

ON SCIENTIFIC RESEARCH OF HUMANS AND NEURAL NETWORKS

ПРО НАУКОВІ ДОСЛІДЖЕННЯ ЛЮДЕЙ ТА НЕЙРОННИХ МЕРЕЖ

prof. KUKLIN Volodymyr Mikhailovich

Prof Olena Shapovalova

ON SCIENTIFIC RESEARCH OF HUMANS AND NEURAL NETWORKS

ПРО НАУКОВІ ДОСЛІДЖЕННЯ ЛЮДЕЙ ТА НЕЙРОННИХ МЕРЕЖ

Volodymyr Kuklin[0000-0002-0310-1582], Olena Shapovalova[0000-0003-4566-6634]

Kharkiv National University of Economics,

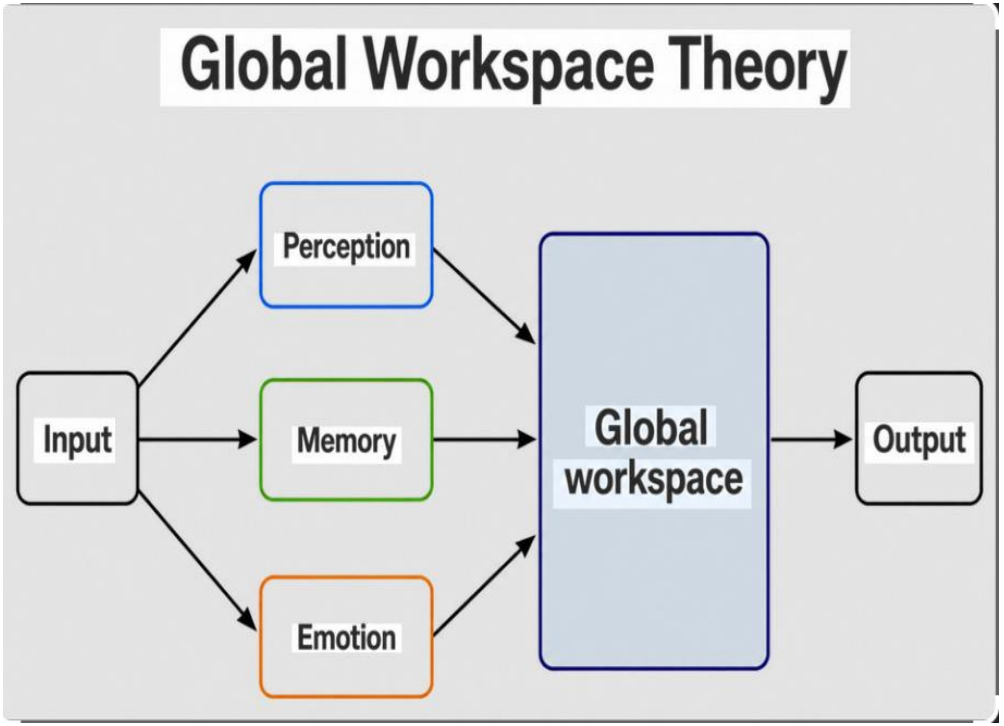
Department of Cybersecurity and Information Technologies,

Nauky Ave., 9A, 61001, Kharkiv, Ukraine

volodymyr.kuklin@hneu.net ; olena.shapovalova@hneu.net

GLOBAL WORKSPACE IN CEREBRAL CORTEX

Neurophysiologists have experimentally found that
Нейрофізіологи експериментально встановили, що
all thoughts and searches for answers posed in each
всі думки, пошуки відповідей на поставлені
problem, details of the history of successes and
завдання, а також подробиці історії успіхів і
failures are collected in a certain area of the cerebral
невдач зосереджуються в певній ділянці кори
cortex, which was called the Global Workspace - the
головного мозку, яку назвали «глобальним робочим простором»
main link of neurobiological theory of consciousness.
ключовою ланкою нейробіологічних теорій людської свідомості.
The work of the brain when solving a problem can take
Процес вирішення завдання мозком може відбуватися
place under the control of consciousness, and outside
як під контролем свідомості, так і поза ним.



