УДК 004.45

СЕРВІС ФІЛЬТРАЦІЇ ПОШТИ НА БАЗІ МАШИННОГО НАВЧАННЯ

Толстопят М.В.[1], к.т.н, доцент; Тройніна А.С.[2]
[1,2] Національний університет «Одеська політехніка»
[1,2] Україна, Одеса
[1] max.tolstopyat@gmail.com; [2] atroinina@op.edu.ua

Анотація. *У ході системного дослідження високоадаптивного сервісу фільтрації електронної пошти "SpamGuard AI" була побудована модель «чорної скриньки» та проведена функціональна декомпозиція системи до четвертого рівня. Визначені ключові вимоги зовнішнього середовища (точність ≤ 0.1%, адаптивність ≤ 4 годин) використані для ідентифікації критичних проблемних ситуацій, пов'язаних із низькою швидкістю адаптації ML-моделі та перевищенням хибно-позитивних спрацьовувань (False Positives). Сформована ієрархічна модель складу дозволила прив'язати ці проблеми до конкретних структурних елементів підсистем ML Core та управління даними. Результати аналізу є основою для проектування оптимізованої архітектури системи..*

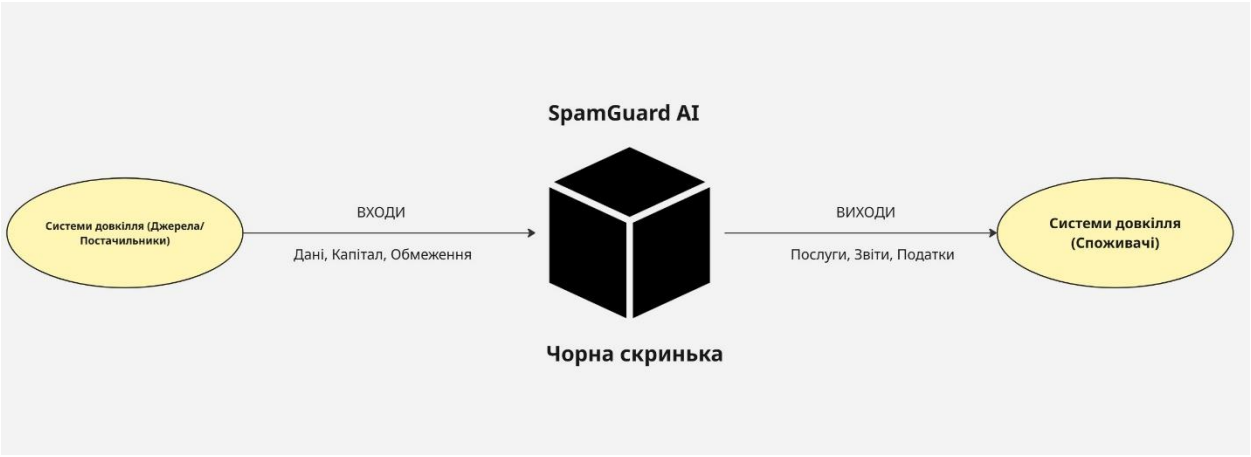
Ключові слова: *системний аналіз; модель «чорної скриньки»; декомпозиція; машинне навчання; антиспам; адаптивність.*

Вступна частина. Актуальність роботи зумовлена постійним зростанням динамічних високотехнологічних кіберзагроз (фішинг, поліморфний спам), проти яких традиційні статичні фільтри є неефективними. Розробка адаптивного рішення на основі інтеграції штучного інтелекту вимагає глибокого системного аналізу для верифікації його надійності, продуктивності та стійкості перед ринковими вимогами.

Метою роботи є перетворення нерегульованої пошти (вхід) на відфільтровані повідомлення (вихід).

Основна частина. Системний аналіз розпочато з формування моделі «чорної скриньки» [1] для об'єкта дослідження "SpamGuard AI".

Малюнок 1 – Модель «Чорної скриньки» системи “SpamGuard AI”



Ключова проблемна ситуація виявлена у точці зіткнення вимог до якості та адаптивності, особливо у потоці адаптивного зворотного зв'язку:

- Вимога споживача - хибно-позитивні спрацьовування (FP) не більше 0.1%;
- Власна вимога -час перенавчання моделі у відповідь на нову загрозу не більше 4 годин.

