

Макаренко Лілія¹

¹Київський інститут сучасної психології та психотерапії

ПСИХОЛОГІЧНІ МЕХАНІЗМИ ВПЛИВУ СИНТЕТИЧНОГО КОНТЕНТУ НА ПОЛІТИЧНУ ДОВІРУ ТА ЕЛЕКТОРАЛЬНИЙ ВИБІР

Анотація

Проаналізовано психологічні механізми впливу синтетичного контенту (діпфейків, AI-генерованих текстів) на політичну довіру та електоральну поведінку. Встановлено, що синтетичний контент апелює до інтуїтивної обробки інформації (Система 1), обходячи аналітичний контроль. Досліджено роль ефекту ілюзорної правди, підтверджувального упередження та феномену «дивідендів брехуна» у формуванні епістемічного нігілізму. Показано, що страх і гнів є ключовими медіаторами поширюваності маніпулятивних матеріалів. Доведено ефективність методу превентивного психологічного щеплення (prebunking) як стратегії когнітивного захисту. Сформульовано рекомендації щодо програм медіаграмотності та виборчого просвітництва.

Ключові слова: синтетичний контент; діпфейки; генеративний штучний інтелект; політична довіра; електоральна поведінка; теорія подвійного процесу; когнітивні упередження; дезінформація; психологічне щеплення; ілюзорна правда.

PSYCHOLOGICAL MECHANISMS OF THE INFLUENCE OF SYNTHETIC CONTENT ON POLITICAL TRUST AND ELECTORAL CHOICE

Abstract

The psychological mechanisms of synthetic content (deepfakes, AI-generated texts) influence on political trust and electoral behavior are analyzed. It is established that synthetic content appeals to intuitive information processing (System 1), bypassing analytical control. The roles of the illusory truth effect, confirmation bias, and the Liar's Dividend phenomenon in fostering epistemic nihilism are examined. Fear and anger are shown to be key mediators of manipulative content virality. The effectiveness of prebunking as a cognitive protection strategy is

demonstrated. Recommendations for media literacy and electoral education programs are formulated.

Keywords: synthetic content; deepfakes; generative artificial intelligence; political trust; electoral behavior; dual process theory; cognitive biases; disinformation; prebunking; illusory truth.

Вступ

Цифровізація публічного простору докорінно змінила умови формування громадської думки та електоральної поведінки. Якщо донедавна основним об'єктом вивчення залишалась традиційна пропаганда – від агітаційних плакатів до телевізійних наративів, – то нині предметом наукового осмислення стає принципово новий феномен: синтетичний контент, вироблений або суттєво модифікований за допомогою генеративних алгоритмів.

Синтетичний контент охоплює широкий спектр медіаматеріалів: аудіо- та відеодіпфейки, автоматично згенеровані тексти, штучно синтезовані зображення, мережі акаунтів, якими керують боти. Поряд із технічно складними deepfakes окремої уваги заслуговують так звані «дешеві підробки» (cheap fakes) – маловартісні маніпуляції зі швидкістю відтворення, монтажем або контекстом, для яких не потрібні передові алгоритми (Paris & Donovan, 2019). Варто зауважити, що дослідники розрізняють три рівні інформаційного розладу: дезінформацію (disinformation), міс-інформацію (misinformation) та шкідливу інформацію (malinformation), кожен з яких має різні психологічні механізми впливу (Wardle & Derakhshan, 2017). На відміну від традиційної дезінформації, синтетичний контент – особливо у своїх найдосконаліших формах – дедалі складніше відрізнити від автентичного без спеціалізованих засобів верифікації, а міра цієї складності зростає відповідно до вдосконалення технологій.

Проблема набуває особливої гостроти в контексті демократичних виборів. Дослідження останніх років фіксують зростання кількості виборців, які повідомляють про труднощі з верифікацією політичних повідомлень, зниження рівня інституційної довіри та поширення «епістемічної втоми» – стану, коли людина припиняє активно розрізняти правдиве й хибне через надмірне когнітивне навантаження. Це формує сприятливе середовище для маніпулятивного впливу на електоральний вибір.

Мета дослідження

Мета статті – систематизувати психологічні механізми, за допомогою яких синтетичний контент впливає на політичну довіру та електоральну поведінку, а також окреслити ефективні стратегії когнітивного захисту.