GLOBAL WORKSPACE
IN CEREBRAL CORTEX

If all the knowledge and history of searching for

Якщо всі знання та історія пошуку

solutions is in the Workspace, this allows ones to

рішень зосереджені в робочому просторі, це дозволяє

constantly reflect on it, make amendments and new

постійно їх осмислювати, вносити корективи та додавати нові

details of the solution procedures. And do not leave

деталі до процедур пошуку рішення. Не полишаючи

the state of reflection, forming in your mind an

стану рефлексії, формуючи у своїй свідомості

environment - a vast puzzle of describing the elements

певне середовище — масштабну мозаїку елементів

of a still poorly defined future solution, where a

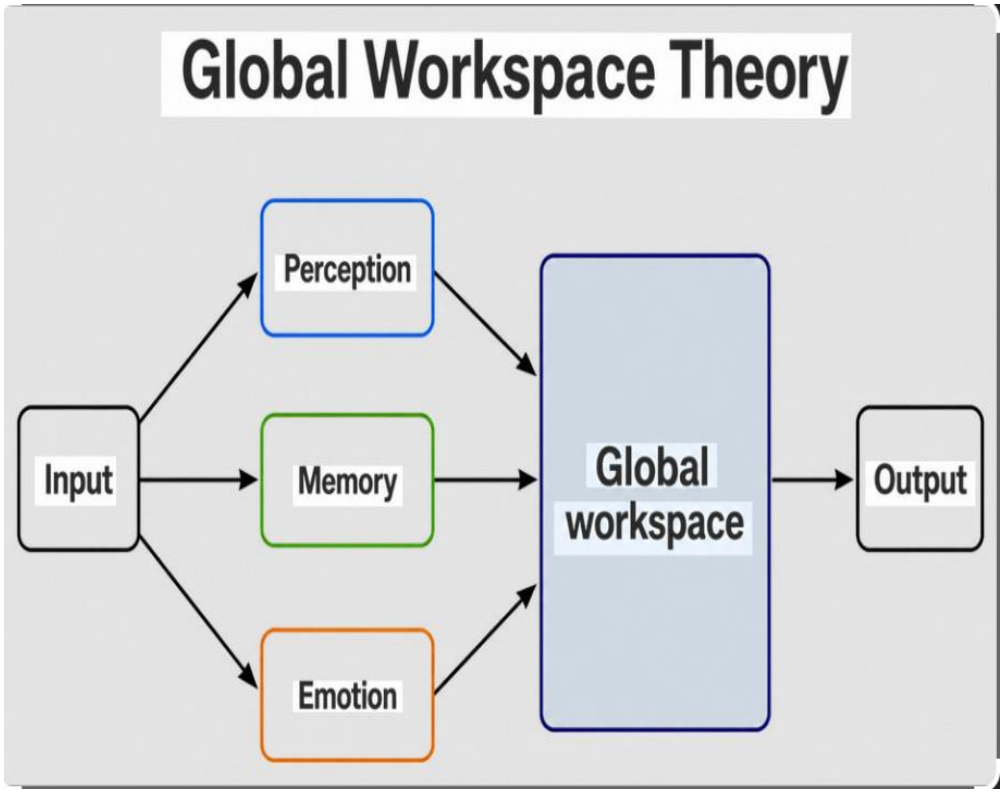
ще нечітко окресленого майбутнього рішення, — де

sophisticated mind is able to first see the goal and

проникливий розум здатен спочатку побачити мету,

then find a solution.

а вже потім знайти шлях до її досягнення.



