

DOI: <https://doi.org/10.36910/6775-2524-0560-2026-64-10>

УДК 004.8:004.932

Гоцалюк Геннадій Петрович, аспірант

<https://orcid.org/0009-0006-4213-7437>

Луцький національний технічний університет, м. Луцьк, Україна

ОЦІНЮВАННЯ ПРОСТОРОВОЇ УЗГОДЖЕНОСТІ КАРТИ МАШИННОЇ ЗНАЧУЩОСТІ ТА ПРОГНОЗОВАНОЇ КАРТИ ЛЮДСЬКОЇ ВІЗУАЛЬНОЇ УВАГИ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Гоцалюк Г. П. Оцінювання просторової узгодженості карти машинної значущості та прогнозованої карти людської уваги в системах підтримки прийняття рішень. У статті запропоновано теоретико-методичний підхід до оцінювання відповідності між картою значущості, сформованою методом пояснюваного штучного інтелекту для конкретного прогнозу моделі комп'ютерного зору, та картою прогнозованої людської візуальної уваги. Наголошено, що прогнозована карта не є еквівалентом емпіричних даних про погляд і має бути попередньо валідована за результатами eye tracking у відповідному сценарії спостереження. Для усунення обмежень однієї метрики запропоновано багатокритеріальне оцінювання, що поєднує коефіцієнт кореляції, міру подібності розподілів, дивергенцію Єнсена–Шеннона та, за наявності фіксацій, Normalized Scanpath Saliency і Area Under the ROC Curve. Часткові оцінки нормалізуються та утворюють інтегральний індекс узгодженості з вагами, які мають визначатися на валідаційній вибірці. Індекс запропоновано використовувати не як доказ правильності чи надійності прогнозу, а як додатковий індикатор необхідності перевірки рекомендації користувачем. Сформовано послідовність підготовки карт, обчислення метрик, перевірки граничних випадків і формування інформації для системи підтримки прийняття рішень. Передбачено окреме подання профілю метрик, вихідної впевненості моделі, візуального пояснення та застереження про межі інтерпретації показників.

Ключові слова: візуальна увага, комп'ютерний зір, карти значущості, пояснюваний штучний інтелект, узгодженість, система підтримки прийняття рішень, eye tracking.

Hotsaliuk H. Assessment of the spatial agreement between a machine saliency map and a predicted human visual attention map in decision support systems. The paper proposes a methodological approach to assessing agreement between a saliency map generated by an explainable artificial intelligence method for a computer vision prediction and a map of predicted human visual attention. The predicted map differs from empirical gaze data and requires validation against eye-tracking observations from a matching viewing task. To overcome the limitations of a single measure, the approach combines the correlation coefficient, distribution similarity, Jensen-Shannon divergence and, when fixation data are available, Normalized Scanpath Saliency and Area Under the ROC Curve. Partial scores are normalized and aggregated into an integral agreement index, with weights estimated on a validation dataset. The index is not interpreted as evidence that a prediction is correct or reliable; instead, it is an additional indicator that may trigger human review. The paper defines map preparation, metric computation, boundary-case treatment and communication of results within a decision support system.

Keywords: visual attention, computer vision, saliency maps, explainable artificial intelligence, agreement, decision support system, eye tracking.

Постановка наукової проблеми. Розвиток методів штучного інтелекту та комп'ютерного зору забезпечив широкі можливості автоматизованого аналізу візуальної інформації. Сучасні моделі здатні виконувати класифікацію, виявлення, локалізацію та сегментацію об'єктів на цифрових зображеннях, що створює передумови для їх використання в системах підтримки прийняття рішень. Водночас результат роботи моделі комп'ютерного зору зазвичай представлений визначеним класом об'єкта та числовою оцінкою впевненості. Такий підхід не завжди дозволяє встановити, на підставі яких саме візуальних ознак було сформовано результат і наскільки ці ознаки відповідають особливостям сприйняття зображення людиною.

Ця проблема є особливо важливою для систем, у яких результат роботи алгоритму використовується не як остаточне автоматичне рішення, а як рекомендація користувачеві. У таких системах необхідно враховувати не лише точність прогнозу, але й можливість його інтерпретації та оцінювання обґрунтованості. Одним із напрямів вирішення цієї проблеми є застосування Explainable Artificial Intelligence (XAI), що дає змогу визначити візуальні області, які найбільше вплинули на результат роботи моделі. Інший напрям пов'язаний із моделюванням людської візуальної уваги та прогнозуванням областей зображення, які привертають увагу людини.