Завдання дослідження

Досягнення поставленої мети передбачає вирішення таких завдань: (1) визначити теоретичний базис дослідження впливу дезінформації; (2) описати ключові когнітивні та емоційні механізми сприйняття підробок; (3) проаналізувати задокументовані випадки використання синтетичного контенту в реальних політичних процесах; (4) дослідити трансформацію електоральної

поведінки в умовах інформаційної невизначеності; (5) обґрунтувати методи протидії маніпуляціям.

Методи та організація дослідження

Дослідження ґрунтується на методах теоретичного аналізу та синтезу: систематизації наукових джерел, порівняльного аналізу теоретичних концепцій і узагальнення результатів емпіричних досліджень у галузях когнітивної психології, політичної комунікації та медіапсихології. Метод системного аналізу застосовано для дослідження задокументованих кейсів використання синтетичного контенту у виборчих кампаніях та збройних конфліктах.

Зразок дослідження

Теоретичну базу дослідження становлять 27 наукових публікацій та офіційних аналітичних документів, відібраних методом цілеспрямованої вибірки з баз даних Scopus, Web of Science та Google Scholar за 1977–2024 роки.

Результати дослідження та їх обговорення

Теоретичний базис дослідження

Теорія подвійного процесу (Dual Process Theory), розроблена у класичній праці Д. Канемана (Kahneman, 2011: 19–30) та поглиблена у дослідженнях Дж. Еванса й К. Становича (Evans, 2008: 255–258; Stanovich & West, 2000: 645–648), виокремлює два режими мислення. Система 1 функціонує автоматично, швидко та асоціативно, спираючись на емоційні реакції. Система 2 – повільна, аналітична, вимагає свідомих зусиль і метакогнітивного контролю. Синтетичний контент, передусім відеодіпфейки, конструюється із розрахунку на активацію Системи 1. Яскраві образи, емоційно насичені сюжети та контекст, що підтверджує наявні переконання, забезпечують майже миттєве сприйняття матеріалу як автентичного до того, як Система 2 встигає здійснити критичний аналіз.

Ефект ілюзорної правди (illusory truth effect), уперше описаний у класичному експерименті Л. Хашер, Д. Голдштайн, Т. Топпіно (Hasher et al., 1977: 107–110) і підтверджений у низці подальших досліджень, зокрема Л.Фаціо та ін. (Fazio et al., 2015: 993–997), полягає в тому, що регулярне повторення твердження – навіть свідомо сприйнятого як хибне при першому контакті – підвищує суб'єктивну оцінку його правдивості. Примітно, що цей ефект не зникає навіть тоді, коли учасники мають достатні знання, що суперечать хибному твердженню: побіжна обробка інформації при повторному контакті активує відчуття знайомості, яке система інтуїтивної обробки хибно інтерпретує як ознаку достовірності. У цифровому середовищі, де одне й те саме повідомлення може з'являтися в стрічці користувача десятки разів, цей ефект набуває системного характеру. С. Левандовські також вважає, що повторення неправдивої інформації у соціальних мережах веде до її поступового прийняття навіть серед аудиторії, початково налаштованої скептично (Lewandowsky et al., 2017: 358–362).

Підтверджувальне упередження (confirmation bias) функціонує як додатковий фільтр: люди схильні приймати інформацію, що відповідає їхній наявній політичній ідентичності й цінностям, і відкидати або критично

переосмислювати дані, що суперечать їх переконанням. Синтетичний контент, ретельно адаптований до певного ідеологічного профілю аудиторії, наштовхується на значно меншу когнітивну резистентність. Персоналізовані алгоритмічні середовища соціальних мереж посилюють цю динаміку, адже «інформаційні бульбашки» послаблюють контакт із дисонуючою інформацією, знижуючи поріг критичності.

Важливим теоретичним конструктом є концепція епістемічної довіри (epistemic trust), розроблена П. Фонагі та П. Алленом у контексті клінічної психології та теорії прив'язаності (Fonagy & Allison, 2014: 372–374). Попри різницю контекстів, перенесення цієї концепції у площину політичної комунікації є обґрунтованим: і в терапевтичному, і в медіасередовищі ключову роль відіграють ті самі психологічні сигнали, що визначають, чи надасть індивід кредит довіри конкретному джерелу – авторитетність, паравербальні ознаки щирості та відповідність культурним очікуванням аудиторії. Синтетичний контент, що технічно відтворює ці сигнали, здатен отримувати незаслужену довіру і проникати у систему переконань людини на правах автентичної інформації.

