ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 12 (2025) No. 11, pp. 146-165
DOI:10.33168/JLISS.2025.1110

NeuroSophic Cognition Model: A Hybrid Framework for Uncertainty and Cognitive Sovereignty

Aboul Ella Hassanien

Faculty of Computer and Artificial Intelligence, Cairo University, Cairo, Egypt.
Scientific Research School of Egypt (SRSEG) https://egyptscience-srge.com/
aboitcairo@cu.edu.eg

Abstract. This paper presents a novel pilot study within the NeuroSophic cognition that combines the Neutrosophic sets with Neurojico philosophy to handle uncertainty and advance cognitive sovereignty that maintains autonomous control over their own mental processes in AI-mediated systems. The NeuroSophic model can effectively and reasonably apply a multidimensional representation that combines the Neutrosophic truth (T), indeterminacy (I), and falsity (F) across five critical pillars cognitive dimensions, including perception (P), cognitive resistance (R), self-consciousness (E), cognitive sovereignty (S), and consciousness (C). Based on these five cognitive pillars, we established the mathematical foundations of the NeuroSophic model, starting with the NeuroSophic cognitive matrix (NCM) that represents the cognitive uncertainty, then presenting each NeuroSophic cognitive dimension that belongs to the NeuroSophic cognitive matrix. The global cognition score (GNS) and the cognitive sovereignty index (CSI) are two new indices presented, focusing on their meaningful implication. Also, aggregation, expected value, and distance metrics, which are basic mathematical operations, are presented. The NeuroSophic model is expanded by introducing a new NeuroSophic probability measure to present uncertainty in a more refined approach. The NeuroSophic cognition model opens pathways for future research, so further research might explore the adaptive weighting, probabilistic risk modeling, and ethical design.

Keywords: Neutrosophic theory — NeuroSophic model — Neurojico Philosophy — Cognitive Indices — AI-mediated systems.

1. Introduction

Cognitive AI refers to artificial intelligence that emulates and simulates human thinking, including incorporating complexity and uncertainty into human decision-making by learning and reasoning from data, then adapting to new knowledge, and refining its process to problem-solving (Zhong, 2006, Gonzalez & Heidari, 2025, George et al. 2022). Traditional cognitive models are built to explain human cognition that is a rule-based structure and sequential processing, which makes them less adaptable and more rigid in dynamic environments.

In contrast, AI-based cognitive models are developed based on machine and deep learning models, focusing on simulating or improving cognition rather than describing it (Fintz et al. 2022). AI-based cognitive models use data-driven probability theory with distributed and parallel processing with high adaptability and the ability to learn from huge datasets (Zhong, 2006). Neutrosophic theory, despite handling uncertainty and indeterminacy in cognition, fails to accurately represent the complex characteristics of dynamic environments' difficulty in understanding output, where indeterminacy and falsity are working simultaneously (Smarandache, 1998; Majumdar, 2015, Das, 2020). This limitation and situation require an innovative strategy that addresses uncertainty and understanding the output in dynamic environments while preserving cognitive autonomy in AI-mediated environments.

The NeuroSophic cognition model introduced in this paper provides a comprehensive and new study that combines Neutrosophic theory (Smarandache, 1998) and Neurojico philosophy (Hassanien, 2025), to modeling AI-based cognition amid uncertainty as well as interpretation of the complex system in a dynamic environment (Franciskus, 2024). The new study enables a comprehensive examination of the five Pillars mentioned before by illustrating cognitive states as triplets of neutrosophic truth (T), indeterminacy (I), and falsity (F). The proposed NeuroSophic cognition allows truth values to be behind the standard range and is enabling the simulation of algorithmic opacity, perceptual distortion, and the losing of sovereignty (Atkinson, 2025).

This paper is organized as follows. Section (2) introduces the mathematical system of the NeuroSophic model. Section (3) presents a numerical example demonstrating the quantification of cognitive states through Neutrosophic principles within the Neurojico philosophical framework Section (4) discusses how weights affect cognitive indices. NeuroSophic sets as a formal framework for cognitive sovereignty supported by a numerical example are discussed in Section (5). Section (6) presents the concept of NeuroSophic Probability, which assesses the likelihood of an event or decision, and is followed by a worked numerical example. The basic three operators, including the aggregation operator, expected value, and distance measure, that the NeuroSophic framework uses to measure and understand cognitive measures are explained in Section (7). Section (8) discusses the reliability and integrity assessment in the NeuroSophic framework. Conclusion and future trends are presented in Section (9).

2. Mathematical Foundation

The NeuroSophic model depicts cognition as a multidimensional system, covering various cognitive dimensions such as enhanced perception (P), cognitive sovereignty (S), cognitive resistance (R), selfhood (E), and consciousness (C), articulated via a Neutrosophic representation. Truth (T), indeterminacy (I), and falsity (F) are the three independent degrees' representations that are assigned to capture the inherent uncertainty of cognitive states in AI-mediated environments (Songlin and Zhang, 2023). Unlike classical or fuzzy logic (Gupta, 2011, Umberto 2008), which only allows for simple modeling of ambiguity, these elements can add up to more than one.

Formally, each NeuroSophic cognitive dimension d belong to the NeuroSophic cognitive matrix (M) is defined as:

$$d = \{T_d, I_d, F_d\}, \quad 0^- \leq T_d, I_d, F_d \leq 1^+ \tag{1}$$

where 0^- and 1^+ indicate lengthy durations to facilitate over- and under-defined states, according to Neutrosophic logic (Umberto 2008).

The algorithmic opacity (Sylvia, 2025) refers to the lack of transparency in making decisions and interrupts the obtained output by the machines. Thus, integrating the algorithmic opacity into the cognitive dynamic environment raises critical concerns about algorithmic opacity (Sylvia,2025), which limits interpretability, accountability, and ethical challenges in decision-making processes (McKinlay, 2020). This concern often correlates with perceptual distortion, where mediated information reshapes human cognition, leading to biased or broken understanding (Gkanatsiou et al. 2025).

In addition, these dynamic environments are accelerating the sovereignty decline (Slawner, 1996), as individual and collective autonomy become progressively dependent on algorithmic governance, which makes the basic ideas of agency and self-determination more challenging (Badawy, 2025; Pop & Pierson, 2025). Addressing these issues requires establishing new theories and philosophies that support transparency, ethical review, and collaborative governance to reduce structural risks and enhance representative resilience.

Each Neutrosophic cognitive triple dimension shows how uncertain a cognitive dimension is. For example, enhanced perception integrates high truth with moderate indeterminacy, which means that the perception is more robust but less clear. For cognitive resistance (R), it has a huge uncertainty, which means people are actively asking questions or the system is unclear. On the other hand, selfhood (E) indicates that misrepresentation is increased, and identity distortion is impacted by algorithms. While consciousness shows a balance between truth and indeterminacy, showing how adaptive consciousness works in situations where events are not clear settings. Cognitive sovereignty (S) combines truth, falsehood, and indeterminacy in a fluid stability, which shows how autonomous thinking and algorithmic governance are at likelihood with each other.

To represent the entire cognitive system, we develop the following NeuroSophic cognitive matrix (***NCM***) and it defined as presented in Equation (2).

$$NCM = \begin{pmatrix} T_C & T_P & T_E & T_R & T_S \\ I_C & I_P & I_E & I_R & I_S \\ F_C & F_P & F_E & F_R & F_S \end{pmatrix} \tag{2}$$

The NeuroSophic cognitive matrix encodes the uncertainty distribution across all dimensions, enabling both static analysis and dynamic simulation of cognitive states. Based on this matrix, two main indices are introduced. The first one is the cognitive sovereignty index (***CSI***) and the NeuroSophic global cognition score (***GNS***).