Поєднання цих двох напрямів створює можливість порівнювати області, значущі для алгоритму штучного інтелекту, з областями, значущими для людського сприйняття. Відповідність або розбіжність між ними може містити додаткову інформацію щодо характеру сформованого моделлю результату та потенційно використовуватися у процесі підтримки прийняття рішень. Тому актуальним є завдання формалізації показників, які дають змогу кількісно оцінити узгодженість

машинної карти значущості з прогнозованою або емпіричною картою людської уваги та використати таку оцінку як додатковий індикатор у процесі підтримки прийняття рішень.

Нехай система комп'ютерного зору аналізує вхідне зображення та формує результат класифікації або розпізнавання разом із рівнем впевненості моделі. За допомогою засобів ХАІ для цього самого результату може бути сформована карта, яка відображає області зображення, що найбільше вплинули на прогноз моделі. Паралельно модель прогнозування людської візуальної уваги формує очікуваний просторовий розподіл уваги. Така карта є модельним наближенням і не повинна ототожнюватися з фактичними фіксаціями погляду конкретного користувача. Виникає задача визначення ступеня відповідності між зазначеними картами, перевірки прогнозованої карти за емпіричними eye-tracking даними та дослідження можливості використання отриманих оцінок у системі підтримки прийняття рішень.

Аналіз останніх досліджень і публікацій. Проблема пояснення результатів роботи моделей штучного інтелекту активно досліджується в межах напряму Explainable Artificial Intelligence. У роботі V. Hassija та ін. [1] систематизовано сучасні методи ХАІ та показано, що пояснюваність є важливою складовою забезпечення прозорості моделей машинного і глибокого навчання. Особливе значення для моделей комп'ютерного зору мають методи, що дозволяють візуально визначити області зображення, які здійснили найбільший вплив на отриманий результат. Використання ХАІ у системах підтримки прийняття рішень розглянуто G. Kostopoulos, G. Davrazos та S. Kotsiantis [2]. Автори відзначають важливість прозорості та інтерпретованості рекомендацій інтелектуальних систем для забезпечення ефективної взаємодії між алгоритмом і користувачем. Це свідчить про необхідність аналізувати не лише результат роботи моделі, але й інформацію, на якій він ґрунтується. У роботі K. Morrison та ін. [3] досліджено вплив способів пояснення результатів штучного інтелекту на прийняття людиною рішень у візуальних задачах. Результати експериментів показали, що спосіб представлення пояснення може впливати на поведінку користувача та рівень його залежності від рекомендації системи. Отже, пояснення результату є складовою процесу взаємодії людини та інтелектуальної системи, а не лише засобом технічної інтерпретації моделі. Паралельно розвиваються методи моделювання людської візуальної уваги. Z. Yang та ін. [4] запропонували Human Attention Transformer для прогнозування послідовностей фіксацій погляду людини. Модель поєднує bottom-up і top-down механізми уваги та дозволяє прогнозувати поведінку користувача як під час вільного перегляду зображення, так і під час цілеспрямованого візуального пошуку.

Перспективність моделей saliency для прогнозування людської візуальної уваги підтверджена розвитком спеціалізованих бенчмарків. Водночас оцінювання таких моделей потребує кількох взаємодоповнювальних метрик, оскільки кожна з них відображає окрему властивість карти або фіксацій [7, 8]. Найближчим до предмета цього дослідження є підхід G. Liu та ін. [5], у якому людська візуальна увага використовується для вдосконалення Explainable Artificial Intelligence моделей комп'ютерного зору. Запропонований авторами підхід Human Attention Guided ХАІ демонструє можливість інтеграції інформації про людську увагу з картами пояснення моделі та підтверджує доцільність дослідження взаємозв'язку між людською візуальною увагою і значущістю ознак зображення для прогнозу моделі. Питання використання ХАІ безпосередньо у процесі прийняття рішень розглядають K. Coussemant та ін. [6]. Автори підкреслюють необхідність оцінювати пояснюваний штучний інтелект з позиції його впливу на якість прийняття рішень людиною, а не лише за характеристиками самого алгоритму.

Таким чином, сучасні дослідження підтверджують перспективність поєднання моделей прогнозування людської візуальної уваги та засобів ХАІ. Недостатньо дослідженим залишається контекстно залежне багатокритеріальне оцінювання відповідності між машинною картою значущості та прогнозованою людською увагою, валідованою за емпіричними фіксаціями, а також використання такої оцінки як індикатора необхідності додаткової перевірки рекомендації.

