УДК 004.056

ОСОБЛИВОСТІ ВИБОРУ ЗАСОБІВ ВИЯВЛЕННЯ DEEPFAKE-АУДІО ДЛЯ СИСТЕМ ПРОТИДІЇ ВІШИНГУ

Веселовська М.Р.[1]; Бабенко Ю.М., доктор філософії[2]; Бабенко М.В., к.т.н., доц.[3]
[1,2] Київський національний університет ім. Т.Шевченко; [3] Дніпровський державний технічний університет
[1,2] Україна, Київ; [3] Україна, Кам'янське
[1] babenkomahalych@gmail.com; [2] yurii.babenko@knu.ua; [3] mvbab@ukr.net

Анотація. *У роботі розглядається проблема вибору засобів виявлення deepfake-аудіо для систем протидії вішингу з урахуванням умов голосового зв'язку. Запропоновано підхід, що передбачає два послідовні етапи: відсіювання варіантів, які не задовольняють жорстким обмеженням конкретного випадку застосування, та багатокритеріальне ранжування рішень, що залишилися.*
Ключові слова: *deepfake-аудіо; вішинг; голосовий зв'язок; засоби виявлення; багатокритеріальний вибір.*

Вступ. Поширення технологій синтезу та клонування голосу супроводжується стрімким зростанням кількості шахрайських голосових атак: за даними галузевого звіту Pindrop, що охоплює аналіз понад 1,2 млрд. телефонних дзвінків, кількість спроб шахрайства з використанням синтезованого голосу в контакт-центрах зросла більш ніж на 1300% протягом 2024 року [1]. Це актуалізує задачу впровадження засобів автоматичного виявлення deepfake-аудіо в системах протидії вішингу.

Основним орієнтиром для порівняння методів виявлення deepfake-аудіо є набір даних ASVspoof, у межах якого регулярно проводяться змагання з розробки детекторів [2]. Запропоновані на цій основі рішення, зокрема архітектура AASIST, демонструють високу точність детектування на еталонних наборах даних [3]. Водночас автори самого ASVspoof зазначають, що введення до протоколу оцінювання варіативності каналу передачі та компресії суттєво ускладнює задачу детектування навіть у межах контрольованого експерименту. Дослідження генералізаційної здатності детекторів на аудіо, зібраному поза лабораторними умовами, демонструє ще суттєвіше падіння якості – зростання частки помилок на 200-1000% порівняно з показниками на еталонних наборах [4].

Мета дослідження. Отже, заявлена ефективність відкритих детекторів deepfake-аудіо здебільшого встановлена на даних, які не відтворюють умов телефонного або VoIP-каналу зв'язку, характерного саме для вішингу (вузькосмугове кодування, стиснення, шум лінії). Крім того, наявні порівняльні дослідження зосереджені переважно на єдиному показнику точності детектування і не пропонують процедури вибору, яка враховувала б одночасно й стійкість до умов реального дзвінка, і практичні обмеження впровадження (доступність обчислювальних ресурсів, допустиму затримку обробки). Це формує проблему, на розв'язання якої спрямована робота: відсутність формалізованого підходу до вибору методу виявлення deepfake-аудіо, придатного для конкретних умов впровадження системи протидії вішингу.

Основний текст. Для розв'язання цієї проблеми пропонується підхід, що передбачає два послідовні етапи прийняття рішення. На першому етапі здійснюється відсіювання варіантів, які не задовольняють обмеженням конкретного випадку застосування – це дозволяє виключити з подальшого розгляду технічно непридатні рішення незалежно від їхньої точності. На другому етапі варіанти, що залишились придатними, ранжуються за сукупністю критері-

їв: якість детектування, стійкість до спотворень телефонного/VoIP-каналу (оцінена як різниця показників якості між чистими та деградованими зразками), швидкодія та ресурсомісткість. Такий порядок – спочатку жорстке відсіювання, а потім гнучке ранжування – відображає практичну логіку прийняття рішення: показник високої точності не має значення, якщо рішення неможливо розгорнути в наявній інфраструктурі.

Для визначення вагових коефіцієнтів критеріїв на етапі ранжування застосовується метод аналізу ієрархій, який дозволяє формалізувати попарне порівняння критеріїв і передбачає вбудовану перевірку узгодженості суджень [5]. Остаточний вибір методу визначення ваг і коло осіб, судження яких використовуватимуться (експерти з інформаційної безпеки чи користувач інструменту), потребують додаткового обґрунтування на наступному етапі роботи.

Висновки. Запропоновано підхід до вибору засобів виявлення deepfake-аудіо для систем протидії вішингу, який передбачає послідовне відсіювання технічно непридатних рішень за жорсткими обмеженнями та багатокритеріальне ранжування решти варіантів. Такий підхід дає змогу враховувати не лише якість детектування, а й стійкість до спотворень голосового каналу, швидкодію та ресурсомісткість, забезпечуючи обґрунтований вибір засобу відповідно до умов його практичного застосування.

Подальша робота передбачає формування набору аудіозразків (справжніх і синтезованих) у чистому вигляді та з імітацією деградації телефонного/VoIP-каналу, експериментальне тестування кількох відкритих детекторів на цьому наборі для отримання значень визначених критеріїв, а також перевірку запропонованої процедури відбору на кількох сценаріях з різними вимогами до впровадження.

Використання інструментів ШІ. Під час підготовки матеріалів для мовного редагування тексту використовувався інструмент штучного інтелекту ChatGPT.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Pindrop. (2025). *Pindrop's 2025 voice intelligence & security report reveals +1,300% surge in deepfake fraud*. PR Newswire. https://www.prnewswire.com/news-releases/pindrops-2025-voice-intelligence--security-report-reveals-1-300-surge-in-deepfake-fraud-302479482.html.

[2] Yamagishi, J., Wang, X., Todisco, M., Sahidullah, M., Patino, J., Nautsch, A., Liu, X., Lee, K. A., Kinnunen, T., Evans, N., & Delgado, H. (2021). ASVspoof 2021: Accelerating progress in spoofed and deepfake speech detection. *2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge*, 47–54. https://doi.org/10.21437/ASVSPOOF.2021-8.

[3] Jung, J. W., Heo, H. S., Tak, H., Shim, H. J., Chung, J. S., Lee, B. J., Yu, H. J., & Evans, N. (2022). AASIST: Audio Anti-Spoofing Using Integrated Spectro-Temporal Graph Attention Networks. In *2022 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2022 - Proceedings* (pp. 2405-2409). https://doi.org/10.1109/ICASSP43922.2022.9747766.

[4] Müller, N. M., Czempin, P., Dieckmann, F., Froghyar, A., & Böttinger, K. (2022). Does audio deepfake detection generalize? *Proceedings of Interspeech 2022*, 2783–2787. https://doi.org/10.21437/Interspeech.2022-108.

[5] Saaty, T. L., & Vargas, L. G. (2012). *Models, methods, concepts & applications of the analytic hierarchy process* (2nd ed.). Springer.

Veselovska Marharyta[1], Babenko Yurii[2]; Babenko Mykhailo[3]

Features of Selecting Deepfake Audio Detection Tools for Anti-Vishing Systems

Abstract. *The paper addresses the problem of selecting deepfake audio detection tools for anti-vishing systems, taking into account voice communication conditions. A selection approach is proposed, comprising two consecutive stages: filtering out options that do not meet the hard constraints of a specific application scenario, followed by multi-criteria ranking of the remaining solutions.*

Keywords: *deepfake audio, vishing, voice communication, detection tools, multi-criteria selection.*