Психологічні феномени сприйняття синтетичних підробок

Феномен «дивідендів брехуна» (Liar's Dividend), введений у науковий обіг Р. Чесні та Д. Ситрон, описує парадоксальний наслідок поширення технологій синтетичного контенту (Chesney & Citron, 2019: 1758–1762): навіть коли певний матеріал є достовірним, сама поінформованість про існування підробок дає зацікавленим акторам підставу оголошувати компрометуючі докази фальсифікацією. Таким чином, технології підробки підривають довіру не лише до синтетичних матеріалів, а й до автентичних. Це продукує стан, який можна визначити як «епістемічний нігілізм»: систематичну відмову від можливості верифікувати будь-яку інформацію, що у підсумку зменшує здатність аудиторії до обґрунтованих суджень.

Дослідження К. Ваккарі та Е. Чедвіка, проведене в умовах реальної виборчої кампанії, встановило, що більше 40% респондентів стикалися з відеоматеріалами, які, на їх думку, могли бути підробленими, і що такий досвід суттєво знижував їхню готовність довіряти будь-яким відеодоказам у майбутньому (Vaccari & Chadwick, 2020: 6–9). Це свідчить про те, що ефект «дивідендів брехуна» (Liar's Dividend) функціонує не лише як тактичний прийом окремих акторів, а й як системний ефект технологізованого середовища в цілому.

Критичною у сприйнятті синтетичного контенту є роль емоцій. Гнів і страх виступають потужними чинниками поширюваності маніпулятивних матеріалів. В. Дж. Бреді на основі вивчення матеріалів Twitter виявив, що кожне додаткове слово морально-емоційного забарвлення у повідомленні підвищує ймовірність його ретвіту приблизно на 20%, причому цей ефект є стабільним у різних ідеологічних спільнотах (Brady et al., 2017: 7314–7315). Страх активує захисні системи, перемикаючи когнітивний контроль з аналітичного режиму на реактивний, – саме тому контент, що апелює до загрози (безпека родини,

зовнішня небезпека, крах ідентичності), обробляється менш критично. Гнів, своєю чергою, знижує критичність при відборі інформації та посилює групову поляризацію. Це спонукає індивіда до миттєвої ідентифікації за принципом «ми–вони», що фактично блокує раціональну перевірку фактів ще на етапі емоційного сприйняття.

Окремої уваги заслуговує явище «глибокого фреймінгу» (deep framing), описане Дж. Лакоффом (Lakoff, 2004: 3–18). Синтетичний контент не просто повідомляє окремий факт, а активує глибокі ціннісні фрейми, вкорінені у культурній та груповій ідентичності реципієнта. Один правдоподібний діпфейк, що зображує кандидата у компрометуючій ситуації, здатен запустити цілий каскад асоціацій, пов'язаних з образом «ворога» або «зрадника», і тим самим вплинути на електоральну позицію значно сильніше, ніж будь-яка аргументована критика.

Синтетичний контент у реальних політичних процесах: задокументовані випадки

Теоретичні конструкти, описані вище, знаходять підтвердження у низці задокументованих випадків використання синтетичного контенту в реальних політичних процесах. Узагальнений аналіз цих прецедентів дозволяє простежити, як психологічні механізми маніпуляції реалізуються в конкретних соціально-політичних умовах.

Одним із перших резонансних прецедентів стала президентська криза у Габоні в січні 2019 року. Після кількох місяців публічної відсутності президента Алі Бонго, який переніс інсульт, державне телебачення поширило відеозвернення, що мало засвідчити його дієздатність. Опозиційні сили та низка міжнародних спостерігачів оголосили відео діпфейком, що спровокувало спробу військового перевороту. Цей випадок є показовим прикладом дії ефекту «дивідендів брехуна» (Liar's Dividend). Незалежно від того, чи був відеозапис підробленим, сам факт його сприйняття як можливої маніпуляції дестабілізував політичну систему (Kietzmann et al., 2021: 558–559).

Суттєвим кроком в осмисленні масштабів проблеми стали дослідження президентських виборів у США 2016 року. Х. Олкотт і М. Гентцков задокументували, що в передвиборчий період поширювалось щонайменше 115 проурядових і 41 пропрезидентська фейкові статті, що отримали в сукупності близько 38 мільйонів репостів у Facebook (Allcott & Gentzkow, 2017: 212–213). Структурний аналіз, проведений Й. Бенклером, Р. Фарісом та Г. Робертсом, показав, що медіасередовище на правому фланзі американської політики виявилось значно більш схильним до дезінформаційної динаміки порівняно з іншими сегментами (Benkler et al., 2018: 12–15). П. Говард, С. Вулі та Р. Кало встановили, що автоматизовані акаунти-боти відповідали за 19% усього політичного трафіку у Twitter в останні тижні виборчої кампанії, формуючи хибне враження масової соціальної підтримки (Howard et al., 2018: 84–86).

