ЕВОЛЮЦІЯ ПРОФАЙЛІНГУ: ІНТЕГРАЦІЯ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ДЛЯ КОМПЛЕКСНОГО ПРОГНОЗУВАННЯ РИЗИКУ АСОЦІАЛЬНОЇ ТА ДЕВІАНТНОЇ ПОВЕДІНКИ

THE EVOLUTION OF PROFILING: INTEGRATION OF LARGE LANGUAGE MODELS FOR COMPREHENSIVE PREDICTION OF THE RISK OF ANTI-SOCIAL AND DEVIANT BEHAVIOR

УДК 159.98
DOI https://doi.org/10.32782/2663-5208.2024.66.50

Шимко В.А.
д.психол.н.,
професор кафедри професійної психології
Національна академія Служби безпеки України

Стаття присвячена дослідженню застосування штучного інтелекту (ШІ) та Великих Мовних Моделей (ВММ) у профайлінгу та поведінковому аналізі. Метою дослідження є визначення можливостей інтеграції ВММ для підвищення точності прогнозування асоціальної та девіантної поведінки в цифровому просторі. Особлива увага приділяється можливостям ВММ у сфері автоматизованого аналізу лінгвістичних та нелінгвістичних даних, що надає правоохоронним органам інструменти для виявлення ризикових осіб та груп у соціальних мережах, форумах та інших цифрових платформах.

У роботі розглядаються такі основні дослідницькі запитання: 1) як використання ВММ може підвищити точність прогнозування на основі цифрових даних, та 2) які методи ВММ ефективні для ідентифікації ризикових осіб у цифровому середовищі. Автори зазначають, що ВММ, такі як GPT-3 і CLIP, завдяки своїм алгоритмам аналізу мовленнєвих та поведінкових патернів, значно перевершують традиційні методи профайлінгу в частині швидкості та точності. ВММ дозволяють автоматизувати процеси обробки великих обсягів цифрових даних і знаходити небезпечні патерни, які залишаються непоміченими за допомогою традиційних методів. Крім того, стаття аналізує перспективи інтеграції ВММ з іншими типами даних, такими як відео та аудіо, що може підвищити точність прогнозування. У роботі також розглядаються етичні аспекти застосування ВММ, зокрема питання конфіденційності, можливості дискримінації та ризики зловживання технологіями для моніторингу та контролю за соціальними групами. Основні висновки дослідження вказують на те, що ВММ забезпечують автоматизацію процесів аналізу, інтеграцію лінгвістичних і нелінгвістичних даних та можливість виявлення прихованих патернів девіантної поведінки. Також визначено кілька перспективних напрямів для подальших досліджень, включаючи розробку адаптованих моделей для різних культурних контекстів, вдосконалення інтеграції ВММ з іншими типами даних та етичне регулювання їх використання в правоохоронній діяльності. Таким чином, ВММ відкривають нові можливості для підвищення точності профайлінгу і прогнозування поведінкових ризиків у цифровому середовищі, однак потребують подальшого дослідження та адаптації для забезпечення їх етичного та ефективного застосування.

Ключові слова: *профайлінг, великі мовні моделі, ВММ, асоціальна поведінка, девіантна поведінка, штучний інтелект, прогнозування ризиків.*

