

Ростислав Чайка

ВИЗНАЧЕННЯ РИС П'ЯТИФАКТОРНОЇ МОДЕЛІ ОСОБИСТОСТІ ЗА УКРАЇНСЬКОМОВНИМ ЕСЕ ПРОФЕСІЙНОЇ САМОПРЕЗЕНТАЦІЇ

Чайка Ростислав. *Визначення рис п'ятифакторної моделі особистості за українськомовним есе професійної самопрезентації.*

Вступ. У практиці управління персоналом, кар'єрного консультування та освіти самопрезентаційні тексти є поширеним джерелом даних про індивідуальні особливості людини. Водночас методологічно обґрунтовані підходи до визначення рис п'ятифакторної моделі особистості за українськомовними есе залишаються недостатньо дослідженими.

Мета. Оцінити можливість автоматичного передбачення рис п'ятифакторної моделі особистості за українськомовним есе самопрезентації та сформулювати відтворюваний початковий метод такого передбачення.

Методи. Дослідження проведено у січні – лютому 2026 року на базі одного з українських університетів. У пілотному аналізі використано 50 валідних відповідей і скорочену 15-пунктову методичку оцінювання рис п'ятифакторної моделі особистості (Soto & John, 2017). Текст есе аналізували як сукупність кількісно представлених мовних ознак, які далі узагальнювали до компактного семантичного подання та використовували для статистичного визначення вираженості рис особистості. Оцінювання стійкості отриманих результатів здійснювали за процедурою п'ятиблокової перехресної перевірки.

Результати дослідження. Найвищі показники відповідності між розрахованими за текстом і самооцінними значеннями рис отримано для «Відкритості до досвіду» ($r = 0.285$) та «Доброзичливості» ($r = 0.235$). Для «Екстраверсії» зв'язок був слабким ($r = 0.101$), тоді як для «Добросовісності» ($r = -0.248$) і «Негативної емоційності» ($r = -0.057$) він виявився низьким або нестійким.

Висновки. Отримані результати свідчать, що українськомовне есе професійної самопрезентації може містити інформацію, пов'язану з окремими рисами особистості, насамперед із «Відкритістю до досвіду» та частково з «Доброзичливістю». Дослідження формує основу для подальшої роботи, яка має передбачати залучення розширених текстових завдань, використання більш досконалих способів аналізу тексту та суворіші процедури перевірки стійкості результатів.

Ключові слова: п'ятифакторна модель особистості, когнітивна лінгвістика, аналіз тексту, прихований семантичний аналіз, регуляризована лінійна регресія, психометрія, прогнозування рис особистості

Chayka, Rostyslav. *Identifying Big Five Personality Traits from a Ukrainian-Language Professional Self-Presentation Essay.*

Introduction. In human resource management, career counseling, and education, self-presentation texts are a common source of data about an individual's characteristics. At the same time, methodologically grounded approaches to identifying Big Five personality traits from Ukrainian-language essays remain insufficiently studied.

Aim. To assess the possibility of automatically predicting Big Five personality traits from a Ukrainian-language self-presentation essay and to develop a reproducible initial method for such prediction.

Methods. The study was conducted in January–February 2026 at a Ukrainian university. The pilot analysis used 50 valid responses and a short 15-item instrument for assessing Big Five personality traits (Soto & John, 2017). The essay texts were analyzed as a set of quantitatively represented linguistic features, which were then generalized into a compact semantic representation and used for the statistical estimation of personality trait expression. The stability of the obtained results was assessed using a five-fold cross-validation procedure.

Results. The highest correspondence between text-based estimates and self-reported trait values was obtained for Openness to Experience ($r = 0.285$) and Agreeableness ($r = 0.235$). For Extraversion, the association was weak ($r = 0.101$), whereas for Conscientiousness ($r = -0.248$) and Negative Emotionality ($r = -0.057$) it was low or unstable.

