УДК 004.7 **doi: 10.26906/SUNZ.2022.2.041**

І. О. Васильєв, В. С. Харченко

Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

ФРЕЙМВОРК ДЛЯ МЕТРИЧНОГО ОЦІНЮВАННЯ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ НА ОСНОВІ МОДЕЛІ ЯКОСТІ

Анотація. Мотивація. На сьогодні вкрай важливо розуміти, чи можна довіряти системам, що базуються на використанні штучного інтелекту (СШІ). Велика кількість сучасних СШІ побудовано за принципом «чорної скриньки», тобто незрозуміло, яким чином вони працюють, а тільки є результати роботи. Також потрібні засоби для порівняння декількох СШІ. У випадку, коли декілька варіантів ШІ конкурують щодо використання у деякій системі, потрібно визначити кращий. **Метою дослідження** є розроблення модель-базованого фреймворку для оцінювання якості СШІ з використанням метрик і методу візуалізації результатів. **Етапи дослідження**. В статті аналізуються моделі якості СШІ, метрики і види згорток для її оцінювання, пропонується метод оцінювання та візуалізації результатів і приклад використання методу. **Висновки.** Для оцінювання СШІ використано базові моделі якості, об'єднані у чотирьохрівневу ієрархію. Для цих характеристик визначено правила формування метрик і метод розрахунку якості з використанням згорток та візуалізації проміжних і кінцевих результатів за допомогою радіальних метричних діаграм. Відповідні моделі якості, метрики і методи оцінювання і візуалізації утворюють фреймворк для автоматизації процесів, який реалізується з використанням розробленого інструментального засобу. Цей засіб дозволяє користувачу створювати модель якості, встановлювати метрики якості, вводити значення показників метрик. Потім на основі цих показників розраховується узагальнена метрика якості системи та візуалізується за допомогою РМД. Засіб є десктопним застосунком, створеним на платформі .Net Framework. **Напрям подальших досліджень.** Майбутні кроки можуть бути присвячено розвитку моделі та інструментарію оцінювання якості для різних доменів з урахуванням аспектів еволюції якості.

Ключові слова: система штучного інтелекту, оцінювання якості, метрики оцінювання, візуалізація, фреймворк.

Вступ

Мотивація. В останній час розробляються і впроваджуються різні засоби штучного інтелекту (ШІ), що виконують відносно прості, на перший погляд, завдання, які займали багато часу та зусиль у минулому, зокрема, оброблення багаторозмірних зображень та відео, анімації статичних зображень, розпізнавання облич, розроблення чат-ботів та інші. Активно досліджуються і створюються засоби ШІ для більш складних задач, а саме, встановлення хвороби пацієнта на основі симптомів, розроблення асистентів для пілотів літака, виявлення зловмисників тощо [1, 2].

Вкрай важливо розуміти, чи можна довіряти системам, що базуються на використанні штучного інтелекту (СШІ). Велика кількість сучасних СШІ побудовано за принципом «чорної скриньки», тобто незрозуміло, яким чином вони працюють, а тільки є результати роботи. Важко перевірити, яким чином ШІ приймає рішення, чи є вони взагалі вірним або помилковим. Також потрібні засоби для порівняння декількох СШІ. У випадку, коли декілька варіантів ШІ конкурують щодо використання у деякій системі, потрібно визначити кращий.

Аналіз публікацій. Модель якості СШІ може бути представлено у вигляді графу типу «дерево», що надає упорядковану ієрархію характеристик [3]. Вона будується за аналогією з моделями якості програмного забезпечення [4-6]. Характеристики відібрано на підставі аналізу документів [6-10], гармонізації їх визначень та пошуку залежностей відповідно до [3].

Ці складові можуть далі розділятися на свої субхарактеристики. У кожної складної характеристики повинно буди щонайменше дві субхарактеристики. Отже, використані в даному досліджені моделі базуються на результатах роботи [3], яка була апробована для побудови моделей якості СШІ, описаних в [11, 12].

Метою дослідження є розроблення модель-базованого фреймворку для оцінювання якості СШІ з використанням метрик і методу візуалізації результатів. Відповідно до мети в статті аналізуються моделі якості СШІ (розділ 1), метрики і види згорток для її оцінювання (розділ 2), пропонується метод оцінювання якості та візуалізації результатів (розділ 3) і приклад використання методу (розділ 4). У висновках аналізуються основні результати дослідження та розроблення відповідного інструментального засобу, а також описуються наступні кроки, спрямовані на розвиток моделей і методів для різних сфер застосування СШІ.