The article explores the application of artificial intelligence (AI) and Large Language Models (LLMs) in profiling and behavioral analysis. The goal of the study is to determine the potential of integrating LLMs to enhance the accuracy of predicting antisocial and deviant behavior in the digital space. Special attention is given to the ability of LLMs to perform automated analysis of linguistic and non-linguistic data, providing law enforcement with tools for identifying at-risk individuals and groups in social networks, forums, and other digital platforms. The paper addresses the following key research questions: 1) how does the use of LLMs enhance the accuracy of predictions based on digital data, and 2) what LLM methods are effective for identifying at-risk individuals in the digital space. The authors highlight that LLMs, such as GPT-3 and CLIP, with their algorithms for analyzing speech and behavioral patterns, significantly outperform traditional profiling methods in terms of speed and accuracy. LLMs enable the automation of large-scale data processing, detecting dangerous patterns that traditional methods often miss. Additionally, the article analyzes the prospects of integrating LLMs with other data types, such as video and audio, which could further improve prediction accuracy. The paper also considers the ethical implications of using LLMs, particularly regarding privacy, the potential for discrimination, and the risk of technology misuse for monitoring and controlling social groups. The main conclusions of the study indicate that LLMs facilitate the automation of analysis processes, the integration of linguistic and non-linguistic data, and the detection of hidden patterns of deviant behavior. Several promising research directions are identified, including the development of models adapted to various cultural contexts, the improvement of LLM integration with other data types, and the ethical regulation of LLM use in law enforcement. Thus, LLMs offer new opportunities to enhance the accuracy of profiling and behavioral risk prediction in the digital space, although further research and adaptation are needed to ensure their ethical and effective application.

Key words: *profiling, large language models, LLM, antisocial behavior, deviant behavior, artificial intelligence, risk prediction.*

Вступ. Застосування штучного інтелекту (ШІ) та Великих Мовних Моделей (ВММ) у профайлінгу й поведінковому аналізі стає ключовим напрямом розвитку сучасних методів прогнозування асоціальної та девіантної поведінки. Ці інструменти забезпечують можливість обробки та аналізу великих масивів даних, що походять із цифрових джерел, таких як соціальні мережі, форуми, повідомлення та інші цифрові комунікації [26].

У той час як традиційні методи профайлінгу обмежуються аналізом видимих поведінкових

проявів, ВММ відкривають нові перспективи для аналізу прихованих моделей комунікації та прогнозування ризику девіантної поведінки [34]. Використання ШІ дозволяє автоматизувати процеси виявлення ризикових осіб і груп, що становить важливу задачу в умовах зростання цифрового злочинного контенту та посилення загроз громадській безпеці [11].

Метою цієї публікації є дослідження можливостей інтеграції ВММ для комплексного прогнозування ризику асоціальної та девіантної поведінки у правоохоронній практиці. У межах цієї мети виокремлюються такі **дослідницькі запитання**:

1. Яким чином використання ВММ може підвищити точність прогнозування асоціальної та девіантної поведінки на основі цифрових лінгвістичних даних?

2. Які конкретні методи аналізу ВММ можна застосувати для ідентифікації ризикових осіб та груп у цифровому просторі?

Наукова новизна цієї роботи полягає у всебічному аналізі новітніх технологій у сфері поведінкового аналізу, зокрема можливостей ВММ у прогнозуванні асоціальної та девіантної поведінки. Зокрема, робота досліджує, як ВММ можуть використовувати великі обсяги лінгвістичних даних для автоматизованого виявлення прихованих моделей поведінки, що можуть свідчити про високий ризик асоціальних дій [4]. Публікація демонструє, як інтеграція ВММ підвищує ефективність профайлінгу, забезпечуючи більш комплексний підхід до аналізу поведінки, що виходить за межі традиційних підходів [26].

Методологія дослідження базується, по-перше, на *аналізі наукових джерел* щодо традиційних методів профайлінгу та технологій на базі ВММ, зокрема «класичний профайлінг» [30], обмеження традиційних методів [6] та сучасні дослідження ВММ [26, 34]. По-друге, застосовано концептуальний аналіз, за допомогою якого систематизовано та порівняно можливості традиційних методів і підходів з ВММ щодо типу даних, методів обробки та точності прогнозування. По-третє, для інтеграції знань із кримінології, психології та лінгвістики використано *системний підхід*. По-четверте, здійснено *порівняльний аналіз*, завдяки якому оцінено ефективність ВММ на фоні традиційних методами профайлінгу, показано перевагу ВММ в аналізі цифрового середовища [16]. По-п'яте, реалізовано а*наліз етичних аспектів* щодо питання конфіденційності, упередженості моделей та можливих зловживань [5].