Conclusions. *The findings indicate that a Ukrainian-language professional self-presentation essay may contain information related to certain personality traits, primarily Openness to Experience and, to a lesser extent, Agreeableness. The study provides a foundation for further research involving more extensive text tasks, more advanced text analysis methods, and stricter procedures for assessing the stability of results.*

Key words: *Big Five personality model, cognitive linguistics, text analysis, latent semantic analysis, ridge regression, psychometrics, personality trait prediction*

Вступ. У психологічних та організаційних дослідженнях дедалі частіше виникає потреба оцінювати індивідуальні відмінності не лише за опитувальниками, а й за вільними текстами – самопрезентаціями, мотиваційними листами, есе, рефлексивними описами досвіду. Для практики це напряду пов'язано із завданнями управління персоналом, кар'єрного консультування та освітньої діагностики: у багатьох реальних процесах саме текст є доступним джерелом інформації про людину, тоді як повні психодіагностичні опитувальники застосовуються рідше. Це формує практичне й наукове питання: чи можна оцінювати особистісні риси за текстовими даними – і якщо так, то якими методами, з якою точністю та за яких обмежень.

Попередні дослідження показують, що лексика та стилістичні особливості письма пов'язані з рисами особистості. На великих масивах текстів (зокрема, у блогах) виявлено системні зв'язки між вживанням певних слів та рисами особистості, причому інформативними можуть бути не лише заздалегідь визначені категорії слів, а й вільні мовні маркери, які виділяються без наперед заданих списків (Yarkoni, 2010). Масштабні дослідження з використанням соціальних мереж підтвердили стійкі зв'язки між мовою та рисами особистості й продемонстрували можливість автоматизованого оцінювання рис на основі мовного профілю (Schwartz et al., 2013; Park et al., 2015). Паралельно в психології індивідуальних відмінностей і психолінгвістичних дослідженнях утвердився підхід, за яким мовленнєві дані можуть використовуватися для визначення рис особистості; ранні праці показали принципову здійсненність такого прогнозування на основі тексту й діалогу (Mairesse & Walker, 2006). У сфері авторського профілювання поширені стандартизовані постановки та протоколи оцінювання, що підтримують відтворюваність і порівнюваність результатів між командами та доменами тексту (Rangel et al., 2016). П'ятифакторна модель особистості (Big Five), як правило, використовується як базова теоретична рамка для задач по визначенню особистісних рис з текстових даних (Goldberg, 1993).

Водночас у практиці польових досліджень часто застосовуються короткі форми вимірювання Big Five (15–30 пунктів), які зменшують навантаження на респондента, але потенційно знижують надійність окремих шкал і, відповідно, обмежують верхню межу точності будь-якого прогнозування, що використовує такі самооцінки як еталон (Soto & John, 2017). Окремим методологічним фактором є якість даних: неуважні або випадкові відповіді системно спотворюють психометричні оцінки та можуть істотно знижувати якість моделей; тому в сучасній методології обґрунтовується застосування спеціальних перевірок уважності і процедур виявлення неякісних відповідей (Meade & Craig, 2012; Oppenheimer et al., 2009).

Попри переконливі результати на великих масивах даних, недостатньо дослідженим залишається питання, наскільки відтворювано можна передбачати риси п'ятифакторної моделі за українськомовними короткими самопрезентаційними текстами (зокрема, навчальними есе), де обсяг тексту від кожного автора обмежений. Також відкритим залишається практичне питання вибору достатньо простого й водночас надійного способу аналізу тексту на невеликих вибірках. Один із підходів полягає в тому, щоб кількісно враховувати, які мовні елементи та з якою частотою використовує людина, а далі на цій основі статистично оцінювати вираженість рис особистості; інший – у тому, щоб не зосереджуватися на великій кількості окремих мовних показників, а узагальнювати їх до більш компактних змістових характеристик тексту. Саме таку функцію виконує прихований семантичний аналіз, який дає змогу перейти від великої кількості окремих мовних ознак до кількох узагальнених смислових компонентів, що відображають основний

зміст текстів (Deerwester et al., 1990). Однак застосування цього підходу до українськомовних есе з самопрезентацій у поєднанні з контролем валідності відповідей потребує емпіричної апробації.