DECISION-MAKING BY SCIENTISTS AND MODELS—LANGUAGE (LINGUISTIC) NEURAL NETWORKS

ПРИЙНЯТТЯ РІШЕНЬ НАУКОВЦЯМИ ТА МОДЕЛЯМИ — МОВНИМИ (ЛІНГВІСТИЧНИМИ) НЕЙРОННИМИ МЕРЕЖАМИ

It is shown that the accumulation of information,

Показано, що накопичення інформації

along with the history of successful and

а також історії успішних і

unsuccessful problem-solving attempts by both

невдалих спроб вирішення завдань як

humans and models, occurs in a specific **local**

людьми, так і моделями, відбувається в певній локальній

area of their intelligent system—the

ділянці їхньої інтелектуальної системи — Workspace. While in **humans**, semantic

«Робочому просторі». У людей семантичні

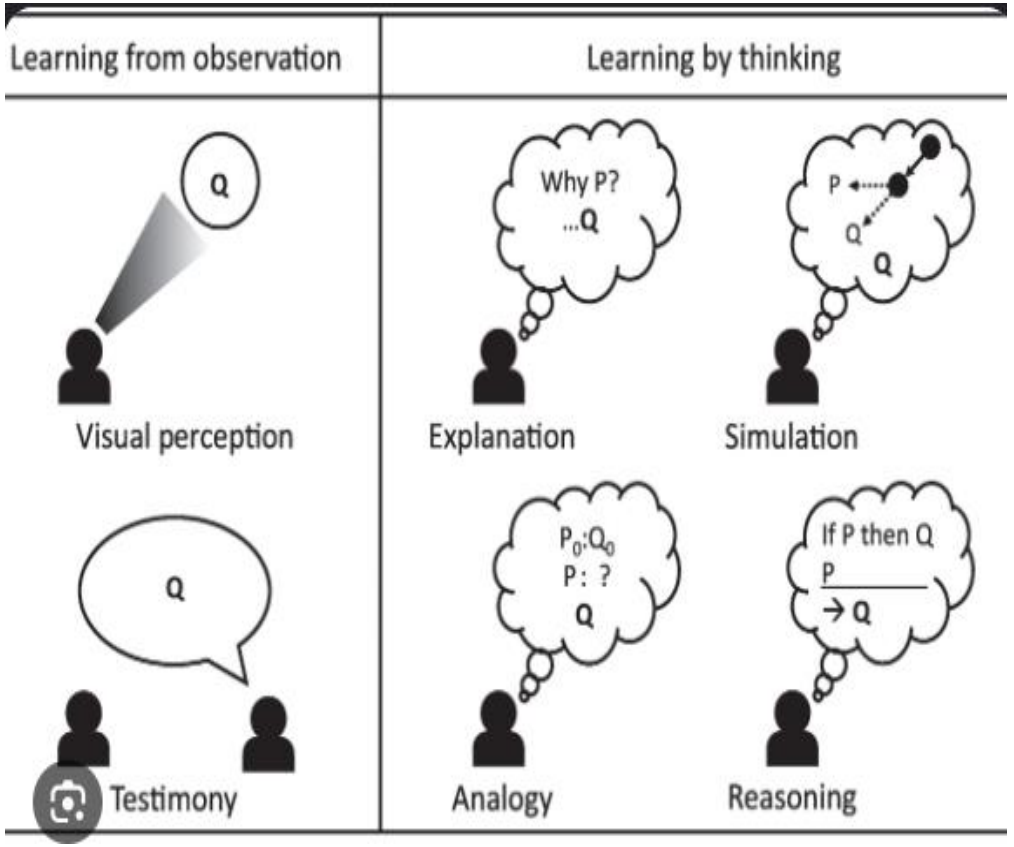
connections between elements of the obtained