1. Моделі якості СШІ

Загальна структура моделі. В основі моделі (рис. 1) лежить оцінка загальної якості системи (artificial intelligence system, AIS).

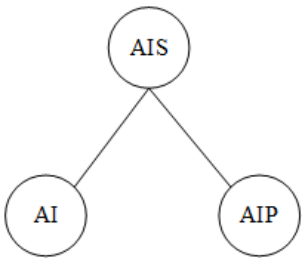

Рис. 1. Модель якості СШІ

Вона розділяється на два складових [3]:

– платформа ШІ (artificial intelligence platform, AIP) – середовище, у якому працює штучний інтелект. Вона відповідає за взаємодію з користувачем, управління штучним інтелектом, передачу йому даних та оброблення результатів. Платформа являє собою звичайну програму, написану людиною. Це може бути додаток на ПК, або хмарний сервіс. Через це, загалом, вимоги до платформи співпадають з вимогами до звичайних програм;

– модель ШІ (artificial intelligence, AI) – навчений штучний інтелект, шо приймає деякі дані на вході, да формує результати обчислень або керуючі впливи (прийняті рішення) на виході.

Модель якості платформи ШІ (рис. 2) включає такі характеристики [3,6,7]:

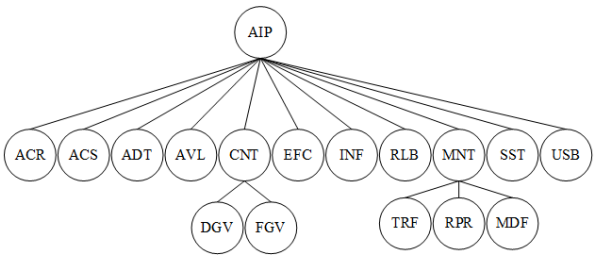

Рис. 2. Платформа ШІ

– доступність (accessibility, ACS);

– точність (accuracy, ACR);

– аудитопридатність (auditability, ADT);

– готовність (availability, AVL);

– керованість (controllability, CNT) – що включає керування даними (data governance, DGV) та керування функціями (function governance, FGV);

– ефективність (effectiveness, EFC);

– інформативність (informativeness, INF);

– надійність (reliability, RLB);

– обслуговуваність (maintainability, MNT) – що включає переносимість (transferability, TRF), ремонтопридатність (repairability, RPR) та модифікованість (modifiability, MDF);

– сталість (sustainability, SST);

– зручність (usability, USB).

Модель якості ШІ (рис. 3) включає п’ять важливих характеристик [3, 6, 7]:

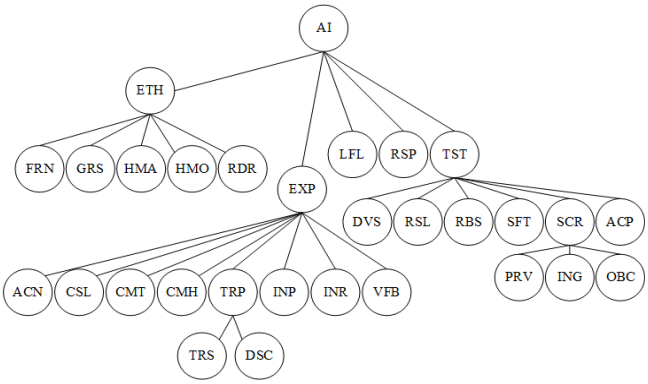

Рис. 3. Модель ШІ

– етичність (ethics, ETH) – здатність ШІ відповідати діючим нормам моралі за результатами функціонування. Вона має субхарактеристики: справедливість (fairness, FRN), сприйнятливість (graspability, GRS), людський нагляд (human oversight, HMO) та відшкодовуваність (redress, RDR);

– пояснюваність (explainability, EXP) – здатність ШІ бути зрозумілим і передбачуваним з точки зору призначення та поведінки. Вона включає субхарактеристики: відстежуваність (accountability, ACN), причинність (causality, CSL), завершеність (completeness, CMT), зрозумілість (comprehensibility, CMH), прозорість (transparency, TRP), простежуваність (traceability, TRC), описуваність (descriptiveness, DSC), інтерпретабельність (interpretability, INP), інтерактивність (interactivity, INR), верифікованість (verifiability, VFB);

– законність (lawfulness, LFL) – здатність ШІ відповідати законодавчим і нормативно-правовим актам;