Мета дослідження. Оцінити можливість автоматичного передбачення рис п'ятифакторної моделі особистості за українськомовним есе самопрезентації та сформулювати відтворюваний початковий метод такого передбачення для розширення та вдосконалення в подальших дослідженнях.

Завдання дослідження.

1. Побудувати статистичний спосіб визначення рис п'ятифакторної моделі особистості на основі мовних особливостей одного есе професійної самопрезентації.
2. Зіставити ефективність двох підходів до аналізу тексту: безпосереднього використання численних мовних показників і їх попереднього узагальнення до компактніших змістових характеристик.
3. З'ясувати, які риси особистості найкраще відображаються у текстах професійної самопрезентації, а також визначити обмеження й напрями подальшого розвитку цього підходу.

Методи та організація дослідження. Дослідження проводилося у листопаді – грудні 2025 року на базі одного з українських університетів, як частина більшого проєкту. Учасниками були студенти, які заповнювали онлайн-анкету в межах навчальної активності та отримували додаткові невеликі бали до поточної успішності за участь. Усі респонденти надавали інформовану згоду, а дані оброблялися в знеособленому вигляді; ідентифікаційні поля (контакти/ПІБ) не використовувалися в аналізі. Психодіагностичний блок включав скорочену «15-пунктову шкалу оцінювання рис п'ятифакторної моделі особистості» (Soto & John, 2017) в її українській адаптації (Кліманська & Галецька, 2019). Додатково учасники писали українською мовою есе на тему «Я – як майбутній фахівець», яке розглядалося як текстова самопрезентація. Для контролю якості відповідей застосовувалися перевірки уважності; відповіді, що не пройшли принаймні одну з перевірок, виключалися з подальшого аналізу, відповідно до рекомендацій щодо виявлення неуважних/випадкових відповідей (Meade & Craig, 2012; Oppenheimer et al., 2009). Всього було отримано 50 валідних записів.

Текст есе розглядали як джерело мовних показників, які можуть бути пов'язані з індивідуальними відмінностями особистості. На першому етапі для кожного тексту фіксували, які мовні елементи в ньому трапляються частіше, а які – рідше, причому більшої ваги надавали тим елементам, що є характерними саме для конкретного есе, а не однаково часто зустрічаються в усіх текстах вибірки (Spärck Jones, 1972). Такий підхід дає змогу перейти від суцільного тексту до системи кількісних мовних показників, придатних для подальшого статистичного аналізу. Для української мови цей спосіб є практичним, оскільки дозволяє враховувати особливості слововживання без складної попередньої мовної обробки тексту (Rangel et al., 2016).

Оскільки кількість таких мовних показників виявляється дуже великою, на наступному етапі їх узагальнювали за допомогою прихованого семантичного аналізу. Це дало змогу перейти від великої кількості окремих мовних ознак до меншої кількості узагальнених змістових характеристик тексту, які відображають основні смислові напрями, спільні для різних есе, і водночас зменшують вплив випадкових коливань та другорядних мовних відмінностей (Deerwester et al., 1990). Іншими словами, замість аналізу сотень і тисяч окремих мовних показників дослідження спиралося на компактніше подання загального змісту й стилю самопрезентації.

Далі ці узагальнені текстові показники використовували для статистичного визначення вираженості рис особистості. Для цього застосовували регуляризовану лінійну регресію, тобто такий спосіб аналізу, який дає змогу встановити, які саме мовні особливості тексту пов'язані з вищими або нижчими значеннями кожної риси, і водночас знижує ризик випадкових або нестійких результатів за великої кількості показників і відносно невеликої вибірки (Hoerl & Kennard, 1970). Для кожної з п'яти рис будували окреме статистичне

рівняння, але в межах єдиного аналітичного підходу, щоб результати можна було безпосередньо зіставляти між собою.