Виклад основного матеріалу. Профайлінг як науковий метод розвинувся на основі психологічних, соціологічних та кримінологічних теорій. Традиційно він використовувався для опису особистості злочинців на основі їхніх поведінкових патернів та злочинних дій. Одним із ранніх підходів є «класичний профайлінг», що зосереджується на методах аналізу демографічних даних і мотивів злочинців [30]. Проте ці методи мають обмеження: вони базуються на обмежених даних, що може призвести до стереотипів і помилок [6], та не враховують нові форми асоціальної поведінки, що виникають в умовах цифрових технологій [15].

ВММ, такі як GPT, використовують глибоке навчання для аналізу природної мови, базуючись на архітектурі трансформера, яка дозволяє моделі розуміти контекст і генерувати текст [31]. ВММ здатні обробляти великі обсяги даних і виконувати такі завдання, як аналіз емоцій, тональності та класифікація текстів [12], що є корисним для виявлення асоціальної поведінки у соціальних мережах [36].

Інтеграція ВММ у поведінковий аналіз дозволяє виявляти приховані патерни, які не можуть бути виявлені традиційними методами. Наприклад, вони допомагають ідентифікувати ознаки агресії через аналіз текстів [1]. Завдяки алгоритмам машинного навчання, ВММ здатні швидко адаптуватися до нових даних та підвищувати точність прогнозування ризиків девіантної поведінки [19], що є важливим для правоохоронної діяльності.

ВММ і обробка лінгвістичних даних. У сучасному цифровому середовищі лінгвістичні дані є важливим джерелом для аналізу поведінкових патернів та ідентифікації загроз. Соціальні мережі, форуми та електронні повідомлення можуть містити ознаки девіантної або асоціальної поведінки. ВММ, як GPT-3 і BERT, автоматизують обробку таких даних, виявляючи небезпечні сигнали, які можуть залишитися непоміченими при традиційних методах [8].

Дані збираються через API (Application Programming Interface) соціальних мереж, таких як Facebook або X, та парсинг відкритих платформ. Наприклад, автоматизовані скрипти збирають публікації та повідомлення для подальшого аналізу [17, 28]. Після цього дані проходять попередню обробку: фільтрацію, лематизацію, токенізацію та нормалізацію [20], що є критично важливим для точності аналізу [22].

ВММ здійснюють семантичний аналіз, виявляючи тривожні слова та вирази, такі як заклики до насильства чи прояви агресії. Це особливо корисно для швидкої обробки великих обсягів текстів та виявлення небезпечних тенденцій [16]. Наприклад, ВММ моніторять соціальні мережі в реальному часі, ідентифікуючи радикальні групи або токсичні коментарі, що дозволяє швидко реагувати на загрози [29].

ВММ і обробка нелінгвістичних даних. Окрім текстів, ВММ можуть обробляти й нелінгвістичні дані, такі як зображення, відео, аудіо,

біографічні дані та активність користувачів у мережі. Цифрові сліди у соціальних мережах можуть вказувати на поведінкові патерни, пов'язані з девіантною поведінкою. Для збору таких даних використовуються ті ж API, що й для текстів, з додатковою обробкою метаданих (геолокація, час публікацій, кількість взаємодій тощо), що дозволяє аналізувати поведінкові моделі у різних контекстах.

ВММ можуть досліджувати не лише текстову інформацію, але й взаємодії користувачів, їхню активність (частоту публікацій, уподобання) та біографічні дані, виявляючи ризики, наприклад, схильність до радикальних дій [27]. Прикладом є аналіз зображень за допомогою OpenAI's CLIP або DALL-E для виявлення радикальних символів чи фотографій з екстремістських заходів [25].

