюансованих контекстно-залежних ознак, таких як використання метафор чи дискурсивних маркерів [7; 16]. Ця еволюція демонструє не лише надійність класичних мовних індикаторів, але

й зростаюче визнання глибших когнітивно-лінгвістичних сигналів.

Варіації мови на різних платформах. Другим важливим відкриттям є те, що специфічні нюанси різних платформ соціальних мереж впливають на вираження та виявлення особистісних рис. Наприклад, у Twitter короткі, часто в режимі реального часу публікації сприяють використанню стислих висловів і емоджі, що може підкреслювати екстраверсію або відкритість [30]. У LinkedIn професійний тон стимулює стилі мовлення, які відповідають добросовісності чи стратегіям самопросування [31]. В Instagram орієнтованість на візуальний контент зміщує акцент із тексту на зображення, але підписи й коментарі все ще виявляють маркери доброзичливості [32]. У Reddit тематично організовані субреддіти дозволяють створювати специфічний дискурс, який може посилювати або приховувати певні маркери залежно від норм спільноти [33].

Отже, хоча загальні мовні індикатори залишаються важливими, прагматичні обмеження та норми кожної платформи формують спосіб вираження особистості та її ідентифікації.

Культурний та крослінгвістичний розвиток. Спочатку більшість робіт у цій галузі були орієнтовані на англійську мову через домінування англо-мовних соціальних платформ і доступності NLP-інструментів для англійської [6]. Однак за останнє десятиліття крослінгвістичні дослідження значно розширилися, виявляючи як універсальні ознаки (наприклад, функціональні слова як загальні маркери когнітивного стилю), так і культурно специфічні вирази (наприклад, культурно зумовлені метафори чи форми ввічливості) [34].

Дослідження іспанської, китайської та арабської мов висвітлили необхідність адаптації або розробки нових мовних ресурсів через відмінності в морфології, писемності та прагматичному використанні [35; 36]. Крім того, культурні норми щодо самовираження, ввічливості та емоційної відкритості можуть суттєво впливати на прояв таких рис, як відкритість чи доброзичливість [37].

Прогрес у точності визначення особистості. Остання ключова тенденція полягає в постійному зростанні точності прогнозування особистісних рис із текстових даних. Ранні дослідження повідомляли про помірні рівні точності через обмежену кількість ознак і невеликі набори даних [12].

Зі зростанням обчислювальних можливостей та доступності даних дослідники вдосконалили моделі за рахунок включення різноманітних лінгвістичних ознак (наприклад, семантичної схожості, прагматичних маркерів) та мультимодальних сигналів (наприклад, використання емоджі, взаємодія із зображеннями) [38].

Поява глибинного навчання й архітектур на основі трансформерів забезпечила значно вищу ефективність у виявленні тональних зрушень, контекстуальних особливостей і дискурсивних патернів користувачів [15; 16]. Крім того, методи

ансамблевого навчання, коли поєднуються прогнози декількох моделей, ще більше підвищили точність [1].

Критична оцінка траскторії розвитку галузі. Виявлення прогалин.

Однією з ключових проблем точності і масштабованості досліджень є виявлення й інтерпретація сарказму, який часто базується на тонких прагматичних сигналах, контексті та спільному культурному знанні, що не завжди враховується стандартними лексичними чи емоційними методами [39].

Іншим викликом є явище перемикування кодів – динамічного поєднання кількох мов чи діалектів у межах однієї розмови, яке дуже поширене в різноманітних онлайн-спільнотах. Оскільки більшість алгоритмів виявлення особистості тренуються на монологічних даних, перемикування кодів може призводити до помилкової класифікації або втрати важливих лінгвістичних сигналів.

Крім того, швидка еволюція онлайн-діалектів, таких як інтернет-жаргон, емоджі та специфічна лексика платформ, продовжує створювати труднощі для лексичних методів. Ці нові мовні форми швидко перевершують статичні словники чи навіть великі мовні моделі, які стають застарілими [40].