Щоб перевірити, наскільки отримані результати є стійкими і чи можуть вони відтворюватися на нових даних, застосовували п'ятиблокову перехресну перевірку. Для цього всю вибірку поділяли на п'ять частин: у кожному циклі аналізу чотири частини використовували для побудови статистичного способу визначення рис, а п'яту – для перевірки його роботи; далі процедуру повторювали доти, доки кожна частина один раз не виступала перевіркою (Hastie et al., 2009). Такий підхід дає змогу оцінити не лише те, наскільки добре спосіб аналізу «пояснює» вже наявні дані, а й те, наскільки він зберігає свою ефективність на підвибірках, яких раніше «не бачив».

Якість визначення рис оцінювали за двома показниками. Першим була кореляція Пірсона між розрахованими за текстом і самооцінними значеннями рис, яка показує, наскільки добре текстові показники відтворюють індивідуальні відмінності між учасниками (Pearson, 1895). Другим показником була середня абсолютна похибка, яка відображає, наскільки в середньому розраховане значення риси відхиляється від самооцінки в одиницях шкали (Willmott & Matsuura, 2005). У поєднанні ці два показники дають більш повне уявлення про результати: кореляція показує, чи правильно відтворюється відносне розташування учасників за вираженістю риси, а похибка – наскільки близькими є самі числові оцінки.

Вибірка дослідження. До пілотного аналізу було включено 58 відповідей респондентів, з яких лише 50 були визнані валідними (пройшли перевірку на уважність). Усі подальші аналізи проводилися лише для валідних відповідей відповідно до протоколу контролю якості даних через перевірки уважності. Вік у вибірці вказали 48 осіб (у 2 записах значення віку було відсутнє). Середній вік учасників становив 18.23 року (стандартне відхилення = 0.63), медіана – 18 років, діапазон – від 17 до 21 року. За статтю вибірка складалася переважно з жінок: 44 особи (88%) вказали жіночу стать і 6 осіб (12%) – чоловічу.

Результати дослідження та їх обговорення. Довжина текстів була достатньою для алгоритмічного аналізу: середня довжина становила 665.80 символів (SD = 188.67) та 102.28 слів (SD = 28.52). Мінімальна довжина есе – 227 символів / 37 слів, максимальна – 1136 символів / 188 слів (див. табл. 1). Отже, у вибірці відсутні надто короткі тексти, які могли б істотно спотворювати результати прогнозування.

Таблиця 1

Довжина есе «Я – як майбутній фахівець» (N=50)

Показник	Символи	Слова
Mean	665.80	102.28
SD	188.67	28.52
Min	227	37
25%	534.00	85.25
Median	654.00	103.50
75%	777.50	117.25
Max	1136	188

Описові статистики п'яти рис особистості наведено в табл. 2. Середні значення перебували в межах 3.29–3.67 бала, що відповідає помірним самооцінкам у більшості

учасників. Внутрішня узгодженість скорочених шкал (по 3 пункти на рису) була неоднорідною: найвищу узгодженість продемонструвала «Негативна емоційність» ($\alpha = 0.735$), помірні значення отримано для «Екстраверсії» ($\alpha = 0.555$), «Добросовісності» ($\alpha = 0.541$) та «Відкритості» ($\alpha = 0.509$), тоді як для «Добррозичливості» узгодженість була нижчою ($\alpha = 0.410$).