The cognitive sovereignty index (***CSI***) measures the independence in the aspect of uncertainty. The cognitive sovereignty index is defined by Equation (3).

$$\boldsymbol{CSI} = w_T \cdot T_s + w_I \cdot (1 - I_s) - w_F \cdot F_s \quad (3)$$

Where w_T, w_I, and w_F are weights that shows how important T,I, and F in AI-mediated context.

A higher cognitive sovereignty index indicates more robust sovereignty, where truth is the most important element, indeterminacy is low, and falsity is very low. A low cognitive sovereignty index indicates a loss of independence, which is often generated by algorithmic control or a leaning on outside systems. This index is essential for assessing ethical AI systems and identifying sovereignty scale in the digital governance or education sector (Umoke et al., 2025).

The second index is the NeuroSophic global cognition score (GNS), which is defined as a sum of all cognitive dimensions and is defined by Equation (4).

$$\boldsymbol{GNS} = \sum_{d \in D} \alpha_d \cdot (T_d - F_d) - \beta_d \cdot I_d \quad (4)$$

where α_d, and β_d are dimension-specific weights that allows customization for different contexts such as education, law, or governance.

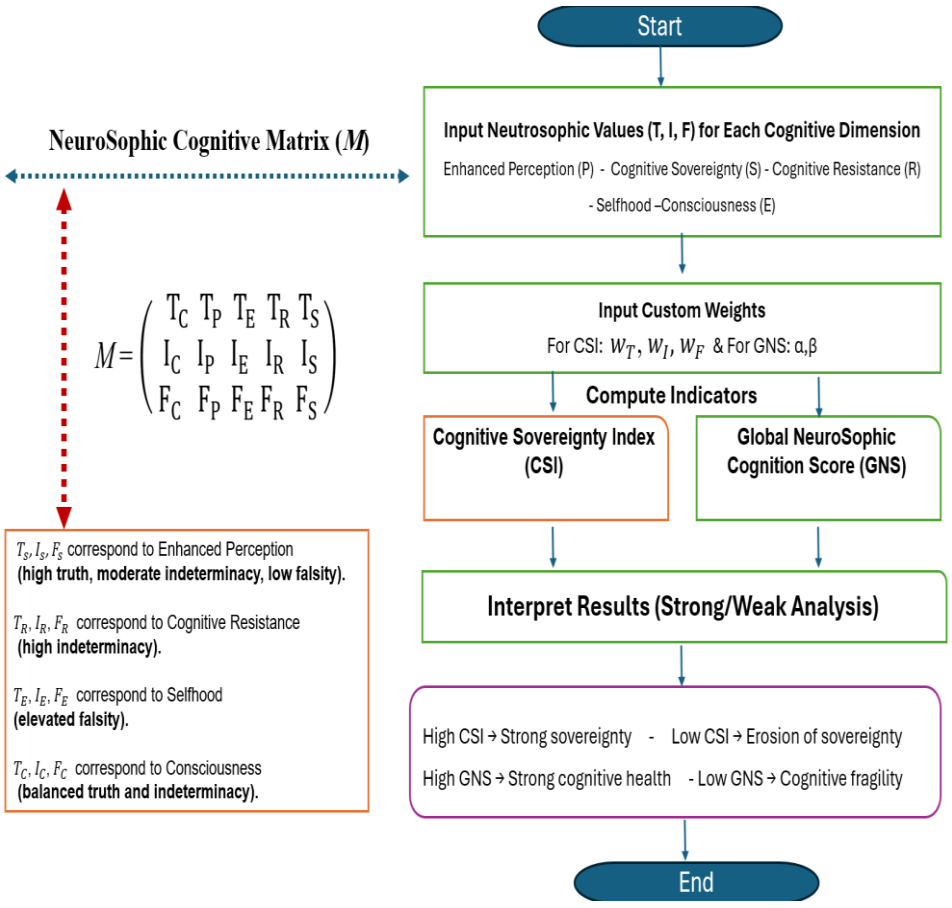

Fig.1: main calculation of the CSI and GNS indices

The NeuroSophic global cognition score (***GNS***) gives a complete view of cognitive context across all Neurojico cognitive pillars. We should note that a higher value of the NeuroSophic global cognition score (***GNS***) means cognition is stable, perception is accurate, and resistance to operation is more robust. A low NeuroSophic global cognition score indicates weakness, expressing an inadequacy in truth, along with falsehood and indeterminacy controls, indicating cognitive vulnerability.

We can use this score to compare between systems or situations (for example, governance vs. law settings).

These mathematical foundations make it possible to simulate how algorithmic interventions change the balance of truth, indeterminacy, and falsehood across cognitive dimensions that model cognitive dynamics in the face of uncertainty. Figure 1 shows the flowchart shows how the matrix encodes values for truth, indeterminacy, and falsity for each cognitive dimension, which are used to figure out ***CSI*** and ***GNS***.

Figure 1 shows the NeuroSophic cognitive matrix (M) and its part in the whole process of figuring out cognitive indicators. The matrix is shown on the left as a 5×3 structure. The rows stand for Truth (T), Indeterminacy (I), and Falsity (F), and the columns stand for the five cognitive dimensions: enhanced perception (P), cognitive resistance (R), selfhood (E), cognitive sovereignty (S), and consciousness (C). Each cell in the matrix holds the uncertainty profile for a certain dimension. Based on the flowchart, we first enter the NeuroSophic values (T, I, and F) for each dimension. Then, you enter custom weights for the ***CSI*** and ***GNS*** calculations. The procedure then figures out two important numbers, the NeuroSophic global cognition score (***GNS***) and the cognitive sovereignty index (***CSI***). Then, we will interpret the results to see how strong or weak sovereignty and cognitive health are, for example. A high ***CSI*** shows that an event is very independent, while a low ***CSI*** shows that it is losing its independence. Low ***GNS*** means that your mind is weak, while high ***GNS*** means that your mind is strong.

3. NeuroSophic Cognitive Matrix: A Numerical example of Cognitive Uncertainty

This section presents a numerical example demonstrating the quantification of cognitive states through Neutrosophic principles within the Neurojico philosophical framework, illustrating the practical application of the NeuroSophic system. Next, apply these NeuroSophic principles by assigning each of the five core components of cognition to the three values that represent truth, indeterminacy, and falsity.

To make the NeuroSophic cognitive matrix, which shows the cognitive system's uncertainty profile, each dimension is first given a Neutrosophic value. Then, there are two more steps to calculate the ***SCI*** and ***GNS***, which are used to measure how much autonomy is kept when faced with falsehood and uncertainty.

Step (1): Make the NeuroSophic Cognitive Matrix

Table 1 shows an example of the cognitive Neutrosophic dimensions and their related uncertainty profiles, which are shown through the three main parts of Neutrosophic (T, I, F). Perception, sovereignty, resistance, selfhood, and consciousness each show a different way that these values are spread out. This shows how truth, indeterminacy, and falsity work together in AI-mediated cognitive environments. This representation offers a quantitative basis for the analysis of cognitive states when there is algorithmic influence and uncertainty.