Фактичний стан системи, який гіпотетично не задовольняє цим вимогам, призводить до ключової проблеми: недостатня адаптивність системи, оскільки процес перенавчання займає більше часу, ніж того вимагає швидкість появи нових загроз.

Розкриття внутрішньої структури «чорної скриньки» здійснено за допомогою Ієрархічної змістовної моделі [2], що забезпечила декомпозицію системи на чотири рівні. Структура "SpamGuard AI" класифікована за функціональним принципом:

1. Основна діяльність - пряма обробка пошти та прийняття рішення (ML Core);
2. Забезпечуюча діяльність - управління даними, навчання моделі, моніторинг інфраструктури;
3. Управлінська діяльність - фінансове та стратегічне планування.

Деталізація структурних елементів (ПД, ЗД, КП) ключової підсистеми представлена у Таблиці 1.

Таблиця 1 – деталізація структурних елементів ключової підсистеми

Підсистема	ПД (Предмет Діяльності)	ЗД (Засоби Діяльності)	КП (Кінцевий Продукт)
ML Core	Парсований E-mail (токени, заголовки).	ML-модель (Transformer/RNN), Обчислювальний кластер.	Класифікаційний Результат (Score, Рішення СПАМ/НЕ-СПАМ).
Управління Даними	Сирий зворотний зв'язок (мітки).	Платформа MLOps [3], Сховище даних.	Валідована ML-модель (готова до розгортання).

Проведений аналіз взаємозв'язків підсистем показав, що проблема низької адаптивності (вимога ≤4 годин) безпосередньо корелює з елементами забезпечуючої діяльності:

- Проблема предмету діяльності - недостатня якість та своєчасність надходження зворотного зв'язку (ПД) від користувачів до підсистеми управління даними;
- Проблема засобу діяльності/процесу - тривалий час повного циклу перенавчання (параметр процесу) через неоптимізований обчислювальний кластер (ЗД).

Взаємозв'язок Постачальники → SpamGuard AI є критичним. Надходження обчислювальних потужностей (вхід) безпосередньо впливає на продуктивність (властивість) і є чинником, що обмежує можливість виконання вимог адаптивності (вихід).

Висновок. Системний аналіз "SpamGuard AI" підтвердив, що ідентифіковані проблеми низької точності та повільної адаптації є наслідком дисбалансу між потребами ключової підсистеми ML Core та можливостями забезпечуючої діяльності. Подальша робота повинна бути зосереджена на оптимізації засобів діяльності (вибір легших ML-моделей, що потребують менше обчислень) та автоматизації предметів діяльності (прискорений збір та валідація зворотного зв'язку) для забезпечення стійкості, рентабельності та виконання всіх вимог систем довкілля.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Gunduz, A., & Ozyer, B**.** (2022). *A Novel MLOps Framework for Explainable AI in Real-Time Systems*, Proceedings of the 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Prague, Czechia: IEEE.

[2] Чорней, Н.Б., & Чорней, Н.К. (2005). *Теорія систем і системний аналіз*. Київ: МАУП.

[3] Rane, J., Mallick, S. K., Kaya, O., & Rane, N. L**.** (2024). Enhancing black-box models: Advances in explainable artificial intelligence for ethical decision-making. *Future Research Opportunities for Artificial Intelligence in Industry 4.0 and 5.0*., 13(15), 136-180.

Tolstopyat M.V, Troynina A.S.

Email filtering service based on machine learning

Abstract. *In the course of a system study of the highly adaptive email filtering service "SpamGuard AI", a black-box model was constructed and a functional decomposition of the system was carried out to the fourth level. Key external environment requirements were defined (accuracy ≤ 0.1%, adaptability ≤4 hours) and used to identify critical problematic situations associated with the low adaptation speed of the ML model and the high rate of False Positives. The developed hierarchical composition model allowed these issues to be linked to specific structural elements of the ML Core and data management subsystems. The results of this analysis form the basis for designing an optimized system architecture.*

Keywords: *system analysis; black-box model; decomposition; machine learning; anti-spam; adaptability.*