зв'язки між елементами отриманого

solution and accumulated knowledge are **formed**

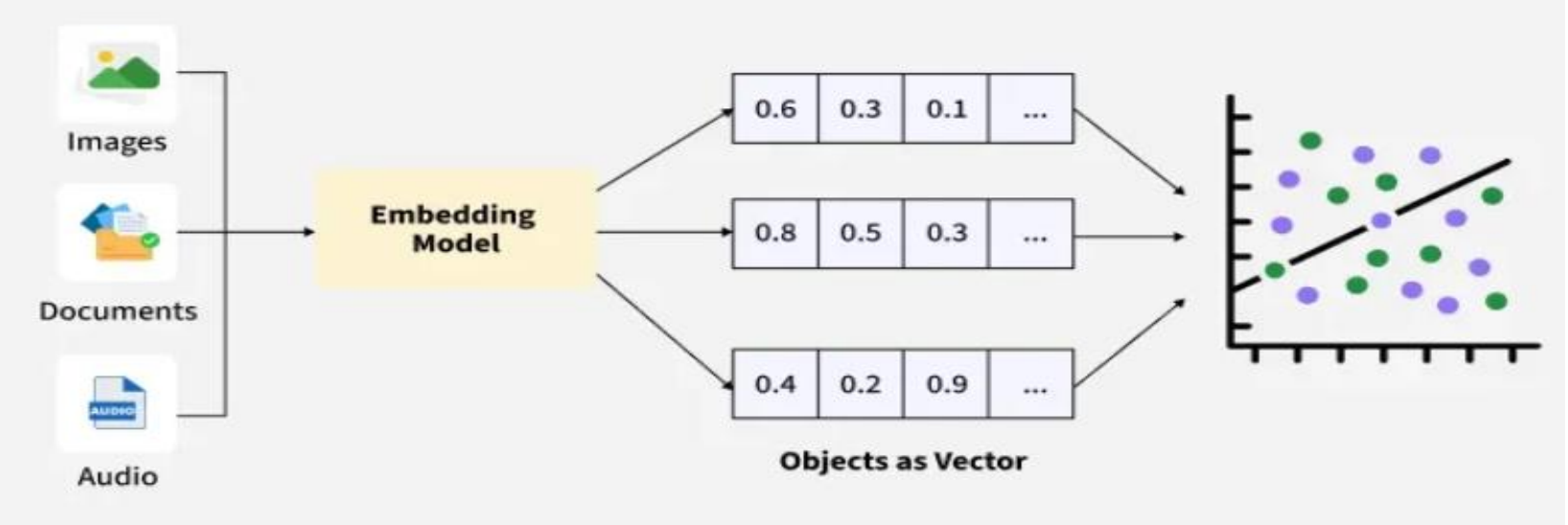
рішення та накопиченими знаннями формуються

through analogies, similarities, association…



DECISION-MAKING BY MODELS

ПРИЙНЯТТЯ РІШЕНЬ



The models use the proximity of comparison fragments

моделі використовують близькість фрагментів зібраного матеріалу

in vector space to form these connections and rely on

для порівняння у векторному просторі

maximizing the probability of the relationships.

та спираються на максимізацію ймовірності отриманих співвідношень.

When matches emerge, the resulting solution is

У разі виявлення збігів отриманий результат

identified with specific, previously studied phenomena.