Table 1: the cognitive NeuroSophic matrix

Dimension	Truth (T)	Indeterminacy (I)	Falsity (F)
Perception	0.80	0.20	0.10
Sovereignty	0.70	0.30	0.20
Resistance	0.60	0.40	0.30
Selfhood	0.75	0.25	0.15
Consciousness	0.85	0.15	0.05

Step (2): Compute CSI

Using the weights $w_T = 0.5, w_I = 0.3$ and $w_F = 0.2$, the cognitive sovereignty index derived from Table (1) is calculated based on Equation (3) as follows.

$$\boldsymbol{CSI} = (0.5 * 0.7) + 0.3 * (1 - 0.3) - 0.2 * 0.2 = 0.52$$

A ***CSI*** value of 0.52 means that cognitive sovereignty is at a moderate level. This means that even though the truth part is strong, the fact that there is uncertainty and falsehood make the whole thing less autonomous the process of making decisions.

Step (3): Compute GNS

Assuming α_d=1, β_d=0.5 for all dimensions, and applying Equation (4), the Global NeuroSophic Score (***GNS***) is computed as follows for each dimension.

P: (0.8 - 0.1) - 0.5 * (0.2) =0.7 - 0.1=0.6

S: (0.7 - 0.2) - 0.5 * (0.3) =0.5 - 0.15=0.35

R: (0.6 - 0.3) - 0.5 * (0.4) =0.3-0.2=0.1

E: (0.75 - 0.15) - 0.5 * (0.25) =0.6-0.125=0.475

C: (0.85 - 0.05) - 0.5 * (0.15) =0.

The NeuroSophic global score (***GNS***) is then estimated as

GNS=0.6+0.35+0.1+0.475+0.725=2.25

A ***GNS*** of 2.25, which is in a range of about 0 to 5, shows that the person's overall cognitive health is good with perception and consciousness being the most important parts of the score.

4. How Weights effect Cognitive Indices

Weights determine how sensitive each index is to truth, indeterminacy, and falsehood for cognitive sovereignty index (***CSI***). We can stress Truth (clarity), correct Indeterminacy (uncertainty), or make Falsity (distortion) more powerful by these changes. This means that the index is based on three

weighted parts for (T, I, and F). The cognitive sovereignty index will increase significantly if the truth value is high, the weight for truth (w_T) will increase rapidly, which will grow the impact of truth on the processing. Increasing the weight for indeterminacy (w_F) supports situations with lower uncertainty; consequently, the cognitive sovereignty index will increase when uncertainty is low. In addition, Increasing the weight for indeterminacy (w_I) tends to benefit situations with low uncertainty; therefore, the cognitive sovereignty index will increase when uncertainty is minimal**.**

Increasing the weight for falsehood (w_F) makes lying more correctly, so if falsifying is high, the cognitive sovereignty index will be smaller. This dynamic context allows the NeuroSophic model to highlight independence and transparency, or in contrast, emphasize how easily elements can be distorted depending on how the weights are customized.

The NeuroSophic global cognition score (***GNS***) considers the difference between what is true and what is false modified for indeterminacy, to combine all cognitive dimensions. Cognitive dimensions with a lot of truth and not much falsehood will greatly improve the overall score because raising the weight for truth-falsity difference (α) makes truth is more powerful and lies are more dangerous. In addition, raising the weight for indeterminacy (β) makes uncertainty more punishing by lowering the score when there is ambiguity. This means that by changing α and β, the model can either focus on strength and clarity or on being open to not sure. A high ***GNS*** shows that cognitive health is good in all areas, while a low *GNS* shows that weakness and the ability to be controlled.

Table 2: Effect of Weight Parameters on ***CSI*** and ***GNS***

Index	**Weight Parameter**	**Effect of Increasing Weight**	**Interpretation Impact**
CSI	w_T (Truth)	**CSI** increases when truth is high	Sovereignty appears more robust
	w_I (Indeterminacy)	**CSI** increases when indeterminacy is low	Rewards clarity, penalizes ambiguity
	w_F (Falsity)	**CSI** decreases sharply when falsity is high	More sensitive to manipulation
GNS	α (Truth-Falsity)	**GNS** increases when truth dominates falsity	Emphasizes truth dominance
	β (Indeterminacy)	**GNS** decreases when indeterminacy is high	Makes uncertainty more damaging

CSI and ***GNS*** are very essential for understanding cognitive resilience and vulnerability. Every weight parameter changes the index's sensitivity to T, I, and F. The Neutrosophic cognitive model can either stress clarity and independence or draw attention to uncertainty and distortion, when changing its weights.

Table 2 illustrates how increasing each weight that affects the explanation. It shows behavior of changing the ***CSI*** and ***GNS*** values when the weights are set up in different ways.

When the truth is the most important event, both ***CSI*** and ***GNS*** go up so much that it shows clarity. In contrast, when falsity is heavily corrected, ***CSI*** goes down because the cognitive model becomes more sensitive to distortion. On the other hand, ***GNS*** stays moderate because truth still matters. Additionally, when the configuration setting uses high weight on indeterminacy, then the ***GNS*** will drop sharply with

a penalty for uncertainty. This issue makes the system very sensitive to ambiguity as well. The default setting gives balanced scores without going too far in sensitivity, and it provides an unbiased understanding of cognitive resilience.

5. NeuroSophic Sets: A Formal Framework for Cognitive Sovereignty

Within the realm of Neurojico, a NeuroSophic set is a distinct variant of a Neutrosophic set (Smarandache, F. 1998) that embodies cognitive states, algorithmic mediation, and neurocognitive sovereignty. This section will present the basic formal NeuroSophic set, followed by a numerical example that trace the definition.

Definition (1): NeuroSophic Sets

Let U be a universal set of cognitive elements (e.g., perceptions, decisions, actions). A NeuroSophic set N_s is characterized as:

$$N_s = \{\langle x: T_c(x), I_a(x), F_m(x)\rangle \mid x \in U\}\} \quad (5)$$

Where:

- $T_c(x)$= Cognitive Truth, which means how real a perception or decision.
- $I_a(x)$ = Algorithmic Indeterminacy, which means that AI mediation/incomplete data can make elements less specific.
- $F_m(x)$ = Manipulation/Falsity, which means how much outside systems distort or specific.

Each part is between 0 and 1, and they don't depend on each other.

Worked numerical example: AI-Driven Medical Diagnosis

An AI-based medical system (Alowais et al. 2023) gives a patient a diagnosis decision. To evaluate the cognitive soundness of this decision within the NeuroSophic framework, we depict the diagnosis as a component:

x = "Diagnosis: Diabetes Type II"

The NeuroSophic components that go with this are defined as follows:

- Cognitive Truth $T_c(x)$ shows how real something is based on clinical evidence and lab results. For this case, $T_c(x)$ is set to 0.80
- Algorithmic Indeterminacy $I_a(x)$ shows how uncertain things are because of missing patient history or gaps in contextual data. In this case $I_a(x)$ is set to 0.15.
- Manipulation/Falsity $F_m(x)$ Shows the chance of bias from the AI model's training data or systemic distortions. $F_m(x)$ is set to 0.05.

Thus, the NeuroSophic representation of the diagnosis is:

$N_s=\langle T_c(x), I_a(x), F_m(x)\rangle$ = (0.80,0.15,0.05)

This example shows how AI-generated medical diagnosis were reliable and uncertain. The diagnosis of "**Diabetes Type II**" is depicted as a NeuroSophic element encompassing three dimensions (cognitive truth, algorithmic indeterminacy, and manipulation/falsity).