Таблиця 2

Опис рис п'ятифакторної моделі особистості за 15-пунктовою скороченою методикою та внутрішня узгодженість шкал (Cronbach's α), N=50

Риса	Mean	SD	Min	Max	α
Екстраверсія	3.34	0.74	1.67	5.00	0.555
Добррозичливість	3.29	0.71	1.00	4.67	0.410
Добросовісність	3.49	0.78	1.00	5.00	0.541
Нейротизм	3.42	0.87	1.67	5.00	0.735
Відкритість	3.67	0.78	2.00	5.00	0.509

Результати дослідження подано в табл. 3. Найвищу якість передбачення продемонстровано для «Відкритості» ($r = 0.285$; MAE = 0.612) та «Добррозичливості» ($r = 0.235$; MAE = 0.542). Для «Екстраверсії» зв'язок між прогнозом і самооцінкою був слабким, але позитивним ($r = 0.101$; MAE = 0.596). Для «Негативної емоційності» кореляція була близькою до нуля й дещо негативною ($r = -0.057$; MAE = 0.765), а для «Добросовісності» спостерігалось негативне значення ($r = -0.248$; MAE = 0.629).

Таблиця 3

**Якість визначення рис особистості за есе
(п'ятиблокова перехресна перевірка, N = 50)**

Основний метод: прихований семантичний аналіз тексту та регуляризована лінійна регресія

Риса	Pearson r	MAE
Відкритість	0.285	0.612
Добррозичливість	0.235	0.542
Екстраверсія	0.101	0.596
Нейротизм	-0.057	0.765
Добросовісність	-0.248	0.629

Водночас базова модель без стиску ознак продемонструвала вищий показник для «Відкритості» ($r = 0.328$; MAE = 0.602), однак нижчі значення для «Добррозичливості» ($r = 0.150$; MAE = 0.551) і гірший результат для «Негативної емоційності» ($r = -0.152$; MAE = 0.788). Для «Екстраверсії» показники були практично однаковими ($r \approx 0.10$), а для «Добросовісності» негативні значення залишалися близькими ($r = -0.231$). Загалом семантичний стиск ознак дав помітніший виграш у прогнозуванні «Добррозичливості» і частково пом'якшив негативний

результат для «Негативної емоційності», тоді як для «Відкритості» перевага залишилася за базовою моделлю (табл. 4).

Таблиця 4

**Якість визначення рис особистості за есе
(п'ятиблокова перехресна перевірка, N = 50)
Основний метод: аналіз зважених частот мовних елементів тексту без попереднього
змістового узагальнення та регуляризована лінійна регресія**

Риса	Pearson r	MAE
Відкритість	0.328	0.602
Доброзичливість	0.150	0.551
Екстраверсія	0.099	0.599
Добросовісність	-0.231	0.632
Нейротизм	-0.152	0.788

Отримані результати свідчать, що українськомовне есе самопрезентації справді може містити інформацію, пов'язану з окремими рисами п'ятифакторної моделі особистості, передусім із «Відкритістю до досвіду» та частково з «Доброзичливістю». Такий висновок узгоджується з попередніми дослідженнями, в яких було показано, що мовлення й письмо людини відображають її індивідуально-психологічні особливості, а мовні закономірності можуть бути пов'язані з рисами особистості (Yarkoni, 2010; Schwartz et al., 2013; Park et al., 2015). У нашому дослідженні саме «Відкритість до досвіду» виявилася рисою, яка найкраще відображалася в есе, що можна пояснити особливостями самого жанру професійної самопрезентації: учасники часто писали про власні інтереси, професійні прагнення, навчання, ідеї та саморозвиток, тобто про зміст, який психологічно близький до цієї риси.

Кращі результати для «Доброзичливості» при використанні узагальненого змістового аналізу тексту можуть означати, що ця риса проявляється не стільки в окремих словах, скільки в загальному способі побудови висловлювання, у характері самопрезентації та в смислових акцентах тексту. Інакше кажучи, для її визначення виявилися важливішими не поодинокі мовні елементи, а цілісніші змістові характеристики есе. Це узгоджується з логікою прихованого семантичного аналізу, який дозволяє перейти від великої кількості окремих мовних ознак до узагальнених змістових компонентів тексту (Deerwester et al., 1990).