– відповідальність (responsibility, RSP) – здатність ШІ функціонувати з урахуванням очікувань замовника (користувача) у відповідності до етичних норм, законодавчих нормативно-правових актів, а також інформувати його при їх порушення;

– довірчоздатність (trustworthiness, TST) - здатність ШІ, яка характеризується ступенем впевненості користувача або іншої зацікавленої особи (розробника, аудитора тощо) в тому, що ШІ відповідає вимогам і виконує функції у передбачуваний спосіб. Вона складається з субхарактеристик: диверсність (diversity, DVS), резильєнтність (resiliency, RSL), робастність (Robustness, RBS), функційна безпечність (safety, SFT), захищеність (інформаційна/кібербезпека) (security, SCR), приватність (privacy, PRV), цілісність (integrity, ING), об’єктивність (objectivity, OBC), прийнятність (acceptability, ACP).

3. Метрики для оцінювання якості СШІ

Метрики якості. Для розрахунку значень характеристик у моделі використовуються різні метрики. Кожна характеристика нижнього рівня повинна мати хоча б одну метрику. Ці метрики встановлюються в залежності від конкретного типу СШІ, вимог до системи, особливостей її розробки тощо. Результатами розрахунку метрик є показник характеристики, який може бути [5]:

– метричний (абсолютний або відносний);

– порядковий (рівневий);

– номінальний (виконується повністю, частково або зовсім не виконується).

Оскільки декілька показників можуть мати різні типи (кількісні та якісні), необхідний спосіб переведення одних показників в інші. Для цього їх потрібно нормалізувати та привести до єдиної шкали від 0 до 1. Для нормалізації метричних показників можна використовувати формулу 1.

$$p = \frac{m - m_{min}}{m_{max} - m_{min}}, \quad (1)$$

де p – нормалізоване значення показника; m – оцінене значення показника; m_{max} – максимальне значення показника (або значення при якому вважається що система повністю відповідає характеристиці);

m_{min} – мінімальне значення показника (або значення, при якому вважається, що система повністю не відповідає вимогам до характеристики).

Для якісних показників кожне з можливих значень шкали відповідає певному метричному значенню, граничні значення повинні дорівнювати 0 і 1. Розподілення між шкалою необов'язково має бути пропорційним. Приклад нормалізації якісної шкали зображено в табл. 1.

Таблиця 1 – **Нормалізація порядкової шкали**

Вихідна порядкова шкала	Нормалізоване значення
Не відповідає	0
Відповідає частково	0,33
Відповідає повністю	1

Оскільки одні характеристики можуть бути більш важливими, ніж інші, вводяться вагові коефіцієнти, які змінюються від 0 до 1.

Вони визначають, наскільки ця субхарактеристика є важливою та яку частину вона складає у характеристиці вищого рівня. Сума вагових коефіцієнтів субхарактеристик однієї характеристики завжди має дорівнювати одиниці.

Згортки. Для характеристик проміжних рівнів та якості в цілому виконується операція згортки. Вона полягає у розрахунку якості характеристики на основі значень метрик її субхарактеристик. Ця операція може мати декілька варіантів реалізації, як це визначено в [5] для оцінки якості ПЗ:

– адитивна згортка;
– згортка на основі нечітких операцій;
– предикативна згортка;
– булева згортка;
– комбінована згортка.

Адитивна згортка полягає у сумі зважених нормалізованих показників усіх субхарактеристик:

$$P = \sum_{i=1}^{n} w_i p_i, \qquad (2)$$

де P – значення показника характеристики; n – кількість субхарактеристик характеристики; w_i – ваговий коефіцієнт i-ої субхарактеристики; p_i – значення показника i-ої субхарактеристики.

Згортка на основі нечітких операцій полягає у формуванні правил розрахунку значення показника характеристики. Такі правила можуть бути будь-якими, головне, щоб у результати був отриманий нормалізований показник.

У предикативній згортці для пошуку значення показника формується набір предикатів – логічних правил, за якими він визначається.

У булевій згортці значення усіх показників мають набувати булевих значень, тобто 0 або 1. Значення показника характеристики виконується за певною булевою функцією, та набуває значення 0 – не відповідає, або 1 – відповідає.

У комбінованій згортці використовуються декілька згаданих методів – це необхідно, якщо у субхарактеристиках задіяні різні види показників, як порядкових, так і метричних. Тоді має виконуватися окрема згортка показників кожного типу, щоб отримати кінцеве значення інтегрованого показника.