ВММ і алгоритми прогнозування поведінки. Інтеграція ВММ у профайлінг включає розробку алгоритмів для прогнозування асоціальної та девіантної поведінки на основі лінгвістичних даних. Вони дозволяють виявляти небезпечних осіб або групи через їхню активність у цифровому просторі. Для прогнозування використовуються моделі машинного навчання, які аналізують великі обсяги текстових даних, шукаючи мовні патерни, що свідчать про радикалізацію, а також враховують соціальні зв'язки та інші ознаки. Прикладом є алгоритми, що використовують GPT-3 для оцінки ризику радикалізації на основі мови ненависті чи екстремістських висловлювань у соціальних мережах [17–19].

Порівняльний аналіз ефективності ВММ та традиційних методів профайлінгу. Порівняння точності прогнозування між традиційними методами профайлінгу і використанням ВММ у сучасних цифрових середовищах демонструє значні переваги останніх за багатьма критеріями (Таблиця 1). Традиційні методи профайлінгу здебільшого базуються на психологічних теоріях, спостереженнях та інтерв'ю, що дозволяє фахівцям виявляти асоціальну або девіантну поведінку на основі прямих взаємодій з індивідуумами або ретроспективного аналізу їхніх дій. ВММ, у свою чергу, забезпечують можливість автоматизованого аналізу величезних обсягів цифрових даних, що дозволяє виявляти ризикові патерни в поведінці людей без необхідності прямого контакту, зокрема через обробку текстових і метаданих в онлайн-середовищі.

Основні переваги ВММ перед класичними методами профайлінгу:

1. *Автоматизація*. ВММ дозволяють автоматично обробляти значні обсяги даних, що забезпечує аналіз поведінкових патернів у реальному часі, тоді як класичні методи вимагають більше часу і ресурсів для збору та обробки інформації [16].

2. *Лінгвістичний аналіз*. Традиційні методи часто залежать від контекстуальних знань експерта і можуть бути упередженими через суб'єктивність. ВММ можуть аналізувати мову більш об'єктивно, відстежуючи тенденції та патерни, які можуть вказувати на девіантну поведінку, наприклад, радикалізацію або агресивні наміри [29].

3. *Нелінгвістичні дані*. Традиційні методи профайлінгу зазвичай фокусуються на безпосередніх поведінкових проявах, які іноді є важкими для вимірювання та узгодженого аналізу. ВММ забезпечують можливість обробляти не лише мовні прояви, але й нелінгвістичні дані, такі як фотографії, відео, метадані активності в соціальних мережах, які також можуть бути важливими індикаторами девіантної поведінки [25].

4. *Точність*. За даними досліджень, моделі на основі глибокого навчання, зокрема ВММ, продемонстрували підвищену точність у прогнозуванні девіантної поведінки порівняно з класичними методами профайлінгу, що базуються на психологічних тестах або ретроспективних даних [18, 19].

5. *Контекстуальна обробка*. ВММ здатні аналізувати широкий контекст, включаючи

Таблиця 1

Можливості ВММ та класичних методів профайлінгу

Критерій порівняння	Традиційні методи профайлінгу	ВММ у профайлінгу
Тип даних	Лінгвістичні (інтерв'ю, тести), спостереження	Лінгвістичні та нелінгвістичні (цифрові сліди)
Швидкість обробки	Порівняно низька, залежить від обсягу даних	Висока, автоматизована обробка великих обсягів даних
Точність	Відносно висока, але залежить від суб'єктивних факторів	Дуже висока, особливо в аналізі текстових даних і прогнозах
Контекстуальна обробка	Лімітована експертом	Глибокий семантичний і контекстуальний аналіз
Нелінгвістичні дані	Обмежено враховуються	Аналізується разом із лінгвістичними даними
Можливість виявлення нових патернів	Обмежена людськими знаннями	Виявлення нових та неочевидних патернів

тональність, семантичні асоціації та взаємодії між користувачами, що дозволяє виявляти небезпечні тенденції, які можуть залишатися прихованими для традиційних методів [22].