Водночас дещо вищий показник для «Відкритості до досвіду» у базовому варіанті аналізу без попереднього узагальнення ознак можна пояснити тим, що ця риса, ймовірно, відображається у більш прямих мовних показниках. У попередніх дослідженнях також показано, що «Відкритість до досвіду» пов'язана з тематикою інтересів, культурних практик, навчання, пізнання нового та інтелектуальної допитливості (Yarkoni, 2010; Schwartz et al., 2013). Саме тому ця риса може краще проявлятися на рівні конкретніших мовних особливостей тексту.

Слабкі або нестійкі результати для «Екстраверсії», а також низькі показники для «Добросовісності» й «Негативної емоційності» потребують обережного тлумачення. Одне з можливих пояснень полягає в тому, що жанр професійної самопрезентації не обов'язково створює умови для достатнього прояву таких характеристик, як: організованість, емоційна напруженість або соціальна активність. Ці риси можуть виразніше відображатися в інших типах текстів, наприклад у розповідях про труднощі, конфлікти, стресові ситуації, невдачі чи необхідність дотримання вимог і строків.

Ще одним важливим чинником є використання дуже короткої шкали для оцінювання рис особистості. Хоча скорочені методики є зручними в умовах обмеженого часу опитування,

вони зазвичай поступаються повним версіям за надійністю та точністю вимірювання, що прямо зазначають автори таких форм (Soto & John, 2017). Унаслідок цього зростає ймовірність вимірювального шуму, а отже зменшується і можливий рівень відповідності між текстовими показниками та самооцінками рис.

Крім того, слід враховувати й невеликий обсяг вибірки. Хоча 50 валідних відповідей достатньо для початкової перевірки можливості такого підходу, цього недостатньо для стійких висновків щодо всіх рис особистості. У таких умовах особливого значення набуває контроль якості даних. Використання лише валідних відповідей, тобто тих, що пройшли перевірки уважності, є методично виправданим, адже неухважні або випадкові відповіді можуть одночасно знижувати якість психодіагностичних показників і послаблювати виявлений зв'язок між текстом та рисами особистості (Meade & Craig, 2012; Oppenheimer et al., 2009). Саме тому застосований підхід до контролю якості даних підвищує довіру до отриманих результатів навіть у межах такого початкового дослідження.

Висновки. У статті представлено початкове дослідження можливості визначення рис п'ятифакторної моделі особистості за українськомовним есе самопрезентації у студентській вибірці з контролем якості відповідей за допомогою перевірок уважності. Аналіз проведено на 50 валідних відповідах; для оцінювання рис використано скорочену 15-пунктову шкалу, яка дає змогу отримати узагальнені показники особистості за умов обмеженого часу обстеження. Текст есе було переведено у систему кількісних мовних показників, далі їх узагальнено до компактніших змістових характеристик тексту, після чого використано для статистичного визначення вираженості рис особистості. Стійкість отриманих результатів перевіряли за допомогою п'ятиблокової перехресної перевірки, що дає змогу оцінити, наскільки підхід зберігає свою ефективність на нових підвибірках.

Результати показали, що в жанрі професійної самопрезентації найкраще відображається «Відкритість до досвіду», а також частково «Доброзичливість». Для «Екстраверсії» виявлено лише слабкий зв'язок між текстовими показниками та самооцінкою, тоді як для «Добросовісності» і «Негативної емоційності» цей зв'язок у межах проведеного дослідження був низьким або нестійким. Порівняння двох способів аналізу тексту показало, що попереднє узагальнення мовних показників може покращувати визначення окремих рис, зокрема «Доброзичливість», і дещо зменшувати нестійкість результатів для емоційних показників. Водночас для «Відкритості до досвіду» кращий результат у цій вибірці дав підхід, що спирався на більш деталізовані мовні характеристики тексту.