4. Метод оцінювання та візуалізації результатів

Радіальні метричні діаграми. Для візуалізації моделі використовуються радіальні метричні діаграми (РМД) [5]. Загальна схема такої діаграми зображена на рис. 4.

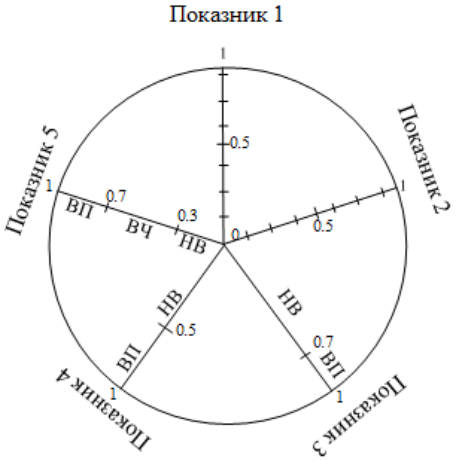

Рис. 4. Схема РМД

Центральна точка діаграми є точкою відліку показників. Для кожної з субхарактеристик проводиться вісь з цієї точки. На цьому промені встановлюється шкала оцінювання, в залежності від типу показника. На рис. 4 показники 1 та 2 показують приклад метричних показників, 3 та 4 – приклад двозначного показника («відповідає», «не відповідає»), 5 – тризначного показника («відповідає», «відповідає частково», «не відповідає»).

Під час оцінки, на шкалі позначається поточне значення показника. Потім усі сусідні позначені точки об'єднуються лініями та отримана фігура замальовується. Така діаграма дозволяє наочно показати рівень оцінки характеристики. Чим більше площа отриманої фігури, тим вище рівень якості. Також одразу видно, які з субхарактеристик не відповідають потребам якості.

Послідовність оцінювання. Оцінювання за допомогою моделі якості виконується у наступні кроки:

1) встановити характеристики, які будуть оцінюватися, з урахуванням стандартів галузі роботи СШІ, вимог до системи тощо;

2) визначити ступінь впливу цих характеристик на якість системи, та встановити вагові коефіцієнти для усіх характеристик;

3) встановити метрики для оцінювання усіх характеристик нижніх рівнів, та задати їх мінімальне та максимальне значення, за якими вони набувають значень 0 та 1 відповідно;

4) обрати тип згортки для розрахунку значень характеристик вищих рівнів;

5) розрахувати значення показників нижчих рівнів за встановленими метриками;

6) за встановленими методами згортки, розрахувати інші характеристики;

7) візуалізувати необхідні характеристики за допомогою РМД.

4. Приклад оцінювання якості СШІ

При оцінюванні моделі ШІ для автопілоту автомобіля [13] були відібрані характеристики, для яких відповідні вершини маркуються світлим кольором на рис. 5.

Для оцінювання довірчоздатності експертним шляхом встановлено вагові коефіцієнти (табл. 2).

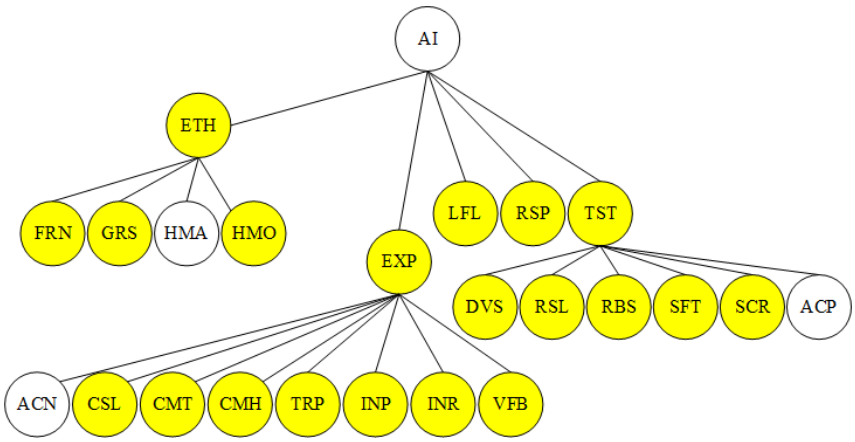

Рис. 5. Модель ШІ автопілоту

Таблиця 2 – **Розрахунок оцінки довірчоздатності**

Довірчоздатність		
Субхарактеристика	Ваговий коефіцієнт	Значення показника
Диверсність	0,1	0,7
Резильєнтність	0,2	0,9
Робастність	0,2	0,9
Функційна безпечність	0,3	0,95
Інформаційна захищеність	0,2	0,8
Значення довірчоздатності		0,875

Потім визначено як приклад фіксовані значення метрик субхарактеристик. Ці значення можуть бути обраховано шляхом обчислення відношення кількості успішних тестів (або кількості вимог, які виконано) для відповідної субхарактеристики до їх загальної кількості.