A truth cognitive value of 0.80 means that the diagnosis is very real because it is supported by powerful clinical and lab evidence. The algorithmic indeterminacy value of 0.15 shows that there is some uncertainty because the patient's history is incomplete or insufficient information about the case, which could make the decision less accurate. Lastly, handling the falsity value of 0.05 indicates that there is only a smaller likelihood that the AI model's data training or systemic distortions will cause bias. When integrated, these values that represent the NeuroSophic representation, like N_s = (0.80, 0.15, 0.05), this will help healthcare system to assess algorithm reliability and stimulate transparency in medical AI.

6. NeuroSophic cognitive probability ($P_N(E)$)

$P_N(E)$ is a specialized form of probability adapted to fit the Neurojico theory. It evaluates the likelihood of an event or decision by considering three distinct Neutrosophic parameters:

$$P_N(E) = \langle T_C(E), I_a(E), F_m(E) \rangle \quad (6)$$

Where:

- $T_C(E)$ [Cognitive Truth Probability]. extent to which the event is accurate or valid based on proof.
- $I_a(E)$ [Algorithmic Indeterminacy Probability]. The level of uncertainty introduced by data that isn't complete or algorithms that aren't clear.
- $F_m(E)$ [Manipulation/Falsity Probability]. The level of bias or distortion that is affecting the event.

Classical probability depicts certainty as a singular numeric value, presuming a binary or ambiguous interpretation of truth (Loucks and Beek, 2017). In AI-mediated cognition (Songlin and Zhang, 2023), however, decisions are affected by many factors, such as truth, uncertainty, and falsity that exist at the same time in the same situation. NeuroSophic Probability tackles this complexity by measuring each dimension separately, which makes it a clearer and more complete depiction of probabilistic reasoning in algorithmic contexts.

Every part of NeuroSophic Probability is in the range [0,1] and works on its own:

$$0 \leq T_C(E), I_a(E), F_m(E) \leq 1 \quad (7)$$

Worked example.

Let's say we want to see how likely it is that an AI-powered recommendation on an educational platform is correct and fair.

Step 1: Give Values

To create a basic model of cognitive uncertainty, we use the baseline values shown in Table 3.

Table 3: The baseline values outline the distribution of (T, I, F) across five principal cognitive dimensions within the NeuroSophic structure.

Cognitive dimension	$T_C(E)$	$I_a(E)$	$F_m(E)$
Resistance	0.60	0.40	0.30
Selfhood	0.75	0.25	0.15
Perception	0.78	0.22	0.10
Sovereignty	0.70	0.30	0.20
Consciousness	0.85	0.15	0.05

Ass shown in Table (3), the assigned values describe the Truth (T), Indeterminacy (I), and Falsity (F) values across five principal cognitive dimensions (perception, sovereignty, resistance, selfhood, and consciousness) within the NeuroSophic. For instance, perception has a high truth value (0.80), a balanced indeterminacy (0.20), and a low falsity (0.10), which means that perception is more robust but not completely clear.

Sovereignty dimension as shown in Table (3) illustrate a truth, indeterminacy, and falsity levels of (0.70, 0.30, 0.20), respectively. These values show the conflict between Self-governance and AI-mediated control. Resistance has a less clear profile with truth and uncertainty at 0.60 and 0.40, which means that people are actively asking questions or elements when the system is unclear.

Selfhood dimension keeps truth and falsehood at (0.75, 0.15), which shows that there is a risk of identity distortion.

Lastly, consciousness dimension shows the most truth and lowest falsehood at values (0.85, 0.05), respectively, which illustrate how adaptable awareness is in situations where uncertainty exists. These benchmark values are used to model cognitive states and calculate advanced metrics like the NeuroSophic cognitive probability ($P_N(E)$) and the cognitive sovereignty index (CSI).

Step (2) Compute $\mathbf{P_N(E)}$

From Equation (5) and the values given in Table (3). The $\mathbf{P_N(E)}$ is calculated as follows:

$\mathbf{P_N(E)} = \langle 0.78, 0.22, 0.10 \rangle$

Step (3): Explanation and Interpretation

The interpretation of the obtained value of $P_N(E)$ is explained as illustrated in Figure 2. The cognitive truth probability value at 0.78, shows there is a high level of confidence in validity. Algorithmic indeterminacy probability at 0.22 shows there is balanced uncertainty due to unclear algorithmic logic. While Manipulation/Falsity probability at 0.10, shows there is a low risk of falsity. This representation shows hidden uncertainty and bias dimensions, which makes decision-making clearer. Different from classical probability, the sum of these cognitive probabilities does not equal 1, which allows for flexible modeling of situations where (T, I, and F) overlapping.

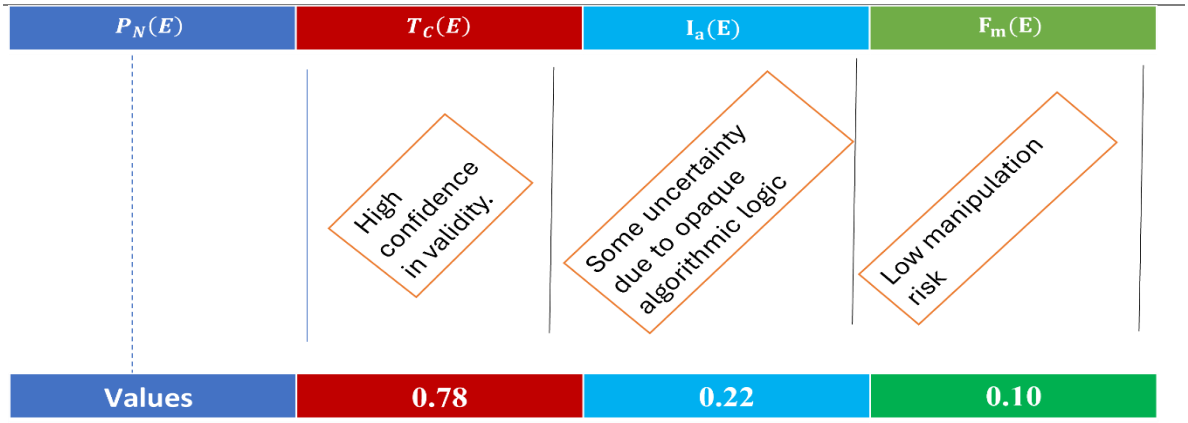

Fig.2: interpretation of the $\mathbf{P_N(E)}$

7. Mathematical Operators

This section presents the basic operators that the NeuroSophic framework uses to measure and understand cognitive parameters. These operators are the aggregation operator, which puts together truth, and the uncertainty and expected value operator, which combines these parts into one decision score. In addition, the distance measure will be presented with an illustration of its application through the concept of the distance matrix that enables comparative analysis between cognitive states.

Definition (2): Aggregation Operator

The aggregation operator is a math function that combines the three main parts of a NeuroSophic set—Truth (T), Indeterminacy (I), and Falsity (F)—into one scalar score that shows how confident someone is in their ability to make decisions. To figure out an aggregation operator for making decisions. We set the following score that represent the aggregation operator.

$$\text{Score(E)} = \alpha * \text{T}_\text{C}(\text{E}) - \beta * \text{I}_\text{a}(\text{E}) - \gamma * \text{I}_\text{a}(\text{E}) \quad (8)$$

In this case, α, β, γ are weights that show how important truth, indeterminacy, and falsehood in the context.