Загалом дослідження підтверджує, що українськомовні тексти самопрезентації можуть бути корисним джерелом інформації для визначення окремих рис особистості та формують початкову основу для подальших досліджень у цьому напрямі. Водночас результати потрібно тлумачити обережно, з огляду на невеликий обсяг вибірки, використання лише одного типу тексту та застосування скороченої шкали рис, що може знижувати точність оцінювання.

Перспективи подальших досліджень. Подальші дослідження доцільно спрямувати як на розширення обсягу текстових даних, так і на вдосконалення способів перевірки стійкості отриманих результатів. Насамперед у межах уже зібраного масиву варто залучити всі доступні текстові завдання, а не лише одне есе, і порівняти, які типи письмових самозвітів – самопрезентація, опис ініціативи, довгострокової мети, успіху чи невдачі, а також ситуації вибору між ризиком і безпекою – є найбільш інформативними для визначення окремих рис особистості. Такий підхід узгоджується з попередніми дослідженнями, які показали, що зв'язок між мовними особливостями тексту та рисами особистості залежить від тематики тексту, його змісту та комунікативного контексту (Yarkoni, 2010; Schwartz et al., 2013). Окремо перспективним є об'єднання кількох текстів одного учасника в єдиний текстовий масив, щоб перевірити, чи підвищується точність визначення рис зі збільшенням обсягу мовного матеріалу про людину (Park et al., 2015).

По-друге, необхідно посилити процедури валідації: окрім перехресної перевірки, варто застосувати відкладену тестову підвибірку або вкладену схему налаштування параметрів, щоб зменшити ризик оптимістичної оцінки якості моделі на невеликих даних. Така практика відповідає загальній логіці порівнювальних постановок у суміжних задачах профілювання автора, де особлива увага приділяється відтворюваності та чесності оцінювання. Також

доцільно провести аналіз стійкості результатів щодо довжини тексту (наприклад, мінімальний поріг слів/символів) і перевірити, як ці умови впливають на якість прогнозування.

По-третє, важливо розширити набір методів подання тексту. У цьому пілоті використано компактні семантичні компоненти на основі прихованого семантичного аналізу, однак у наступній роботі варто перевірити сучасні семантичні подання, що будуються нейромережевими мовними моделями, і порівняти їх із класичними підходами за однакових протоколів валідації. Паралельно можна дослідити гібридні моделі, що поєднують семантичні подання з інтерпретованими показниками тексту (довжина, різноманітність словника, частка займенників тощо), щоб підвищити як точність, так і пояснюваність.

По-четверте, перспективним є удосконалення еталонного вимірювання рис. Оскільки короткі шкали можуть знижувати надійність і тим самим обмежувати верхню межу досяжної точності передбачення, надалі варто розглянути використання більш розгорнутих форм оцінки п'ятифакторної моделі або додаткових зовнішніх критеріїв для перевірки валідності прогнозів. Окремим напрямом може стати зіставлення результатів текстових моделей із поведінковими або навчальними показниками, якщо такі доступні, для оцінки практичної значущості прогнозів.

По-п'яте, необхідно продовжувати роботу з якістю даних: застосування перевірок уважності та виключення невалідних відповідей є важливим кроком, але в майбутніх дослідженнях доцільно порівняти різні стратегії виявлення неуважних відповідей і оцінити їх вплив на точність моделей та психометричні характеристики шкал.