ВММ і процес ухвалення рішень. Впровадження ВММ у правоохоронну діяльність значно підвищує ефективність виявлення загроз і прогнозування ризиків у цифровому просторі [3]. ВММ швидко аналізують великі обсяги даних, дозволяючи своєчасно ідентифікувати загрози. На відміну від ручного аналізу, ВММ забезпечують вищу точність завдяки автоматизованому моніторингу цифрових платформ. Наприклад, GPT-3 виявляє агресивні патерни та радикальні висловлювання у форумах або закритих групах [29].

Ключова перевага ВММ – швидкість обробки даних, що дає змогу оперативно реагувати на загрози [16]. Вони також забезпечують раннє виявлення небезпек, дозволяючи проактивно запобігати кримінальним діям [17–19].

Виклики та етичні аспекти застосування ВММ. Однією з головних проблем використання ВММ у правоохоронній практиці є їхня обмежена здатність до інтерпретації контексту. Хоча ВММ досягають значних успіхів у роботі з великими обсягами даних, вони часто не враховують соціокультурні аспекти, що призводить до хибних інтерпретацій [5]. Моделі, що базуються на статистичних патернах, можуть помилково трактувати іронію чи сарказм як загрози, якщо контекст не враховується [7]. Це часто спричиняє неправильні прогнози, як це сталося у 2020 році, коли GPT-2, застосована правоохоронними органами Канади, видала понад 500 хибно позитивних результатів через нездатність правильно інтерпретувати невинні комунікації [35].

Крім того, ВММ можуть демонструвати упередженість через неякісне навчання або обмежені набори даних, що призводить до помилкових висновків про поведінку певних груп. Дослідження показують, що моделі, натреновані на даних західних соціальних мереж, не завжди коректно обробляють тексти з інших культурних контекстів, що також може призводити до хибних ідентифікацій [5]. Однією з основних проблем залишається складність розпізнавання іронії та сарказму, що часто призводить до помилок у правоохоронній практиці [10].

Впровадження ВММ у правоохоронну діяльність порушує важливі етичні питання, зокрема конфіденційності, дискримінації та можливостей зловживання технологією. ВММ здатні аналізувати великі масиви даних, включаючи особисті комунікації та соціальні мережі, що піднімає питання вторгнення в приватне життя [14]. Ризик зловживання технологіями високий, особливо у відсутності чітких правових регламентів.

Помилкові ідентифікації можуть виникати через упередженість або помилки моделі, коли невинні люди підозрюються у злочинних намірах через неправильне розуміння контексту чи сарказму [10]. ВММ, якщо не враховують соціокультурні особливості, можуть відтворювати стереотипи та сприяти дискримінації, спрямовуючи прогнози на певні етнічні або соціальні групи [23]. Крім того, ВММ можуть бути використані для масового спостереження або контролю за опозиційними групами, що загрожує демократичним принципам [32, 33].

ВММ у правоохоронній практиці. Застосування ВММ у правоохоронній практиці продемонструвало значні успіхи в ідентифікації асоціальної або девіантної поведінки у цифровому просторі. Одним із таких прикладів є використання ВММ для моніторингу соціальних мереж та онлайн-форумів з метою виявлення потенційних загроз терористичної діяльності. Так, у 2021 році правоохоронні органи Європейського Союзу впровадили спеціалізовану платформу, яка базується на GPT-3 для моніторингу та аналізу відкритих джерел даних, включаючи Twitter, Telegram, і закриті форуми. Система була розроблена для виявлення мовних патернів, що можуть свідчити про радикалізацію або підготовку до насильницьких акцій. За перші шість місяців роботи системи було зафіксовано понад 200 випадків виявлення загроз, що дозволило правоохоронним органам попередити кілька терористичних атак [21, 24, 26]. Це дозволило правоохоронним органам вчасно втрутитися та запобігти потенційним атакам.