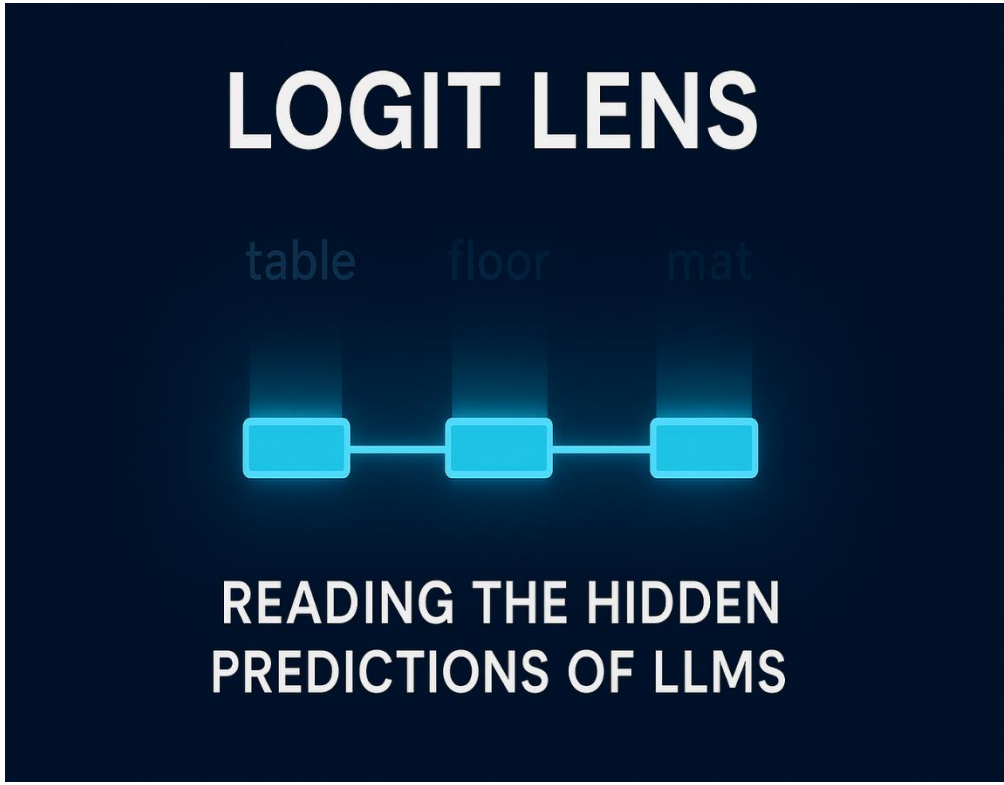
зіставляється з конкретними, раніше дослідженими явищами.

HOW THE WORKING SPACE OF THE MODELS WAS DISCOVERED

1.LOGIT LENS METHOD

ЯК БУЛО ДОСЛІДЖЕНО РОБОЧИЙ ПРОСТІР МОДЕЛЕЙ: 1. МЕТОД LOGIT LENS

Previously, Logit Lens methods only revealed
Раніше методи Logit Lens дозволяли виявити лише те,
how a specific keyword related to the text,
як саме ключове слово пов’язане з текстом,
thereby explaining the model's decision-making
що пояснювало механізм прийняття рішень моделлю.
mechanism. Developers then realized they could
механізм. Потім розробники зрозуміли, що можуть
identify a set of keywords that would lead them to
знайти набір ключових слів, які могли б привести їх
the text fragments—known as activations—
до фрагментів тексту — так званих активацій —
underlying the decision.
що лягли в основу для формування рішення.



HOW THE WORKING SPACE OF THE MODELS WAS DISCOVERED **J-SPACE** METHOD

ЯК БУЛО ВИЯВЛЕНО РОБОЧИЙ ПРОСТІР МОДЕЛЕЙ — МЕТОД J-ПРОСТОРУ

Identifying this set of

Виявлення цього набору

keywords and their associated activations (the

ключових слів і пов’язаних із ними активацій

building blocks of the decision) required

(майбутніх елементів рішення) вимагало

extending the search for "keyword-activation"

розширення пошуку «ключове слово — активація»