Література

1. Кліманська, М., & Галецька, І. (2019). Українська адаптація короткого п'ятифакторного опитувальника особистості TIPI (TIPI-UKR). *Психологічний часопис*, 5(9), 57-74. <https://doi.org/10.31108/1.2019.5.9.4>
2. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407. [http://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3C391::AID-ASII%3E3.0.CO;2-9](http://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASII%3E3.0.CO;2-9)
3. Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26-34. <https://doi.org/10.1037/0003-066X.48.1.26>
4. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
5. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67. <https://doi.org/10.1080/00401706.1970.10488634>
6. Mairesse, F., & Walker, M. A. (2006, June). Automatic recognition of personality in conversation. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers* (pp. 85-88). <https://aclanthology.org/N06-2022/>
7. Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437-455. <https://doi.org/10.1037/a0028085>
8. Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867-872. <https://doi.org/10.1016/j.jesp.2009.03.009>
9. Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934-952. <https://doi.org/10.1037/pspp0000020>
10. Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *proceedings of the royal society of London*, 58(347-352), 240-242. <https://doi.org/10.1098/rspl.1895.0041>
11. Rangel, F., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., & Stein, B. (2016). Overview of the 4th author profiling task at PAN 2016: cross-genre evaluations. In *Working Notes Papers of the CLEF 2016 Evaluation Labs. CEUR Workshop Proceedings/Balog, Krisztian [edit.]; et al.* (pp. 750-784). <https://ceur-ws.org/Vol-1609/>
12. Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9), e73791. <https://doi.org/10.1371/journal.pone.0073791>
13. Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69-81. <https://doi.org/10.1016/j.jrp.2017.02.004>
14. Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 11-21. <https://doi.org/10.1108/eb026526>
15. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. <https://doi.org/10.3354/cr030079>
16. Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44(3), 363-373. <https://doi.org/10.1016/j.jrp.2010.04.001>

References

1. Klimanska, M., & Haletska, I. (2019). Ukrainianska adaptatsiia korotkoho p'iatyfaktorного opytuvalnyka osobystosti TIPI (TIPI-UKR) [Ukrainian adaptation of the short five-factor personality inventory TIPI (TIPI-UKR)]. *Psykhologichnyi chasopys*, 5(9), 57–74. [In Ukrainian] <https://doi.org/10.31108/1.2019.5.9.4>
2. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407. [http://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3C391::AID-ASI%3E3.0.CO;2-9](http://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASI%3E3.0.CO;2-9)
3. Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26–34. <https://doi.org/10.1037/0003-066X.48.1.26>
4. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1–758). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
5. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
6. Mairesse, F., & Walker, M. A. (2006, June). Automatic recognition of personality in conversation. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers* (pp. 85–88). <https://aclanthology.org/N06-2022/>
7. Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
8. Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867–872. <https://doi.org/10.1016/j.jesp.2009.03.009>
9. Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934–952. <https://doi.org/10.1037/pspp0000020>
10. Pearson, K. (1895). VII. *Note on regression and inheritance in the case of two parents*. *proceedings of the royal society of London*, 58(347-352), 240–242. <https://doi.org/10.1098/rspl.1895.0041>
11. Rangel, F., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., & Stein, B. (2016). Overview of the 4th author profiling task at PAN 2016: cross-genre evaluations. In *Working Notes Papers of the CLEF 2016 Evaluation Labs. CEUR Workshop Proceedings/Balog, Krisztian [edit.]; et al.* (pp. 750–784). <https://ceur-ws.org/Vol-1609/>
12. Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9), e73791. <https://doi.org/10.1371/journal.pone.0073791>
13. Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68, 69–81. <https://doi.org/10.1016/j.jrp.2017.02.004>
14. Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 11–21. <https://doi.org/10.1108/eb026526>
15. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. <https://doi.org/10.3354/cr030079>
16. Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44(3), 363–373. <https://doi.org/10.1016/j.jrp.2010.04.001>

Відомості про автора

Чайка Ростислав Михайлович, доктор філософії в галузі психології, доцент кафедри теоретичної та практичної психології Національного університету «Львівська Політехніка», м. Львів, Україна.

Chayka, Rostyslav, PhD, Associate Professor, Department of theoretical and applied psychology, Lviv National Polytechnic University, Lviv, Ukraine.

ORCID ID: <https://orcid.org/0000-0003-1979-0738>

Email: rchayka@gmail.com

Отримано 11 лютого 2026 р.
Рецензовано 19 березня 2026 р.
Прийнято 23 березня 2026 р.