Тести і вимоги можуть, залежно від їх важливості, також бути зваженими, що підвищує точність метрик.

За допомогою адитивної згортки (3) розраховується показник довірчоздатності. На цій підставі розробляється РМД, для відображення рівня довірчоздатності СШІ (рис. 6).

Її значення визначає відповідну складову для РМД якості ШІ в цілому (рис. 7).

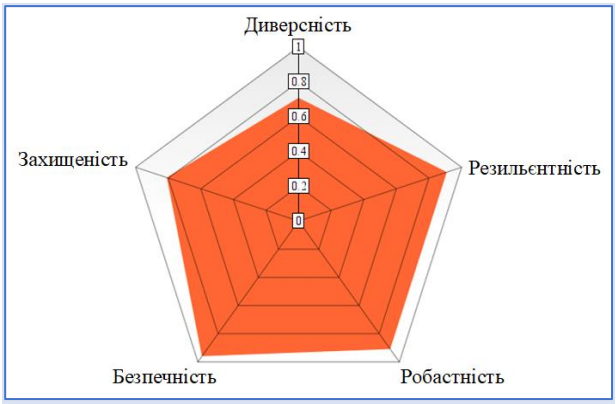

Рис. 6. РМД довірчоздатності

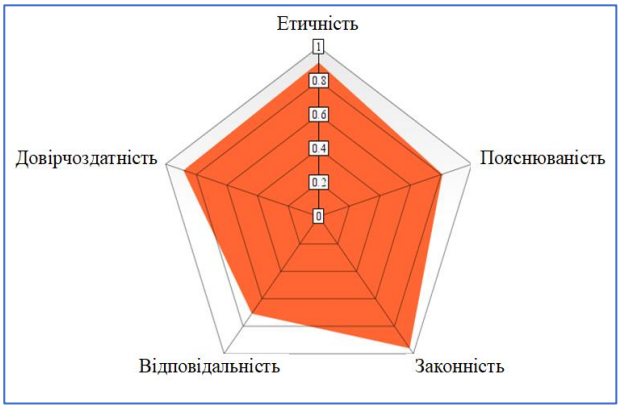

Рис 7. РМД моделі ШІ

Висновки

Результати. Для оцінювання СШІ використано базові моделі якості, запропоновані в [3] і об'єднані у чотирьохрівневу ієрархію.

Для цих характеристик визначено правила формування метрик і метод розрахунку якості з використанням згорток та візуалізації проміжних і кінцевих результатів за допомогою радіальних метричних діаграм.

Відповідні моделі якості, метрики і методи оцінювання і візуалізації утворюють фреймворк для автоматизації процесів, який реалізується з використанням розробленого інструментального засобу.

Цей засіб дозволяє користувачу створювати модель якості СШІ (або використовувати запропоновану

в [3] і адаптовану у цій статті), встановлювати метрики якості, вводити значення показників метрик. Потім на основі цих показників розраховується узагальнена метрика якості системи та візуалізується за допомогою РМД. Засіб є десктопним застосунком, створеним на платформі .Net Framework.

Майбутні кроки можуть бути присвячено розвитку моделі та інструментарію (метрик, методик і засобів) оцінювання якості для різних доменів (оборона, медицина, юриспруденція, інтерактивне мистецтво тощо) з урахуванням аспектів еволюції якості [14].