Додатковим аспектом є питання конфіденційності даних і етичних міркувань. Розвиток аналітичних методик підвищує ризик ненавмисного розкриття чи зловживання чутливими атрибутами користувачів. Тому дослідники все більше прагнуть поєднувати методологічні інновації із забезпеченням автономії учасників, поінформованої згоди та відповідального поводження з даними.

Висновки. Дослідження еволюції методів аналізу особистісних рис із текстових даних соціальних мереж у контексті когнітивної лінгвістики показало значний прогрес у цій галузі за останні два десятиліття.

Еволюція теоретичних рамок. Від лексичних і словникових підходів ранніх досліджень галузь перейшла до складних когнітивно-лінгвістичних моделей, які враховують контекстуальне використання мови, прагматику та концептуальні структури. Ця еволюція створила більш цілісне уявлення про те, як особистість відображається у мовленні, відкривши шлях до інтеграції додаткових моделей рис, таких як HEXACO чи Темна тріада.

Методологічний прогрес. Зміна від ручного кодування до автоматизованих підходів значно розширила можливості аналізу великих обсягів текстових даних. Впровадження глибинного навчання та сучасних трансформерних моделей підвищило точність прогнозування особистісних рис, дозволяючи враховувати складні контекстуальні залежності та інтегрувати мультимодальні сигнали, включаючи зображення та емоджі.

Розширення джерел даних. Еволюція джерел даних – від невеликих вибірок зручного доступу до масштабних багатомовних та мультимодальних корпусів, таких як Twitter, Reddit і TikTok – дозволила отримувати більш різноманітні й репрезента-

тивні дані. Однак зростання обсягів даних створює нові виклики для їхньої стандартизації та обробки.

Валідація та надійність. Більш складні підходи до валідації, такі як інтеграція поведінкових даних і лонгitudних дизайнів, значно підвищили валідність досліджень. Лонгitudні підходи дозволяють враховувати динамічні зміни мовлення та глибше розуміти, як особистість взаємодіє з контекстуальними факторами, включаючи стрес чи життєві переходи.

Інтеграція когнітивно-лінгвістичних підходів. Використання метафоричних і концептуальних аналізів та врахування контекстуальних факторів дало нові можливості для розуміння того, як мова відображає особистість. Ці підходи дозволяють виявляти тонкі когнітивно-лінгвістичні сигнали, які виходять за межі поверхневого аналізу тексту.

Ключові результати та виклики. Залишаються виклики, такі як інтерпретація сарказму, перемикання кодів і еволюція онлайн-діалектів. Окрім того, питання конфіденційності даних та етичні аспекти роботи з текстовими даними потребують подальшої уваги.

ЛІТЕРАТУРА:

1. Schwartz H.A., Eichstaedt J.C., Kern M.L. et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*. 2013. 8(9). e73791.
2. Kulkarni V., Kern M.L., Stillwell D. et al. Latent human traits in the language of social media: An open-vocabulary approach. *PLoS ONE*. 2018. 13(11).
3. Plank B., Hovy D. Personality traits on Twitter – or – how to get 1.500 personality tests in a week. In: *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 2015. P. 92–98.
4. Pennebaker J.W., Mehl M.R., Niederhoffer K.G. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*. 2003. 54(1). 547–577.
5. Groom C.J., Pennebaker J.W. Words. *Journal of Research in Personality*. 2002. 36(6): 615–621.
6. Tausczik Y.R., Pennebaker J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*. 2010. 29(1). 24–54.
7. Hart C. Moving beyond metaphor in the cognitive linguistic approach to CDA. *Critical Discourse Studies in Context and Cognition*. 2011. 43. 171–192.
8. Lakoff G., Johnson M. The metaphorical structure of the human conceptual system. *Cognitive Science*. 1980. 4(2). 195–208.
9. Hart C. Cognitive linguistic critical discourse studies: connecting language and image. In: *The Routledge Handbook of Language and Politics*. Routledge. 2017. p. 187–201.
10. Ashton M.C., Lee K. Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*. 2007. 11(2). 150–166.
11. Paulhus D.L., Williams K.M. The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*. 2002. 36(6). 556–563.
12. Mairesse F., Walker M.A., Mehl M.R., Moore R.K. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*. 2007. 30. 457–500.
13. Ruths D., Pfeffer J. Social media for large studies of behavior. *Science*. 2014. 346(6213). 1063–1064.
14. Majumder N., Poria S., Gelbukh A., Cambria E. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*. 2017. 32(2). 74–79.
15. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv. 1810.04805. 2018.
16. Mehta Y., Majumder N., Gelbukh A., Cambria E. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*. 2020. 53(4). 2313–2339.
17. Joshi A., Sharma V., Bhattacharyya P. Harnessing context incongruity for sarcasm detection. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 2015. P. 757–762.
18. Gill A., Nowson S., Oberlander J. What are they blogging about? Personality, topic and motivation in blogs. In: *Proceedings of the International AAAI Conference on Web and Social Media*. 2009. Vol. 3, No. 1. P. 18–25.
19. Vazire S., Carlson E.N. Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*. 2011. 20(2): 104–108.
20. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. 2013. 110(15): 5802–5805.
21. Back M.D., Stopfer J.M., Vazire S. et al. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*. 2010. 21(3): 372–374.
22. Chung C.K., Pennebaker J.W. Textual analysis. In: *Measurement in Social Psychology*. Routledge. 2018. P. 153–173.
23. Park J., Baek Y.M., Cha M. Cross-cultural comparison of nonverbal cues in emoticons on Twitter: Evidence from big data analysis. *Journal of Communication*. 2014. 64(2). 333–354.
24. Cohn M.A., Mehl M.R., Pennebaker J.W. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*. 2004. 15(10). 687–693.
25. Fauconnier G., Turner M. Conceptual blending, form and meaning. *Recherches en communication*. 2003. 19. 57–86.
26. Ritchie L.D. *Metaphor*. Cambridge University Press. 2013.
27. Fillmore C.J. Ideal readers and real readers. *Analyzing Discourse: Text and Talk*. 1982. P. 248–270.
28. Langacker R.W. *Foundations of Cognitive Grammar: Volume I: Theoretical Prerequisites*. Stanford University Press. 1987.
29. Qiu L., Lin H., Ramsay J., Yang F. You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*. 2012. 46(6). 710–718.

30. Marengo D., Giannotta F., Settanni M. Assessing personality using emoji: An exploratory study. *Personality and Individual Differences*. 2017. 112. 74–78.
31. Vasilescu B., Capiluppi A., Serebrenik A. Gender, representation and online participation: A quantitative study. *Interacting with Computers*. 2014. 26(5). 488–511.
32. Liu L., Preotiuc-Pietro D., Samani Z.R., Moghaddam M.E., Ungar L. Analyzing personality through social media profile picture choice. In: *Proceedings of the International AAAI Conference on Web and Social Media*. 2016. Vol. 10, No. 1. P. 211–220.
33. Kumar S., Hamilton W.L., Leskovec J., Jurafsky D. Community interaction and conflict on the web. In: *Proceedings of the 2018 World Wide Web Conference*. 2018. P. 933–943.
34. Wang N., Kosinski M., Stillwell D.J., Rust J. Can well-being be measured using Facebook status updates? Validation of Facebook's Gross National Happiness Index. *Social Indicators Research*. 2014. 115. 483–491.
35. El-Haj M., Kruschwitz U., Fox C. Creating language resources for under-resourced languages: methodologies, and experiments with Arabic. *Language Resources and Evaluation*. 2015. 49. 549–580.
36. Miranda C.H., Guzman J. A review of Sentiment Analysis in Spanish. *Tecciencia*. 2017. 12(22). 35–48.
37. Davidovitch N., Khyhniak K. Language Personality in the Conditions of Cross-Cultural Communication: Case-Study Experience. *International Education Studies*. 2018. 11(2). 13–26.
38. Park G., Schwartz H.A., Eichstaedt J.C. et al. Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*. 2015. 108(6). 934–952.
39. Ghosh A., Veale T. Fracking sarcasm using neural network. In: *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 2016. P. 161–169.
40. Hovy D., Spruit S.L. The social impact of natural language processing. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2016. P. 591–598.