Інший успішний кейс стосується використання ВММ у боротьбі з кіберзлочинністю. Правоохоронні органи США використовували ВММ від OpenAI для аналізу комунікацій на темних веб-ринках, де відбувався продаж наркотиків, зброї та інших заборонених товарів. Використовуючи алгоритми класифікації тексту, моделі ідентифікували незаконні транзакції та відстежували комунікації між злочинними угрупованнями. За два роки система допомогла заарештувати понад 150 осіб та зупинити кілька великих угод із продажу зброї [2].

Крім того, ВММ використовуються для виявлення педофільських мереж через аналіз комунікацій у чатах і соціальних мережах. За допомогою алгоритмів аналізу мовленнєвих патернів розпізнаються специфічні фрази та сленг, які використовуються в середовищі зловмисників, що сприяє ідентифікації злочинців і захисту потенційних жертв [8, 9].

З огляду на успіхи, досягнуті за допомогою ВММ, існує низка рекомендацій, що можуть покращити впровадження цих технологій у майбутніх дослідженнях та практичній роботі правоохоронних органів.

1. Розробка адаптованих мовних моделей для різних контекстів. Однією з ключових рекомендацій є необхідність створення спеціалізованих ВММ, що будуть адаптовані до конкретних культурних і мовних контекстів. Наприклад, моделі, які працюють із комунікаціями у західних країнах, можуть бути менш ефективними в аналізі даних у країнах з іншими культурними і мовними особливостями [24].

2. Інтеграція ВММ з іншими типами даних. Для підвищення точності прогнозів рекомендується інтегрувати ВММ з іншими типами даних, такими як аналіз поведінкових патернів у відео або аудіо. Це дозволить створити багатовимірний профіль підозрілих осіб і підвищити ефективність моніторингу у цифровому просторі [37].

3. Навчання персоналу для правильного використання ВММ. Правоохоронним органам важливо навчати спеціалістів, які будуть відповідальні за роботу з ВММ, щоб уникнути помилкових висновків та мінімізувати ризики помилкової ідентифікації загроз. Фахівці повинні розуміти як технічні, так і етичні аспекти використання таких систем [33].

4. Розробка етичних норм та правових регламентів. Для запобігання зловживанням технологіями ВММ необхідно розробити чіткі етичні норми та правові рамки, що будуть регулювати використання цих моделей. Це включає захист конфіденційності громадян та мінімізацію ризиків дискримінації [13].

Таким чином, ВММ належить важлива роль у сучасній правоохоронній практиці, однак їх ефективність залежить від правильного впровадження, налаштування і дотримання етичних стандартів.

Висновки. Інтеграція ВММ у профайлінг і поведінковий аналіз відкриває нові можливості для підвищення точності та ефективності прогнозування ризику асоціальної та девіантної поведінки. Основний здобуток цієї публікації полягає у комплексному розгляді можливостей ВММ для обробки лінгвістичних та нелінгвістичних даних у цифровому середовищі, а також у аналізі алгоритмів, які дозволяють виявляти приховані патерни девіантної поведінки на основі цих даних.

Досягнення мети дослідження полягає у конкретизації відповідей на дослідницькі запитання:

Як використання ВММ підвищує точність прогнозування на основі цифрових даних? Аналіз показав, що ВММ значно перевершують традиційні методи профайлінгу завдяки автоматизації процесів збору й аналізу даних, що дозволяє оперативно і точно прогнозувати ризики на основі великих обсягів цифрової інформації. ВММ демонструють високу точність у виявленні небезпечних патернів, як у лінгвістичних даних, так і в нелінгвістичних (фото, відео, поведінкові метадані).

Які методи ВММ можна застосувати для ідентифікації ризикових осіб та груп у цифровому просторі? ВММ, такі як GPT-3 та CLIP, продемонстрували здатність ефективно аналізувати мовні патерни, що можуть свідчити про радикалізацію або асоціальну поведінку. Алгоритми, які використовуються для автоматизованого аналізу цифрових слідів (текст, метадані активності, зображення), забезпечують правоохоронним органам можливість раннього виявлення загроз у цифровому середовищі, особливо у випадках тероризму та кіберзлочинності.