Numerical example: By applying equation (8) on Table (3) with α, β, γ = (0.6, 0.3, 0.1). Then using the score. Table 4 shows the aggregation operator score for each dimension.

$$\text{Score(E)} = 0.6 * \text{T}_\text{C}(\text{E}) - 0.3 * \text{I}_\text{a}(\text{E}) - 0.1 * \text{I}_\text{a}(\text{E})$$

Table 4: Aggregation Operator score

Cognitive Dimension	Score(E)
Perception	0.416
Sovereignty	0.330
Resistance	0.230
Selfhood	0.380
Consciousness	0.480

As we can see from Table (4), consciousness has the highest score of 0.480, which means it is very true and very little false. The lowest score for resistance is 0.230, which shows that there is more false

information and uncertainty, which makes people less confident. These scores help us rank cognitive dimensions or set limits for making decisions.

Definition (3): Expected Value

For a set of events $\{E_1, E_2, \dots, E_n\}$, the NeuroSophic expected value (EVN) is defined as:

$$\boldsymbol{EVN} = \frac{1}{n}\sum_{i=1}^{n}(T_C(E),\ I_a(E), F_a(E)) \quad (9)$$

This operator computes the average vector of the three components (T, I, F). Where

$$\boldsymbol{EVN}=(\overline{T}, \overline{I}, \overline{F}) = \left(\sum\frac{T_C(E_i)}{n}, \sum\frac{I_a(E_i)}{n}, \sum\frac{F_m(E_i)}{n}\right)$$

Worked Example: Using Table (3), we find the average vector of the three parts (T, I, F) as follows.

$$\overline{T} = \frac{0.78+0.70+0.60+0.75+0.85}{5} = 0.736$$

$$\overline{I} = \frac{0.22+0.30+0.40+0.25+0.155}{5} = 0.264$$

$$\overline{F} = \frac{0.10 + 0.20 + 0.30 + 0.15 + 0.05}{5} = 0.160$$

Then, the ***EVN*** is estimated as

EVN= (0.736,0.264,0.160)

The average truth level for all dimensions is 0.736, which shows a high level of confidence. Indeterminacy averages is 0.264, which means there is some uncertainty, and falsity averages 0.160, which means there is a low risk of not right.

Definition (4): Distance Measure

To compare two NeuroSophic probabilities $P_N(E_1)$ and $P_N(E_2)$ we define the distance function as follows: where T_C, I_a, and F_m represents truth probability, algorithmic indeterminacy, and falsity.

$$d(E_1, E_2) = \sqrt{[T_c(E_1) - T_c(E_2)]^2 + [I_a(E_1) - I_a(E_2)]^2 + [F_m(E_1) - F_m(E_2)]^2} \quad (10)$$

This helps measure similarity or divergence between two AI decisions.

Worked example: To show how NeuroSophic probability can be used in real life applications for judgment AI-mediated systems. Let's consider three different areas such as healthcare, law, and education to make decisions.

Traditional probability theory fails to be capturing this complexity, as it reduces certainty to a single number (Hennig, 2024). Cognitive probability based NeuroSophic, on the other hand, gives a more complete representation by offering autonomous metrics for T_C, I_a, and F_m. This description shows how

these elements can be quantified for real-world situations and combined into a NeuroSophic theory that supporting a transparent and ethical assessment of AI-medicated results.

Let us consider as for example, the cognitive probabilities for three different areas such as healthcare, law, and education as well as the values of truth probability, algorithmic indeterminacy, and falsity, illustrated on Table 5.

Table 5: NeuroSophic Cognitive Score Components by Application Area

Area| cognitive score	T_C	I_a	F_m
Law	0.85	0.10	0.05
Recommendation	0.70	0.20	0.10
Healthcare	0.80	0.15	0.05

The NeuroSophic distance between E_2 and E_1 is calculated as follows:

The NeuroSophic distance between E_1, E_2 is calculated based on Equation (10). It as follows:

$$d(E_1, E_2) = Sqr\{(0.80 - 0.85)^2 + (0.15 - 0.10)^2 + (0.05 - 0.05)^2\} = 0.0707$$

This minor distance (approximately 0.07) suggests that the two occurrences—Healthcare diagnosis and law —are quite alike regarding their NeuroSophic elements (truth, indeterminacy, and falsity).

8. Reliability and Integrity Assessment in the NeuroSophic Framework

The NeuroSophic theory presents a multidimensional approach that transforms conventional accuracy metrics into cognitive metrics by integrating truth, indeterminacy, and falsity into its assessment and error patterns. This helps in detailed assessment of cognitive integrity and reliability in the case of uncertainty. NeuroSophic reliability scores are a quantitative measure of integrity based on the NeuroSophic. The NeuroSophic reliability score (***NRS***) is defined as:

$$\boldsymbol{NRS} = \text{clip}(w_T * T + \alpha * (1\text{-}w_I * I) + \beta * (1 - w_F * F), 0, 1) \quad (11)$$

Where:

- T, I, F ∈ [0,1] = (0.7, 0.4, 0.3), for example
- w_T, w_I, w_F are weights (e.g., 0.6, 0.25,0.15) and $\boldsymbol{\alpha}$ = 0.8, β = 0.5
- Where the function clip refers to the obtained values should be between the range

We must note that a higher NRS means more reliable AI-decision.

Using Equation (10) and the basic setting of the weights, the NeuroSophic reliability scores become 1.0. This result indicates that the model has high cognitive state. We must note that the NeuroSophic reliability cognitive scores are read on a continuous way as presented in the following Figure 3.