pairs across a wide range of contexts. This step

пар на значну кількість контекстів. Цей крок

proved crucial for understanding the process by

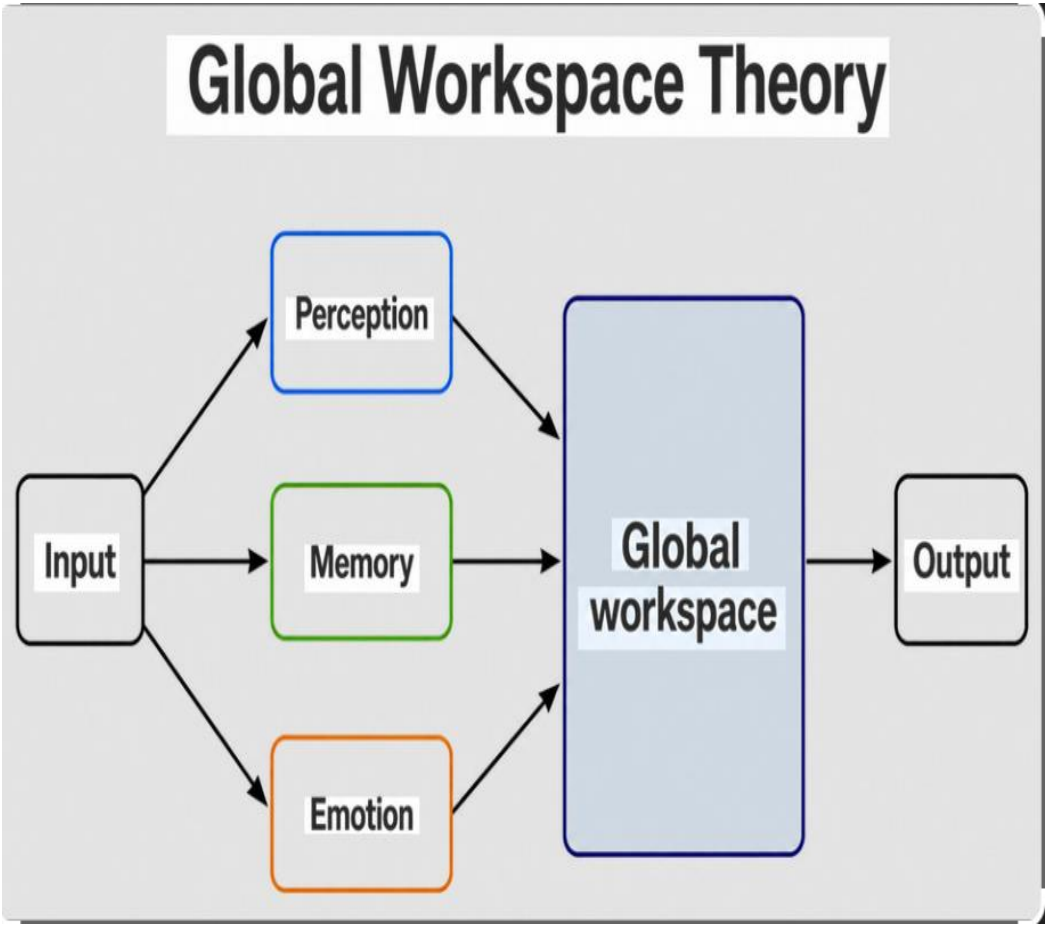
виявився вирішальним для розуміння процесу

which the model forms its conclusions.

формування висновків моделлю.

CLAUDE-5 MODEL CREATED THE J-SPACE

During training the Claude-5 model already
Під час навчання модель Claude-5 вже
created a space of such crucial internal patterns—
створила простір таких важливих внутрішніх моделей
—
the J-space. When asked where she was located,
J-простір. На запитання, де вона знаходиться,
ones received the answer she was in this space.
отримали відповідь, що вона знаходиться в J-простірі
Both humans and models use the knowledge
І люди, і моделі використовують знання
collected in such Workspace areas for solution
зібрані в робочому просторі, для вирішення
within the context of semantic manipulation.
в контексті семантичної маніпуляції
The solution search may include assortativity:
Пошук рішення може включати асортативність:
fragments are glued together based on their
фрагменти склеюються на основі їх
semantic similarity and maximizing the overall
семантичної подібності та максимізації загальної
probability of the resulting sentence based on the
words in the entire corpus.
ймовірності результуючого речення на основі слів у



CLAUDE-5 MODEL CREATED THE J-SPACE

Incidentally, intermediate steps also appear in the J-space, even if the model doesn't indicate this.

Проміжні кроки також з'являються в J-просторі.

J-space patterns are not involved in most of the model's functions—speech, mentioning simple facts, and using syntax.

Шаблони J-простору не задіяні в мовленні, згадуванні фактів і в використання синтаксису.

Models utilize opinions that reinforce the solution.

Як і люди, моделі використовують думки, які підкріплюють рішення.