Основні висновки проведеного аналізу пов'язані з такими очевидними перевагами ВММ порівняно з традиційними підходами:

1. *Автоматизація та масштабованість* – ВММ забезпечують можливість аналізувати величезні обсяги даних у реальному часі, що значно підвищує швидкість і точність ухвалення рішень у правоохоронній практиці.

2. *Інтеграція лінгвістичних і нелінгвістичних даних* – використання ВММ дозволяє комплексно аналізувати як текстову інформацію, так і метадані, зображення та інші цифрові сліди, що розширює можливості традиційного профайлінгу.

3. *Виявлення прихованих патернів* – алгоритми на основі ВММ дозволяють виявляти складні мовні та поведінкові патерни, які можуть залишатися непоміченими під час застосування традиційних методів профайлінгу.

Перспективні напрями досліджень:

1. Розробка адаптованих моделей для різних культурних контекстів. ВММ, розроблені для конкретних мовних та культурних середовищ, могли б підвищити точність прогнозування у глобальному контексті.

2. Вдосконалення інтеграції ВММ з іншими типами даних. Важливо досліджувати подальші можливості інтеграції ВММ із даними, отриманими з відео, аудіо та інших цифрових джерел.

3. Етичні аспекти та правові рамки. Необхідно продовжити дослідження у сфері етичного регулювання використання ВММ у правоохоронній діяльності для запобігання зловживанням технологіями.

Таким чином, ВММ відіграють ключову роль у сучасному профайлінгу, проте для їх ефективного використання необхідно враховувати як технічні, так і етичні аспекти, що дозволить максимально використовувати потенціал цих моделей у майбутньому.

ЛІТЕРАТУРА:

1. Alm CO, Sproat R. Emotions from text: A pilot study of the emote project. In: Proceedings of the 5th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2004); 2005.

2. Bagdasaryan E, Poursaeed O, Shmatikov V. Cloak: Protecting Confidentiality against Adversarial Inference through Representation Cloaking. In: Proceedings of the 2022 IEEE Symposium on Security and Privacy; 2022. https://doi.org/10.1109/SP.2022.00123

3. Baker J, McKenzie R. The Role of AI in Criminal Profiling: Insights from Law Enforcement. J Crim Justice Res. 2021;45(3):234-50.

4. Bavelas JB, Chovil N. Nonverbal and verbal communication: Hand gestures and facial displays as part of language use in face-to-face dialogue. Hum Commun Res. 2006;28(3):312-48. https://doi.org/10.1111/j.1468-2958.2002.tb00809.x

5. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021:610-23. https://doi.org/10.1145/3442188.3445922

6. Bennell C, Jones N. The role of profiling in the investigation of serious crime: A review of the literature. J Investig Psychol Offender Profiling. 2005;2(2):113–29.

7. Bommasani R, Hudson D, Adeli E, Altman R, Arora S, von Arx S, et al. On the Opportunities and Risks of Foundation Models. 2021; arXiv preprint arXiv:2108.07258. https://arxiv.org/abs/2108.07258

8. Chen L, Wang Y, Zhao H. Behavioral Analysis Using Large Language Models in Digital Communication. J Data Sci Anal. 2022;18(4):45-60.

9. Chen Z, Zhang H, Jiang B, Cao C. Detecting Hidden Cyber-Pedophilia Networks through Language Patterns in Digital Communications. ACM Trans Internet Technol. 2021;21(3):15-24. https://doi.org/10.1145/3471234

10. Crawford K. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven: Yale University Press; 2021.

11. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. Proceedings of the 5th annual ACM web science conference. 2013:47-56. https://doi.org/10.1145/2464464.2464471

12. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. 2018; arXiv preprint arXiv:1810.04805.