Якісно новий вимір проблема набула у 2022 році з початком повномасштабного вторгнення Росії в Україну. У березні 2022 року в українських

соціальних мережах та на зламаних телевізійних каналах поширилось відео із зображенням президента В. Зеленського, який нібито закликає українських солдатів скласти зброю. Відео було швидко ідентифіковане як дідфейк низької якості та спростоване, проте сам факт його появи засвідчив готовність ворожих акторів використовувати синтетичний контент для психологічного впливу на особовий склад збройних сил і цивільне населення в умовах збройного конфлікту (Goldstein et al., 2023: 12–14). Цей випадок підтверджує ефект ілюзорної правди: навіть спростований дідфейк може залишати емоційний осад і підвищувати рівень тривожності, знижуючи здатність аудиторії до критичного сприйняття наступних повідомлень.

Особливу загрозу для електоральних процесів продемонстрував аудіальний дідфейк, розповсюджений у Словаччині у вересні 2023 року – за два дні до парламентських виборів. У соціальних мережах з'явився звуковий запис, на якому лідер ліберальної партії «Прогресивна Словаччина» М. Шімечка нібито обговорює схему фальсифікації виборів шляхом підкупу голосів представників ромської меншини. Обидва учасники діалогу, що ніби відбувся, спростували його автентичність, а незалежний аналіз ідентифікував запис як синтезований. Проте часовий розрив між поширенням запису і спростуванням виявився психологічно критичним, адже матеріал вийшов у суботу, коли більшість верифікаційних медіаорганізацій не працювали, а рішення виборців формувалося в умовах інформаційної невизначеності (Mehrotra, 2023). Аналіз вказує на те, що цілеспрямоване використання «часового вікна» стає самостійним тактичним інструментом дезінформаційних операцій.

На початку 2024 року у Сполучених Штатах зафіксовано прецедент масової операції із застосуванням синтезованого голосу: незадовго до праймеріз Демократичної партії в Нью-Гемпширі тисячі виборців отримали рободзвінки з голосом, синтезованим на основі публічних виступів президента Дж. Байдена, який закликав не брати участі у голосуванні. Федеральна комісія із зв'язку США кваліфікувала застосування AI-синтезованих голосів у рободзвінках як незаконне і зобов'язала операторів вживати заходів для їх блокування (Federal Communications Commission, 2024). Цей випадок засвідчує, що генеративний штучний інтелект суттєво знижує технічний і фінансовий поріг входу для організації дезінформаційних кампаній. Якщо раніше для виробництва переконливого синтетичного контенту були потрібні значні ресурси, сьогодні подібні операції доступні широкому колу акторів.

Окремим напрямом досліджень є вивчення ефективності мікротаргетованих дідфейків – матеріалів, адаптованих до конкретних психографічних профілів виборців. Т. Добер та ін. в контрольованому експерименті показали, що персоналізовані дідфейки, сконструйовані відповідно до ціннісного профілю реципієнта, значно ефективніше змінювали ставлення до кандидатів порівняно з неадаптованими матеріалами – навіть коли учасники отримували попередження про можливість маніпуляції (Dobber et al., 2021: 223–228). Цей результат є особливо тривожним з огляду на зростаючу

доступність даних соціальних мереж для конструювання детальних психографічних профілів виборців.

Трансформація політичної довіри та електоральної поведінки

Традиційна модель електорального вибору передбачає, що виборці формують свої переваги на основі раціонального аналізу програм кандидатів та оцінки їхньої компетентності. Ця модель, однак, неодноразово критикувалась задовго до появи генеративного штучного інтелекту – на підставі досліджень евристичного мислення та ролі партійної ідентичності. Поширення синтетичного контенту ще більше ускладнює ці й без того неоднозначні механізми.