Виділення невирішених раніше частин загальної проблеми. Недостатньо дослідженим залишається кількісне оцінювання узгодженості карт машинної та прогнозованої людської візуальної уваги. Потребують обґрунтування добір і поєднання відповідних метрик, а також використання отриманої оцінки як додаткового індикатора перевірки рекомендацій у системах підтримки прийняття рішень.

Мета та завдання дослідження. Метою роботи є розроблення теоретико-методичного підходу до багатокритеріального оцінювання узгодженості між картою машинної значущості та

валідованою картою прогнозованої людської візуальної уваги для використання отриманого індексу як додаткового індикатора у системах підтримки прийняття рішень. Завданнями є формалізація порівнюваних представлень, добір взаємодоповнювальних метрик, визначення правил агрегування та граничних випадків, а також формування протоколу майбутньої експериментальної перевірки.

Основна частина дослідження.

Представлення машинної та прогнозованої людської візуальної уваги.

Нехай кольорове вхідне зображення задано тензором $I \in \mathbb{R}^N$, де N і W – висота та ширина зображення, K – кількість каналів. Після оброблення зображення моделлю комп'ютерного зору отримується результат $R_{CV} = (c, q)$, де c – визначений моделлю клас, а $q \in [0,1]$ – вихідна оцінка впевненості щодо класу. Значення q не слід трактувати як імовірність правильності без окремої перевірки калібрування.

Для визначення областей зображення, які найбільше вплинули на прогноз, засобами ХАІ формується карта машинної значущості $S^M = \{m_{ij}\}$, $m_{ij} \in [0,1]$, де m_{ij} характеризує внесок відповідної області у сформований прогноз згідно з обраним ХАІ-методом.

Для того самого зображення за допомогою моделі людської візуальної уваги формується карта $S^H = \{h_{ij}\}$, $h_{ij} \in [0,1]$, де h_{ij} характеризує прогнозований рівень уваги людини до відповідної області. Для тверджень про людську увагу карту S^H необхідно валідовувати за емпіричною картою фіксацій E , отриманою в тому самому завданні перегляду.

Для зіставлення карти приводять до однакової просторової розмірності та нормалізують: $m'_{ij} = \frac{m_{ij}}{\sum_{u=1}^N \sum_{v=1}^W m_{uv}}$, $h'_{ij} = \frac{h_{ij}}{\sum_{u=1}^N \sum_{v=1}^W h_{uv}}$. Отримані карти M' і H' є невід'ємними та мають одиничну суму. Значенники мають бути більшими за нуль. Якщо ХАІ-карта містить від'ємні значення, спосіб їх перетворення потрібно визначити до експерименту.

Отже, S^M відображає значущість областей для конкретного прогнозу моделі, а S^H – прогноз людської уваги. Ці представлення мають різну семантику, тому їхній збіг інтерпретується лише як просторова узгодженість, а не як причинна тотожність способів прийняття рішення.

Багатокритеріальне оцінювання узгодженості

Для кількісного оцінювання введемо вектор взаємодоповнювальних метрик $z = \Phi(M', H', F)$, де F – множина емпіричних фіксацій, якщо вона доступна; Φ – процедура обчислення метрик відповідно до типу еталонних даних.

Коефіцієнт кореляції CC оцінює лінійну просторову залежність між двома неперервними картами: $CC = \frac{\sum_{i,j} (m'_{ij} - \mu_M)(h'_{ij} - \mu_H)}{NW \sigma_M \sigma_H}$, де μ_M , μ_H та σ_M , σ_H – відповідно середні значення і стандартні відхилення елементів нормованих карт. $CC \in [-1,1]$. Якщо дисперсія будь-якої карти дорівнює нулю, CC не визначається.

Для нормованих розподілів додатково використовуємо Similarity та дивергенцію Єнсена–Шеннона: $SIM = \sum_{i,j} \min(m'_{ij}, h'_{ij})$, $Q = \frac{M'+H'}{2}$, $JS = \frac{1}{2} KL(M' \parallel Q) + \frac{1}{2} KL(H' \parallel Q)$. Позначення Q використано, щоб не змішувати проміжний розподіл з інтегральним індексом узгодженості A .