13. Floridi L, Cowls J. A Unified Framework of Five Principles for AI in Society. Harv Data Sci Rev. 2019;1(1). https://doi.org/10.1162/99608f92.8cd550d1

14. Floridi L, Taddeo M. What is data ethics? Philos Trans A Math Phys Eng Sci. 2016;374(2083):20160360. https://doi.org/10.1098/rsta.2016.0360

15. García M, Rojas J. Assessing the effectiveness of psychological profiling in criminal investigations. Crim Justice Stud. 2018;31(1):79–96.

16. Huang C, Huang Y, Zhu Z. Exploring deep learning methods for sentiment analysis in social media. J Adv Comput Intell Intell Inform. 2021;25(4):673-80.

17. Kumar A, Dutta A. Web Scraping and Data Collection: Techniques for Gathering Information from Social Media. J Comput Sci Inf Technol. 2020;12(2):100-12.

18. Kumar S, Hamilton WL, Leskovec J, Jurafsky D. Community Interaction and Conflict in Online Social Networks. Proceedings of the 13th International Conference on Web Search and Data Mining. 2020:317-25.

19. Kumar A, Singh S. Applications of NLP in criminal profiling: A review. J Appl Res Intellect Disabil. 2020;33(4):900–14.

20. López F, Martinez G, Rodríguez A. Text Preprocessing in Machine Learning: A Comprehensive Review. J Mach Learn Res. 2022;23(5):211-40.

21. Mackenzie R. Identifying Radicalization Online: A Study of Language Patterns in Social Media. J Terror Res. 2022;13(2):45-58. https://doi.org/10.15664/jtr.1586

22. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge: Cambridge University Press; 2014.

23. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. ACM Comput Surv. 2021;54(6):1-35. https://doi.org/10.1145/3457607

24. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The Ethics of Algorithms: Mapping the Debate. Big Data Soc. 2019;3(2):2053951716679679. https://doi.org/10.1177/2053951716679679

25. Radford A, Kim JW, Hallacy C, Ramesh A. Learning Transferable Visual Models from Natural Language Supervision. Proceedings of the 38th International Conference on Machine Learning. 2021:8748-56.

26. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI Blog. 2019;1(8):9-12. Retrieved from https://openai.com/research/language-models

27. Sharma R, Gupta S, Roy S. Detecting Threats in Social Media: A Case Study on Data Mining Techniques. Int J Inf Secur. 2021;20(1):99-115.

28. Shymko V. Scripts for automated data collection from Facebook, Instagram and Twitter. Zenodo. 2024. https://doi.org/10.5281/zenodo.13981396

29. Skeppstedt M, Eklund A, Gustafsson J. Detection of violent extremist rhetoric in social media with language models. Comput Hum Behav. 2022;130:107211.

30. Turvey BE. Criminal profiling: An introduction to behavioral analysis. Academic Press; 2011.

31. Vaswani A, Shardlow M, Wolf T. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017); 2017.

32. Vincent J. Ethical Considerations in AI Deployment in Law Enforcement. AI Soc. 2021;37(4):583-95. https://doi.org/10.1007/s00146-021-01269-0

33. Vincent J. The Metaverse has a groping problem already. The Verge. 2021. https://www.theverge.com/2021

34. Vitale JE, Gervasio MT, Mastrogiovanni F. Augmented behavior understanding through AI-based profiling tools: Risks and potentials. IEEE Trans Affect

Comput. 2021;12(3):528-40. https://doi.org/10.1109/TAFFC.2019.2959444

35. Weidinger L, Mellor J, Rauh M, Griffin C, Uesato J, Huang P, et al. Ethical and social risks of harm from language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022:547-58. https://doi.org/10.1145/3531146.3533088

36. Zhao Y, Li Q, Yan Y. Natural language processing in criminal investigations: A review of recent advancements. Int J Law Inf Technol. 2021;29(1): 1–24.

37. Zhong C, Zheng P, Li D. Multi-modal Integration of Large Language Models for Risk Assessment. IEEE Trans Inf Forensics Secur. 2022;17(6):1289-302. https://doi.org/10.1109/TIFS.2022.3140096