Дослідники фіксують щонайменше три виразні трансформації в умовах інформаційно забрудненого середовища. По-перше, відбувається відхід від довіри інституціям: традиційні медіа та державні інституції поступово втрачають функцію верифікуючого посередника. Їх місце частково посідає «довіра до алгоритму» – нераціональне припущення, що платформа вже відфільтрувала хибне, – або ж, навпаки, загальна апатія щодо будь-яких офіційних джерел. По-друге, зростає залежність від персоналізованих мереж довіри. Роль мікроінфлюенсерів та онлайн-спільнот у верифікації інформації підвищується, що відтворює до-масмедійні механізми «сарафанного радіо» у значно динамічнішому й масштабнішому форматі. По-третє, спостерігається перехід від раціонального до евристичного прийняття рішень: «кандидат виглядає порядно», «він так говорить – значить, це правда».

Р. Чалдіні описав класичні принципи соціального впливу – соціальний доказ, авторитет, дефіцит, прив'язаність (Cialdini, 2007: 21–145), – які синтетичний контент здатен відтворювати штучно. Автоматично згенеровані коментарі, подроблені рейтинги підтримки, фейкові ролики «виборців, що підтримують» – усе це активує ті самі механізми переконання, що й реальні соціальні сигнали, і може суттєво спотворювати уявлення про настрої в суспільстві.

Г. Пеннікук і Д. Ренд у серії експериментальних досліджень довели, що сприйнятливість до дезінформації більшою мірою зумовлена не мотивованим мисленням, а недостатнім рівнем аналітичного опрацювання інформації. Люди поширюють дезінформацію не стільки тому, що хочуть цього, скільки тому, що не зупиняються, щоб перевірити (Pennycook & Rand, 2019: 44–47). Ключовим чинником, що визначає стійкість до маніпуляцій у синтетичному медіасередовищі, є рівень критичного мислення та уваги до джерел (Pennycook & Rand, 2021: 394–398). Водночас навіть поінформовані користувачі не позбавлені вразливості. При тривалому повторенні інформації ефект ілюзорної правди виявляється стійким навіть на тлі сформованих навичок критичного аналізу.

Стратегії когнітивного захисту: психологічне щеплення

На противагу методам спростування дезінформації після її поширення, у науковій літературі останнього десятиліття активно розвивається концепція

превентивного інформування (prebunking), яку також описують як психологічне щеплення. Як біологічна вакцина вводить ослаблений патоген для формування імунітету, так і психологічне щеплення ознайомлює людину з ослабленою версією маніпулятивного прийому, формуючи когнітивну резистентність до його повноцінного застосування.

С. Левандовські та С. ван дер Лінден у серії експериментів показали, що превентивне пояснення риторичних прийомів маніпуляції (апеляція до неіснуючого консенсусу, хибні дихотомії, апеляція до авторитету без підстав, використання цифр і статистичних даних) знижує ефективність наступних дезінформаційних повідомлень від 10 до 25% залежно від теми й аудиторії (Lewandowsky & van der Linden, 2021: 354–368). Важливо, що ефект зберігається без зазначення конкретних «ворожих» акторів і без ідеологічного забарвлення – достатньо нейтрального пояснення механіки маніпуляції. Дж. Рузенбек вважає, що щеплення зберігає ефективність протягом кількох тижнів і переноситься на нові, не використані в навчанні типи маніпуляцій (Roosenbeek et al., 2020: 143–147).

Ефективні програми медіаграмотності, засновані на принципах психологічного щеплення, містять кілька компонентів. Перший – навчання розпізнаванню технік маніпуляції без прив'язки до конкретних повідомлень чи кандидатів. Другий – формування звички «зупинки» перед репостом: прості підказки на зразок «чи перевіряв ти це джерело?» знижують поширення неперевіреної інформації (Lutzke et al., 2019). Третій – розвиток алгоритмічної грамотності, тобто розуміння того, що рекомендаційні системи максимізують залученість, а не достовірність, суттєво змінює суб'єктивне ставлення до контенту.

Водночас необхідно визнати обмеженість підходу. По-перше, щеплення виявляється менш ефективним для людей зі сталою ідеологічною прихильністю: якщо маніпулятивне повідомлення підтверджує їхні переконання, мотивація до критичного перегляду знижується. По-друге, швидкість технологічного розвитку може випереджати освітні програми. По-третє, щеплення вимагає масштабного охоплення – точкові програми не здатні зупинити системний вплив на медіасередовище загалом.