SIM набуває значень від 0 до 1, де більше значення означає вищу подібність. Якщо в KL використано логарифм з основою 2, $JS \in [0,1]$, а оцінка подібності для агрегування дорівнює $z_{JS} = 1 - JS$.

За наявності емпіричних фіксацій F показник NSS для карти S визначається як $NSS(S, F) = \frac{1}{|F|} \sum_{(i,j) \in F} \frac{s_{ij} - \mu_S}{\sigma_S}$, де μ_S і σ_S – середнє значення та стандартне відхилення карти. Якщо $\sigma_S = 0$, NSS не визначається. Окремо можна обчислювати $NSS(M', F)$ і $NSS(H', F)$. AUC або $sAUC$ оцінюють здатність карти ранжувати фіксовані та нефіксовані області [7, 8].

Одна метрика не є достатньою: CC і SIM порівнюють неперервні карти, JS оцінює розбіжність розподілів, а NSS та AUC використовують дискретні фіксації. Тому висновок слід формувати за профілем метрик і перевіряти його стійкість до вибору способу нормалізації. Порівняльну характеристику використаних показників наведено в табл. 1.

Таблиця 1. Порівняльна характеристика метрик узгодженості

Метрика	Вхідні дані	Діапазон	Інтерпретація	Основне обмеження
CC	дві неперервні карти	$[-1; 1]$	лінійна просторова залежність	невизначена для сталої карти; чутлива до форми розподілу
SIM	два нормовані розподіли	$[0; 1]$	перетин маси розподілів; більше – краще	не враховує порядок фіксацій і причинність ознак
JS	два нормовані розподіли	$[0; 1]$	симетрична розбіжність; менше – краще	потребує узгодженої нормалізації та оброблення нульових значень
NSS	карта та емпіричні фіксації	$(-\infty; +\infty)$	середня стандартизована значущість у фіксаціях	залежить від дисперсії карти та якості eye-tracking даних
AUC/sAUC	карта, фіксації та негативні точки	$[0; 1]$	якість ранжування фіксованих і нефіксованих областей	нечутлива до абсолютних значень; результат залежить від вибору негативних точок

Висока узгодженість не є доказом правильності прогнозу. Модель і людина можуть зосереджуватися на тій самій області та водночас помилятися; навпаки, різні області можуть містити рівнозначні діагностичні ознаки.

Використання узгодженості як індикатора додаткової перевірки

Вихідна оцінка впевненості моделі q та індекс узгодженості відображають різні властивості. Перша характеризує вихід моделі, друга – подібність просторових представлень. Жодна з них окремо не є оцінкою ймовірності правильності рекомендації.

Для узагальнення часткових оцінок введемо інтегральний індекс узгодженості A . Нехай $K_{\text{дост}}$ – множина доступних і визначених метрик. Тоді $A = \frac{\sum_{k \in K_{\text{дост}}} \alpha_k z_k}{\sum_{k \in K_{\text{дост}}} \alpha_k}$, $\alpha_k \geq 0$, де $z_k \in [0, 1]$ – значення метрик, приведені до однакового діапазону та орієнтації «більше – краще», а ваги визначені на валідаційній вибірці. Для повного набору метрик сума ваг дорівнює 1. Індекс A є описовою характеристикою просторової узгодженості. Правило необхідності додаткової перевірки задається так: $V = \begin{cases} 1 & \text{якщо } q \geq T_q \wedge A < T_A \\ 0 & \text{в інших випадках} \end{cases}$, де T_q і T_A – пороги, визначені за результатами експериментальної калібрації.

Значення $V = 1$ вказує на суперечливий випадок: модель має високу вихідну впевненість, але її просторове пояснення недостатньо узгоджується з валідованим прогнозом людської уваги.

Запропоноване правило не передбачає автоматичного відхилення прогнозу та не підтверджує його помилковості. Воно лише ініціює додаткову перевірку користувачем.

Інформацію для користувача доцільно формувати як структурований результат $R = (c, q, A, V, E)$, де c – прогнозований клас; q – вихідна оцінка впевненості моделі; A – індекс узгодженості; V – індикатор потреби в додатковій перевірці; E – візуальне пояснення.

Алгоритм оцінювання узгодженості та протокол валідації

Реалізація запропонованого підходу передбачає виконання таких етапів.