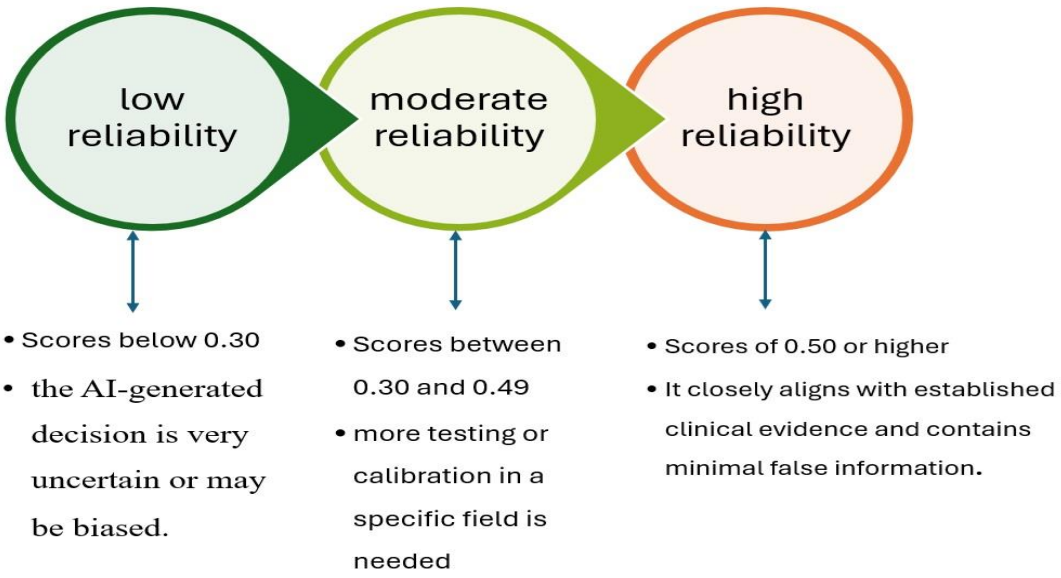

Fig.3: NeuroSophic reliability cognitive scores

8.1 Comparative Analysis of NeuroSophic Reliability Across AI Models for Clinical Diagnostics

Table 6 shows a synthetic dataset generated for 10 diagnostic medical area (A-J) tested on 10 different artificial intelligence (AI) models (Diabetes-Stroke/TIA). The values represent the cognitive NeuroSophic reliability scores (NRS), with scores ranging from 0 to 1. These scores present information about how reliability the AI-generated decisions are. In the case of higher scores of ***NRS***, the decisions will be more consistent with clinical evidence, while ***NRS*** lower scores refer to the model having more uncertainty or bias. For example, and as shown in the table, the analysis indicates that Model F achieves the highest overall average performance, whereas Model H demonstrates superior consistency across domains. Also, Model D performs robustly, securing leading positions in multiple areas. Domain-level difficulty varies significantly. For example, Cardio Risk emerges as the least challenging, while Sepsis represents the most complex environment for predictive modeling. Performance patterns reveal domain-specific specialization among models. For example, Model F excels in diabetes and COPD but underperforms in breast cancer, while Model H dominates in sepsis and stroke/TIA yet exhibits weakness in CKD. These findings support the adoption of a mixture-of-experts, assigning the most effective model to each domain rather than relying on a single universal solution. Considerations of stability and variability are critical—Model H offers balanced reliability, Model F provides peak performance with notable fluctuation, and Models A and B remain consistently weak. Future work should focus on metric normalization, implementation of domain-aware gating mechanisms, and exploration of weighted ensemble strategies to enhance robustness and generalizability.

Table 6: The synthetic dataset for Ten diagnostic areas tested on ten different AI models.

Table 6-1: Domains 1–5

Domain	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Diabetes	0.466463	0.296021	0.772003	0.702014	0.541750	1.000000	0.782163	0.784521	0.472494	0.736959
Cardio Risk	0.759865	0.663819	0.490210	0.984037	0.713979	0.977186	0.566905	0.819571	0.528055	0.692568
Hypertension	0.314593	0.483311	0.744020	0.401689	0.922823	0.802072	0.756336	0.584226	0.908108	0.525898
CKD	0.372067	0.403967	0.814414	0.877303	0.679726	0.823003	0.395707	0.227462	0.793356	0.316510
COPD	0.143519	0.349628	0.789792	0.863812	0.708665	0.896175	0.308335	0.643463	0.578873	0.530633

Table 6-2: Domains 6–10

Domain	A	B	C	D	E	F	G	H	I	J
Liver Disease	0.514926	0.531614	0.588131	0.726520	0.431540	0.575221	0.811530	0.784183	0.618350	0.480545
Breast CA	0.554451	0.510365	0.832958	0.569006	0.456513	0.373746	0.405965	0.569494	0.608361	0.670729
Lung CA	0.484684	0.435475	0.696710	0.685659	0.922566	0.693695	0.490007	0.833254	0.777280	0.920621
Sepsis	0.288813	0.442860	0.356066	0.501335	0.364260	0.647641	0.618798	0.736399	0.718932	0.322803
Stroke/TIA	0.570972	0.678720	0.580902	0.549905	0.458521	0.527055	0.549539	0.851869	0.550122	0.478827

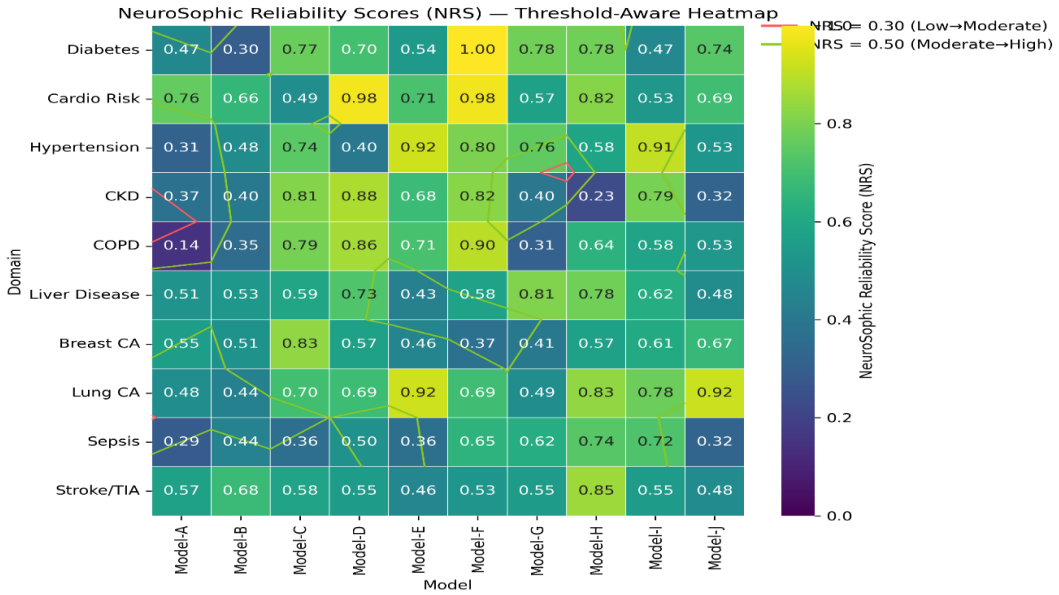

Fig.4: Threshold-Aware Heatmap

Figure 4 shows a heatmap of NeuroSophic Reliability Scores (***NRS***) across diagnostic area/domains (rows) and AI model variants (columns). Each cell has a score from 0 (least reliable) to 1 (most reliable), with darker colors showing less reliability and brighter colors showing more reliability. There are two threshold contours (1) red at ***NRS*** = 0.30 (low to moderate) and (2) green at ***NRS*** = 0.50 (moderate to high). Cells above the green contour show high reliability that matches clinical evidence. Cells between the contours show moderate reliability that needs to be validated, and cells below the red contour show low reliability with a more uncertainty or bias. This representation makes it easy to compare different

domains and models faster, which helps with threshold-based screening and finding patterns for focused improvements.

Definition (5): NeuroSophic distance integrity score

NeuroSophic distance integrity score is a new score that measures how different the integrity and reliability profiles of AI-mediated systems are across fields. It comes from Neutrosophic logic, which treats truth, indeterminacy, and falsity that makes the measure good for checking reliability in complex and unstable dynamic environment like healthcare sector.

For example, let us consider two different domains or fields i and j, each domain represented by an integrity profile, say $S_i(T_i, I_i, F_i)$ and $S_j(T_j, j, j)$. So, the mathematical form that measure the integrity distance between two profiles is presented in Equation (12).

$$D_{ij} = w_T * \left|\left|T_i - T_j\right|\right| + w_I * \left|\left|I_i - I_j\right|\right| + w_F * \left|\left|F_i - F_j\right|\right| \quad (12)$$

- T, I, F$\in$ [0,1] represent Truth**,** Indeterminacy, and Falsity components.
- w_T, w_I, w_F are weights reflecting the relative importance of each dimension.

A low NeuroSophic integrity distance, where D_{ij} close to zero, that means the domains are very consistent with each other. In this case, the AI-mediated system shows similar reliability profiles in different cases, which means robust generalizability and stable in performance.