Висновки

Синтетичний контент є не просто новою формою дезінформації, а якісно іншим психологічним подразником, що системно використовує обмеження людського пізнання. Теорія подвійного процесу пояснює, чому навіть критично налаштовані реципієнти лишаються вразливими: контент, що апелює до Системи 1, отримує «пропуск» до переконань ще до того, як аналітична Система 2 встигає спрацювати. Когнітивні упередження – ілюзорна правда та підтверджувальне упередження – посилюють цю вразливість у середовищах із повторюваним і персоналізованим контентом.

Феномен «дивідендів брехуна» (Liar's Dividend) свідчить про те, що загроза виходить за межі конкретних підробок. Саме знання про можливість фальсифікації руйнує епістемічну інфраструктуру публічного дискурсу, відкриваючи шлях до заперечення автентичних доказів. Емоційні чинники – страх і гнів – прискорюють поширення маніпулятивних матеріалів, обходячи раціональну верифікацію.

Аналіз задокументованих кейсів – від президентської кризи у Габоні (2019) до аудіального дідфейку в Словаччині (2023) та синтезованих рободзвінків у США (2024) – підтверджує, що синтетичний контент уже застосовується як системний інструмент впливу на виборчі процеси. Показово, що його ефективність не залежить виключно від технічної досконалості, адже навіть «дешеві підробки» або невдалі дідфейки здатні формувати масштабну інформаційну невизначеність та підривати довіру до інституцій.

Трансформація електоральної поведінки під впливом синтетичного контенту виявляється у відході від інституційної довіри, поширенні евристичного прийняття рішень і зростанні ролі персоналізованих мереж довіри. Ці процеси не є детермінованими – на них суттєво впливає рівень алгоритмічної грамотності та критичного мислення аудиторії.

Стратегія психологічного щеплення є науково підтвердженим і перспективним інструментом протидії маніпуляціям, однак вимагає системного впровадження на рівні освіти та медіаполітики. Подальші дослідження мають зосередитися на розробленні масштабованих програм пребанкінгу, адаптованих до конкретних культурних і технологічних контекстів, а також на моніторингу їхньої ефективності в реальних виборчих кампаніях.

REFERENCES

1. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
2. Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780190923624.001.0001>
3. Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318. <https://doi.org/10.1073/pnas.1618923114>
4. Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1820. <https://doi.org/10.15779/Z38RV0D15J>
5. Cialdini, R. B. (2007). *Influence: The psychology of persuasion* (Rev. ed.). HarperCollins Publishers.

6. Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2020). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, 26(1), 69–91. <https://doi.org/10.1177/1940161220944364>
7. Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
8. Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002. <https://doi.org/10.1037/xge0000098>
9. Federal Communications Commission. (2024, February 8). *FCC makes AI-generated voices in robocalls illegal* [Press release]. <https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal>
10. Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, 51(3), 372–380. <https://doi.org/10.1037/a0036505>
11. Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). *Generative language models and automated influence operations: Emerging threats and potential mitigations*. arXiv. <https://doi.org/10.48550/arXiv.2301.04246>
12. Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)
13. Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81–93. <https://doi.org/10.1080/19331681.2018.1448735>
14. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
15. Lakoff, G. (2004). *Don't think of an elephant! Know your values and frame the debate*. Chelsea Green Publishing.
16. Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
17. Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
18. Lutzke, L., Drummond, C., Slovic, P., & Arvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, 58, Article 101964. <https://doi.org/10.1016/j.gloenvcha.2019.101964>

19. Paris, B., & Donovan, J. (2019). *Deepfakes and cheap fakes: The Manipulation of Audio and Visual Evidence*. *Data & Society*. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>
20. Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
21. Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
22. Roozenbeek, J., & van der Linden, S. (2022). Psychological Inoculation against Misinformation: Current Evidence and Future Directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 136–151. <https://doi.org/10.1177/000271622211087936>
23. Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5), 645–665. <https://doi.org/10.1017/S0140525X00003435>
24. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1–13. <https://doi.org/10.1177/2056305120903408>
25. Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe. <https://rm.coe.int/information-disorder-report-november-2017/1680764666>

Відомості про автора

Лілія МАКАРЕНКО, доктор політичних наук, доцент, завідувачка кафедри загальних та соціально-гуманітарних дисциплін, Київський інститут сучасної психології та психотерапії (КІСПП), м. Київ, Україна.

Liliia Makarenko, Doctor of Political Sciences, Associate Professor, Head of the Department of General, Social and Humanitarian Disciplines, Kyiv Institute of Modern Psychology and Psychotherapy (KIMPP), Kyiv, Ukraine

ORCID: 0000-0002-6676-2126

E-mail: makarenkolilia35@gmail.com