Етап 1. Отримання вхідного зображення. Виконується завантаження та попереднє оброблення візуальних даних. Етап 2. Аналіз зображення моделлю комп'ютерного зору. Визначається клас об'єкта або події та вихідна оцінка впевненості q . Етап 3. Формування карти машинної значущості. Засобами ХАІ визначаються області зображення, що найбільше вплинули на результат моделі, та формується карта S^M . Етап 4. Формування прогнозованої карти людської уваги. Для того самого зображення та за відповідного завдання перегляду незалежно формується карта S^H . Етап 5. Валідація та підготовка карт. S^H перевіряється за емпіричними фіксаціями; карти приводяться до спільної розмірності, узгодженого діапазону й однакової області аналізу. Етап 6. Обчислення профілю метрик. Визначаються CC , SIM , JS та, за наявності фіксацій, NSS і $AUC/sAUC$; перевіряються сталі карти, пропущені дані й чутливість до нормалізації. Етап 7. Формування індексу узгодженості. Після експериментального визначення ваг обчислюється A ; до калібрації подається лише профіль часткових метрик. Етап 8. Перевірка суперечливого випадку. За попередньо валідованими порогоми визначається V . Етап 9. Формування інформації для користувача. Система подає прогноз, вихідну впевненість, ХАІ-пояснення, профіль узгодженості, обмеження оцінки та ознаку необхідності перевірки.

ХАІ-карта та прогнозована карта людської уваги мають формуватися незалежними гілками з того самого вхідного зображення; після валідації та нормалізації вони надходять до багатокритеріального блоку оцінювання. Послідовність і незалежність гілок показано на рис. 1.

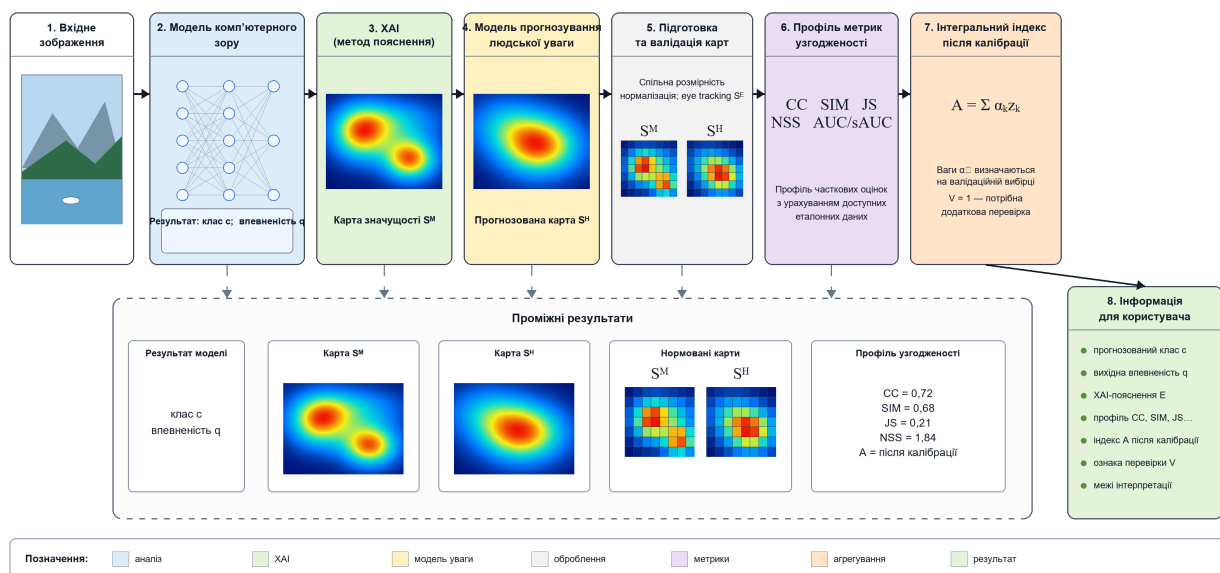


Рис. 1. Схема багатокритеріального оцінювання узгодженості машинної та прогнозованої людської візуальної уваги

Висновки та перспективи подальшого дослідження. У роботі сформульовано теоретико-методичний підхід до оцінювання просторової відповідності між областями, значущими для конкретного прогнозу моделі комп'ютерного зору, та валідованою картою прогнозованої людської візуальної уваги. Уточнено, що прогнозована моделлю карта не є безпосереднім вимірюванням людської уваги. Тому її використання потребує зіставлення з емпіричними eye-tracking даними, отриманими за узгодженого завдання перегляду. Запропоновано багатокритеріальний профіль із CC , SIM , JS та, за наявності фіксацій, NSS і $AUC/sAUC$. Наукова новизна полягає у контекстно залежній формалізації узгодженості з чітким розмежуванням просторової подібності, правильності прогнозу та каліброваної впевненості моделі.