In contrast, a high NeuroSophic distance integrity, where D_{ij} reflects significant divergence between domains. This difference recommended that the system might have weaknesses or adjustment problems that are specific to a certain area, which they need to be carefully analyzed and improvement to ensure reliable decision-making in every dynamic environment.

Worked Example: Suppose we have two domains: Diabetes with $S_i = (T_i = 0.80, I_i = 0.15, F_i = 0{,}05)$ and Cardio Risk with $S_i = (T_j = 0.70, I_j = 0.20, F_j = 0.10)$

Using Equation (11) and assuming $w_t = 0.5, w_i = 0.3, w_t = 0.10$

$$D_{ij} = 0.5 * \left|\left|0.80 - 0.70\right|\right| + 0.20 * \left|\left|0.15 - 0.20\right|\right| + 0.10 * \left|\left|0.05 - 0.10\right|\right| = 0.065$$

A distance of 0.065 shows that Diabetes and Cardio Risk are very similar, which recommend that the AI-mediated system acts the same way in all domains.

In another example, consider the synthetic pairwise NeuroSophic distance values for Ten diagnostic domain as show in Table 7. For analyzed the values in the Table, a lower value like Diabetes and Hypertension domains is equal (0.130) that shows robust consistency, while higher values like Sepsis and COPD is equal (1,000) shows a huge difference. Diabetes, Cardio Risk, Hypertension domains, and other metabolic conditions (CKD) represent a set with short distances, which represents the reliability profiles are applicable. Sepsis and Stroke/TIA domains show large distances from most domain areas, which indicates that there are domain-specific weaknesses that may need specific adjustment.

Table 7: Synthetic Data Table (NeuroSophic Distance)

Domain	Diabetes	Cardio Risk	Hypertension	CKD	COPD	Liver Disease	Breast CA	Lung CA	Sepsis	Stroke/ TIA
Diabetes	0.000	0.195	0.130	0.188	0.402	0.732	0.387	0.818	0.727	0.172
Cardio Risk	0.195	0.000	0.128	0.268	0.531	0.389	0.483	0.623	0.687	0.448
Hypertension	0.130	0.128	0.000	0.103	0.450	0.517	0.711	0.604	0.693	0.578
CKD	0.188	0.268	0.103	0.000	0.880	0.305	0.654	0.615	0.516	0.635
COPD	0.402	0.531	0.450	0.880	0.000	0.433	0.560	0.105	0.805	0.275
Liver Disease	0.732	0.389	0.517	0.305	0.433	0.000	0.562	0.206	0.165	0.509
Breast CA	0.387	0.483	0.711	0.654	0.560	0.562	0.000	0.400	0.208	0.510
Lung CA	0.818	0.623	0.604	0.615	0.105	0.206	0.400	0.000	0.489	0.340
Sepsis	0.937	0.890	0.897	0.693	1.000	0.290	0.339	0.663	0.000	0.993
Stroke/TIA	0.297	0.615	0.764	0.830	0.417	0.686	0.687	0.492	0.993	0.000

Figure 5 illustrates clear clustering within metabolic domains Diabetes, Cardio Risk, Hypertension, CKD, with low integrity distances that represent a robust internal consistency and the potential for generalizable adjustment. Sepsis and Stroke/TIA, on the other hand, are very different from most domain areas, which may have specific weaknesses that require unique modeling. The intermediate distances noted for domains such as COPD and Liver Disease indicate a certain level of overlap, requiring selective modifications rather than total separation.

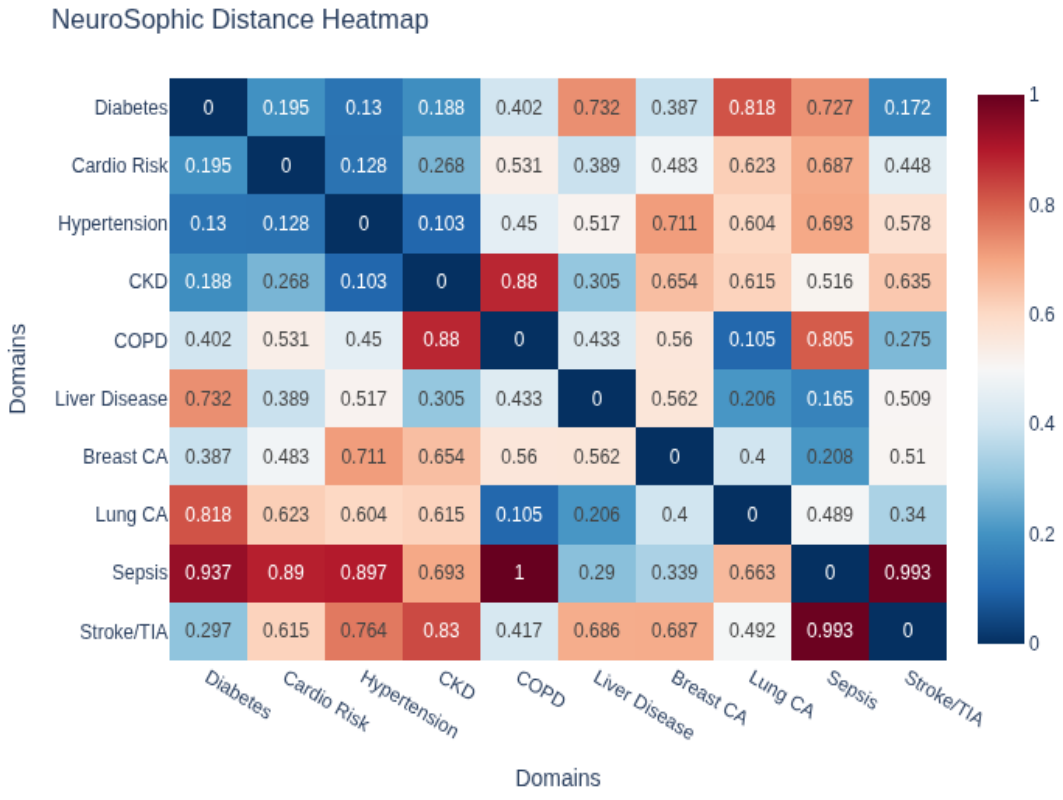

Fig.5: NeuroSophic Distance Heatmap among Ten Diagnostic Domains

9. Conclusion and Future Work

NeuroSophic cognition theory provides a comprehensive framework for simulating human cognition in AI-mediated environments that merge Neutrosophic theory with Neurojico philosophy to create AI-mediated dynamic environments. This paper introduces a novel view of the AI-based cognitive theory, a multidimensional representation of Neutrosophic parameters, truth, indeterminacy, and falsity. The basic NeuroSophic cognition matrix is introduced and defined to illustrate the complexities of cognitive states and enables the calculation of fundamental indices, such as the Cognitive Sovereignty Index (***CSI***) and the NeuroSophic Global Cognition Score (***GNS***). These metrics offer significant insights into autonomy, adaptability, and vulnerability within the algorithmic influence approach. AI-assisted context diagnosis is one example of how the theory could make intelligent systems more transparent, ethical, and trustworthy. Some new operators are presented, such as the aggregation, expected value, cognitive distance integrity, and proof using traced numerical examples.

Following research may enhance this model to incorporate dynamic simulations and cross-domain evaluations, thus contributing to the advancement of more AI-mediated dynamic systems that are flexible and focus on people or systems.