LOGIT LENS AND J-LENS METHODS IN MODEL SPACE

In the resulting array of pairs, it became possible to identify semantic conscts that the model had collected and stored for similar cases, including solving a specific problem. This method was called the J-lens (Jacobi lens), which, it was hoped, and not without reason, would reveal a set of specific internal neural patterns.

For each word in the dictionary, the J-lens searches for and finds a pattern of activity that increases the likelihood that this word and the corresponding pattern will be frequently used and prove significant in constructing a solution.

On the other hand, a list of such meaningful words are clues to the current J-space content, which is easily extracted. This allows us to localize the J-space itself. Interestingly, introducing other different words and meanings into the relatively small J-space changes the responses, while making similar changes elsewhere in the network had no effect.

В отриманому масиві пар вдалося виявити семантичні конструкти, які модель накопичила та зберегла для схожих ситуацій, зокрема й для розв’язання конкретних задач. Цей метод отримав назву «J-лінза» (лінза Якобі); очікувалося — і небезпідставно, — що він дозволить розкрити низку специфічних внутрішніх нейронних патернів. Для кожного слова зі словника J-лінза знаходить патерн активності, який підвищує ймовірність того, що це слово (разом із відповідним патерном) часто використовуватиметься і відіграватиме важливу роль у формуванні розв’язку.З іншого боку, перелік таких значущих слів слугує підказкою щодо поточного вмісту J-простору, який можна легко видобути. Це дозволяє локалізувати сам J-простір. Цікаво, що введення інших слів і значень у відносно невеликий J-простір змінює реакції системи, тоді як аналогічні зміни в інших частинах мережі жодного ефекту не дають.

CONCLUSION

Then, of course, the conformity of the obtained result must be verified.

The principles of coordinating individual semantic fragments, their "gluing" together,

MAY WELL DIFFER IN HUMANS AND MODELS,

but

- THIS DIFFERENCE MAY NOT BE SO IMPORTANT FOR THE RESULT.
- Принципи узгодження окремих смислових фрагментів, їхнього «склеювання» між собою цілком можуть відрізнятися у людей і моделей,алеця відмінність може бути не такою вже й важливою для результату.

INFORMATION SOURCES

1.Abramov, G., Kuklin, V., Melnyk, I., Luntovskyy, A. (2025). Evolution and Sustainability of Neural Networks Theory. Conference paper/ First Online: 01 **October 2025** pp 217–253. In book: Luntovskyy, A., Klymash, M., Melnyk, I., Beshley, M., Gütter, D. (eds) Networks and Sustainability. TCSET 2024. Lecture Notes in Electrical Engineering, vol 1473. Springer, Cham. (Scopus) https://doi.org/10.1007/978-3-032-02272-1_10 ***https://link.springer.com/book/10.1007/978-3-032-02272-1?page=2#toc***

2.Abramov, G., Kuklin, V., Shapovalova, O., Melnyk, I. (2025). On the Use of Artificial Intelligence Systems for Scientific Search. First Online: 01 **October 2025**. pp 706–733. In book : Luntovskyy, A.,at.all. (eds) Networks and Sustainability. TCSET 2024. Lecture Notes in Electrical Engineering, vol. 1473. Springer, Cham. https://doi.org/10.1007/978-3-032-02272-1_32 ***https://link.springer.com/book/10.1007/978-3-032-02272-1?page=2#toc***

3.Kuklin V.M., Shapovalova O. O. Sirenka T. O. (2025). On systems and methods for solving problems. DOI: 10.30890/2709-2313.2025-41-07-028 P. 119-169 in book ‘Scientific thought development ‘ 2025 sge 41-07 (2025). Kuklin V.M., Osadchuk O.V., Novoselchuk N. et al. 282 p. https://

4.Kuklin V. On the evolution of neural networks» https://desymp.promonograph.org/index.php/sge/issue/archive DOI: 10.30890/2709-2313.2025-42-04-001 Number of Author's sheets: 3,38. In book: Kuklin V.M., Zhuravel I.M., Babchuk S., et al. The level of development of science and technology in the XXI century: Computer