Інтегральний індекс A запропоновано використовувати лише після експериментального визначення ваг і порогів. Він є додатковим індикатором для перевірки, а не доказом надійності або правильності рекомендації. Сформовано алгоритм незалежного отримання карт, їх валідації, нормалізації, обчислення профілю метрик, контролю граничних випадків і подання результату користувачеві.

Подальші дослідження мають охопити збір eye-tracking даних, перевірку міжспостерігачевої узгодженості, порівняння ХАІ-методів і saliency-моделей, калібрування q , визначення ваг та порогів

на окремій валідаційній вибірці, а також експериментальну оцінку впливу індикатора на точність і час прийняття рішень людиною.

Список бібліографічного опису

1. Hassija V., Chamola V., Mahapatra A. et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*. 2024. Vol. 16. P. 45-74. DOI: 10.1007/s12559-023-10179-8.
2. Kostopoulos G., Davrazos G., Kotsiantis S. Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. *Electronics*. 2024. Vol. 13, No. 14. Art. 2842. DOI: 10.3390/electronics13142842.
3. Morrison K., Shin D., Holstein K., Perer A. Evaluating the Impact of Human Explanation Strategies on Human-AI Visual Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*. 2023. Vol. 7, CSCW1. Art. 48. P. 1–37. DOI: 10.1145/3579481.
4. Yang Z., Mondal S., Ahn S. et al. Unifying Top-down and Bottom-up Scanpath Prediction Using Transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024. P. 1683–1693.
5. Liu G., Zhang J., Chan A. B., Hsiao J. H. Human Attention Guided Explainable Artificial Intelligence for Computer Vision Models. *Neural Networks*. 2024. Vol. 177. Art. 106392. DOI: 10.1016/j.neunet.2024.106392.
6. Coussemment K., Abedin M. Z., Kraus M., Maldonado S., Topuz K. Explainable AI for Enhanced Decision-Making. *Decision Support Systems*. 2024. Vol. 184. Art. 114276. DOI: 10.1016/j.dss.2024.114276.
7. Bylinskii Z., Judd T., Oliva A., Torralba A., Durand F. What Do Different Evaluation Metrics Tell Us About Saliency Models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019. Vol. 41, No. 3. P. 740–757. DOI: 10.1109/TPAMI.2018.2815601.
8. Kümmerer M., Wallis T. S. A., Bethge M. Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics. *Proceedings of the European Conference on Computer Vision*. 2018. P. 798-814.

References

1. Hassija, V., Chamola, V., Mahapatra, A., et al. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, 16, 45-74. <https://doi.org/10.1007/s12559-023-10179-8>.
2. Kostopoulos, G., Davrazos, G., & Kotsiantis, S. (2024). Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. *Electronics*, 13(14), 2842. <https://doi.org/10.3390/electronics13142842>.
3. Morrison, K., Shin, D., Holstein, K., & Perer, A. (2023). Evaluating the Impact of Human Explanation Strategies on Human-AI Visual Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), Article 48, 1–37. <https://doi.org/10.1145/3579481>.
4. Yang, Z., Mondal, S., Ahn, S., Xue, R., Zelinsky, G., Hoai, M., & Samaras, D. (2024). Unifying Top-down and Bottom-up Scanpath Prediction Using Transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1683–1693.
5. Liu, G., Zhang, J., Chan, A. B., & Hsiao, J. H. (2024). Human Attention Guided Explainable Artificial Intelligence for Computer Vision Models. *Neural Networks*, 177, 106392. <https://doi.org/10.1016/j.neunet.2024.106392>.
6. Coussemment, K., Abedin, M. Z., Kraus, M., Maldonado, S., & Topuz, K. (2024). Explainable AI for Enhanced Decision-Making. *Decision Support Systems*, 184, 114276. <https://doi.org/10.1016/j.dss.2024.114276>.
7. Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What Do Different Evaluation Metrics Tell Us About Saliency Models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3), 740–757. <https://doi.org/10.1109/TPAMI.2018.2815601>.
8. Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2018). Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics. *Proceedings of the European Conference on Computer Vision*, 798-814.

Історія статті:

Отримано: 19.05.2026 Доопрацьовано: 19.07.2026 Прийнято до друку: 23.09.2026 Опубліковано: 29.09.2026