The NeuroSophic cognition model offers numerous promising directions for future research. Here we will list some future points for research around NeuroSophic theory. (1) Improving the NeuroSophic cognitive matrix to show how cognitive states change over time when algorithms are still running, so that sovereignty and resilience can be measured in real time, (2) Using the proposed framework in different domain (Humphreys, 2024)), like cybersecurity, social systems, and governance, to check how well AI-based decision-making follows moral and cognitive rules, (3) Combine NeuroSophic indices with Explainable AI methods for making AI systems easier to understand, making them more open and trustworthy. (4) Make machine learning algorithms that can change ***CSI*** and GNS weight parameters based on priorities that are specific to the situation. (5) Expand probabilistic modeling for critical situations like self-driving cars and financial systems, where bias and uncertainty must be precisely controlled (Ping et al. 2025). (6) Use ***CSI*** and ***GNS*** as examples when making ethical AI policies and governance structures that put a lot of priority on human autonomy, and (7) Hybrid NeuroSophic with learning that examines the integration of deep learning. architectures to create AI systems that are aware of how people think and can find the right balance between performance and keeping sovereignty safe.

Acknowledgement

The author conceived all core ideas, mathematical models, and research presented in this manuscript. Generative AI was used a little in some parts within the papers an assistive tool for polishing the English prose. The author has reviewed and takes full intellectual responsibility for all final content.

References

Alowais, S.A., Alghamdi, S.S., Alsuhebany, N. et al. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 23, 689. https://doi.org/10.1186/s12909-023-04698-z

Atkinson, R. (2025). Cognitive sovereignty and neurocomputational harm in predictive digital platforms. Ethics Inf Technol 27, 66. https://doi.org/10.1007/s10676-025-09873-y

Badawy, W. (2025) Algorithmic sovereignty and democratic resilience: rethinking AI governance in the age of generative AI. AI Ethics 5, 4855–4862. https://doi.org/10.1007/s43681-025-00739-z

Das, S., Roy, B.K., Kar, M.B. et al. (2020). Neutrosophic fuzzy set and its application in decision making. J Ambient Intell Human Comput 11, 5017–5029. https://doi.org/10.1007/s12652-020-01808-3

Fintz, M., Osadchy, M. & Hertz, U. Using deep learning to predict human decisions and using cognitive models to explain deep learning models. Sci Rep 12, 4736 (2022). https://doi.org/10.1038/s41598-022-08863-0

Franciskus Antonius Alijoyo, S. Janani, Kathari Santosh, Safa N. Shweihat, Nizal Alshammry, Janjhyam Venkata Naga Ramesh, Yousef A. Baker El-Ebiary, (2024). Enhancing AI interpretation and decision-making: Integrating cognitive computational models with deep learning for advanced uncertain reasoning systems, Alexandria Engineering Journal, Vol. (99) pp. 17-30, https://doi.org/10.1016/j.aej.2024.04.073.

George Siemens, Fernando Marmolejo-Ramos, Florence Gabriel, Kelsey Medeiros, Rebecca Marrone, Srecko Joksimovic, Maarten de Laat, (2022). Human and artificial cognition, Computers and Education: Artificial Intelligence, Vol. (3),100107, https://doi.org/10.1016/j.caeai.2022.100107.

Gkanatsiou, M.A.; Triantari, S.; Tzartzas, G.; Kotopoulos, T.; Gkanatsios, S. Rewired (2025). Leadership: Integrating AI-Powered Mediation and Decision-Making in Higher Education Institutions. Technologies, 13, 396. https://doi.org/10.3390/technologies13090396

Gonzalez, C., & Heidari, H. (2025). A cognitive approach to human–AI complementarity in dynamic decision-making. Nat Rev Psychol 4, 808–82. https://doi.org/10.1038/s44159-025-00499-x

Gupta M.M., (2011) On fuzzy logic and cognitive computing: Some perspectives, Scientia Iranica, 18(3), 2011, Pages 590-592, ISSN 1026-3098,https://doi.org/10.1016/j.scient.2011.04.010.

Hassanien, Aboul Ella (2025). Neurojico: When Artificial Intelligence Reshapes Humanity 2.0. Amer Publishing. ISBN: **978-633-8332-80-0**

Hennig, C. (2024). Probability Models in Statistical Data Analysis: Uses, Interpretations, Frequentism-as-Model. In: Sriraman, B. (eds) Handbook of the History and Philosophy of Mathematical Practice. Springer, Cham. https://doi.org/10.1007/978-3-031-40846-5_105

Humphreys, D., Koay, A., Desmond, D. et al. (2024). AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business. AI Ethics 4, 791–804. https://doi.org/10.1007/s43681-024-00443-4

Loucks, D.P. & van Beek, E. (2017). An Introduction to Probability, Statistics, and Uncertainty. In: Water Resource Systems Planning and Management. Springer, Cham. https://doi.org/10.1007/978-3-319-44234-1_6

Majumdar, P. (2015). Neutrosophic Sets and Its Applications to Decision-Making. In: Acharjya, D., Dehuri, S. & Sanyal, S. (eds) Computational Intelligence for Big Data Analysis. Adaptation, Learning, and Optimization, vol 19. Springer, Cham. https://doi.org/10.1007/978-3-319-16598-1_4

McKinlay, Steve (2020), Trust and Algorithmic Opacity, in Kevin Macnish, and Jai Galliott (eds), Big Data and Democracy. https://doi.org/10.3366/edinburgh/9781474463522.003.0011

Ping Lu, Sunan Zhang, Feihong Tan, Fulin Zhang, Yuxiang Feng, Bo Hu,(2025), An uncertainty-aware safe-evolving reinforcement learning algorithm for decision-making and control in highway autonomous driving, Engineering Applications of Artificial Intelligence, vol.(161), Part A,112108,https://doi.org/10.1016/j.engappai.2025.112108.

Pop, A., & Pierson, J. (2025). Algorithmic Governmentality, Digital Sovereignty, and Agency Affordances. Weizenbaum Journal of the Digital Society. Vol. 3 No. 2 (2023) https://ojs.weizenbaum-institut.de/index.php/wjds/article/view/87/80

Slawner, K. (1996). The Decline of Sovereignty? In: Denham, M.E., Lombardi, M.O. (eds) Perspectives on Third-World Sovereignty. International Political Economy Series. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-349-24937-4_9

Smarandache, F. (1998). Neutrosophy: Neutrosophic Probability, Set, and Logic. American Research Press. ISBN: 978-1-879585-33-8

Songlin Xu, & Xinyu Zhang (2023) Augmenting Human Cognition with an AI-Mediated Intelligent Visual Feedback. CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems Article No.: 26, pp 1 - 16 https://doi.org/10.1145/3544548.3580905

Sylvia Lu (2021), Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence, vol. 23 Vanderbilt Journal of Entertainment and Technology Law 99.

Umoke C.C., Sunday O. N., and Oroke A. O. (2025). The Governance of AI in Education: Developing Ethical Policy Frameworks for Adaptive Learning Technologies. International Journal of Applied Science and Mathematical Theory E- ISSN 2489-009X P-ISSN 2695-1908, 11(2), 71–88.

Umberto R., (2008) Neutrosophic logics: Prospects and problems, Fuzzy Sets and Systems, 159(14), 2008, pp.1860-1868, ISSN 0165-0114, https://doi.org/10.1016/j.fss.2007.11.011.

Zhong Y. X., (2006) A Cognitive Approach to Artificial Intelligence Research, 2006 5th IEEE International Conference on Cognitive Informatics, Beijing, China, 2006, pp. 90-100, doi: 10.1109/COGINF.2006.365682.