СПИСОК ЛІТЕРАТУРИ

1. Trustworthy AI [Text] / R. Chatila, V. Dignum, M. Fisher, F. Giannotti, K. Morik, S. Russell, K. Yeung // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): collective monograph, edited by B. Braunschweig, M. Ghallab. – Cham: Springer International Publishing, 2021. – Vol. 12600. – P. 13-39. DOI: 10.1007/978-3-030-69128-8.
2. A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks [Text] / M. R. Islam, M. U. Ahmed, S. Barua, S. Begum // Applied Sciences. – 2022. – Vol. *12*. – Article Id: 1353. DOI: 10.3390/app12031353
3. Харченко В. С., Фесенко Г. В., Ілляшенко О. О. (2022), Базова модель нефункційних характеристик для оцінки якості штучного інтелекту // Радіоелектронні і комп'ютерні системи 2(102). с. 1-14.
4. ISO/IEC 25010 (2011). ISO/IEC 25010:2011, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models.
5. Харченко В.С., Жихарев В.Я., Іллюшко В.М. та ін. (2004), Основи надійності цифрових систем, Харків: Нац. Аерокосм. Ун-т. «ХАІ».
6. NIST Four Principles of Explainable Artificial Intelligence: Draft NISTIR 8312 / P. J. Phillips, C. A. Hahn, P. C. Fontana, D. A. Broniatowski, M. A. Przybocki, C. A. Hahn, P. C. Fontana. – Gaithersburg: National Institute of Standards and Technology, 2020. – 24 p. DOI: 10.6028/NIST.IR.8312.
7. European Commission, Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI, Publications Office, (2019), https://data.europa.eu/doi/10.2759/346720
8. UNESCO (2021), Recommendation on the Ethics of Artificial Intelligence, https://unesdoc.unesco.org/ark:/48223/pf0000380455, Дата звернення 21.05.2022
9. ISO/IEC TR 24028:2020. Information technology. Artificial intelligence. Overview of trustworthiness in artificial intelligence [Electronic resource]. – Available at: https://www.iso.org/standard/77608.html. – 10.03.2022.
10. OECD. Tools for Trustworthy AI: A Framework to Compare Implementation Tools [Electronic resource]. – Available at: https://www.oecd.org/science/tools-for-trustworthy-ai-008232ec-en.htm. – 10.03.2022.
11. Москаленко В. В. Багатоетапний метод глибинного навчання з попереднім самонавчанням для класифікаційного аналізу дефектів стічних труб [Текст] / В. В. Москаленко, М. О. Зарецький, А. С. Москаленко, А. Г. Коробов, Я. Ю. Ковальський // Радіоелектронні і комп'ютерні системи. – 2021. – № 4. – С. 71-81. DOI: 10.32620/reks.2021.4.06.
12. Kuchuk, H. System of license plate recognition considering large camera shooting angles [Text] / H. Kuchuk, A. Podorozhniak, N. Liubchenko, D. Onischenko // Radioelectronic and Computer Systems. – 2021. – No. 4. – P. 82-91. DOI: 10.32620/reks.2021.4.07
13. Some, Evariste & Gondwe, Greg & Rowe, Evan. (2019). Cybersecurity and Driverless Cars: In Search for a Normative Way of Safety. 352-357. 10.1109/IOTSMS48152.2019.8939168.
14. Gordieiev, O. IT-oriented software quality models and evolution of the prevailing characteristics [Text] / O. Gordieiev, V. Kharchenko // Dependable Systems, Services and Technologies (DESSERT): Proceeding of 9th Int. Conf., 2018. – P. 375-380. DOI: 10.1109/DESSERT.2018.8409162.

Received (Надійшла) 30.03.2022
Accepted for publication (Прийнята до друку) 18.05.2022

A framework for metric evaluation of artificial intelligence systems based on quality model

Ihor Vasyliev, Vyacheslav Kharchenko

Abstract. Motivation. Nowadays it is crucially important to understand whether systems based on artificial intelligence (AI) can be trusted. Many modern AI systems are built according to the "black box" principle, i.e. it is not clear how they work, but we see only the results of their work. Besides, it is needed tools to compare different AI solutions. When several AIs are competing for use in some system, it is required to determine the best one. **The goal of the research** is to develop a model-based framework to evaluate the quality of an AI system (AIS) using metrics and a method for visualizing the evaluation results. **Research stages.** The article analyzes the models of AIS quality, metrics and types of convolution for its evaluation, proposes a method for evaluating and visualizing the results and describes an example of applying the method. **Conclusions.** The basic models of quality, combined into a four-level hierarchy are used to assess AIS. The rules of metrics formation and the method of quality calculation using convolutions and visualization of intermediate and final results using radial metric diagrams have been defined for these characteristics. Corresponding quality models, metrics, and evaluation and visualization methods provide implementing automation framework by use of the developed tool. This tool allows the user to create a quality model, set metrics, and enter metrics values. Then, based on these metrics, a generalized quality metric for the system is calculated and visualized using the radar diagrams. The tool is a desktop application created on .Net Framework platform**. The direction of further research.** Forthcoming steps can be devoted to development of the model and tools for quality assessment for different domains, considering the aspects of quality evolution.

Keywords: artificial intelligence system, quality assessment, evaluation metrics, visualization, framework.