



# INTERDISZCIPLINÁRIS MI-KUTATÁSI MÓDSZERTAN

## INTERDISCIPLINARY AI RESEARCH METHODOLOGY

Kolcsár Levente,<sup>1</sup> Bakó László<sup>2</sup>

<sup>1</sup> Sapiientia EMTE, Marosvásárhelyi Kar, Villamosmérnöki Tanszék, Marosvásárhely, Románia, [kolcsar.levente@ms.sapiientia.ro](mailto:kolcsar.levente@ms.sapiientia.ro)

<sup>2</sup> Sapiientia EMTE, Marosvásárhelyi Kar, Villamosmérnöki Tanszék, Marosvásárhely, Románia, [lbako@ms.sapiientia.ro](mailto:lbako@ms.sapiientia.ro)

### Abstract

The use of artificial intelligence (AI) in our time is becoming more and more extensive, so it is no question that there is a need for new AI concepts. We have developed our research methodology to accommodate new cognitive science theories, which we present through a concrete example: how Integrated Information Theory (IIT) was born and what it is good for. IIT is one of the main theories of consciousness in cognitive neuroscience and phenomenology. The main pillar of our methodology is the philosophy of science, which puts IIT into context, examines its origin and usability. This serves the purpose of a framework to be able to interpret ideas from the most diverse fields and to rank them according to their engineering applicability, so that we can create new, useful AI toolkits from them.

**Keywords:** *AI, research, philosophy of science, IIT.*

### Összefoglalás

Korunk mesterségesintelligencia- (MI) felhasználása egyre jobban kiterjed, ezért nem kérdés, hogy új MI-konceptiókra igény van. Kutatási módszertanunkat a kognitív tudományokból érkező új elméletek befogadására dolgoztuk ki, amelyet egy konkrét példán keresztül be is mutatunk: hogyan született, és mire jó az integrált-információ-teória (IIT), amely a kognitív idegtudományok és a fenomenológia egyik fő elmélete a tudatról. Módszertanunk alappillére a tudományfilozófiai gondolkodásmód, amely kontextusba helyezi az IIT-t, megvizsgálja annak eredetét és felhasználhatóságát. Módszertanunk azt a célt szolgálja, hogy e keretrendszerrel a legkülönbözőbb területekről érkező ötleteket is képesek legyünk értelmezni és azokat rangsorolni mérnöki felhasználhatóságuk szerint, hogy ezekből új, hasznos MI-eszköztárakat készíthessünk.

**Kulcsszavak:** *MI, kutatás, tudományfilozófia, IIT.*

## 1. Bevezetés

A dolgozatunkban tárgyalt tudományfilozófiai módszertannal az MI-kutató a kognitív tudományok területéről érkező teóriákat kontextusba helyezi, és csak utána kezd dolgozni vele. Erre azért van szükség, mert az interdiszciplináris munkához a rendezettség és az egységes értelmezés igényli a jól körülírt keretrendszert. Ezzel a módszerrel a kutató megismeri magát a teóriát, az eredetét, a matematikai leírását és felhasználhatóságát. A módszertan szerkezetét megalkotó három fő tárgyat definiáljuk:

**α. A filozófia** olyan szisztematikus, nyelvi és gondolati struktúrák együttese, amelyek a világról alkotott elképzeléseinkből állnak. A filozófia fő kérdései: Mi a természet? Mi az eredete? Mi a fundamentuma? Kik vagyunk? Ezek az elképzelések nemcsak a megismerő módszereinkben, hanem a hétköznapi életvitelünkben is megjelennek.

**β. A tudomány** definíciónk szerint olyan elágazó folyamat, amely a filozófiához hasonlóan a természetet kívánja megismerni, de ezt a megismerést szisztematikus megfigyelésre és kísérletekre alapozza. Ebből következetesen látszik, hogy a két

terület átfedése az új tudományos elméletek megalkotásakor történik, hiszen a legelső fázisban, kísérlet nélkül csak a képzeteletére hivatkozhat a kutató. Ebből is látszik, de historikusan is evidens [1], hogy a tudomány a filozófiából ered.

**γ. A tudományfilozófia** a filozófia azon ága, amely a tudományos gondolkodással és módszerrel foglalkozik. Fő célja: rendezetten irányítani és javítani a tudományos kutatást és módszertant, szem előtt tartva a tudományos gondolkodás problémáit és a humán részrehajlást. Néhány példa, ahol szerintünk fontos: atomfizika, kozmológia, biológia, kognitív tudományok, MI-kutatás, pszichológia, etika, humán tudományok. Néhány példa, ahol szerintünk kevésbé fontos: alkalmazott tudományok, technológiák, ipar és minden olyan terület, ahol a meglévő teóriákat érintetlenül felhasználják.

## 2. A módszertan szerkezete

Az MI-kutatónak a talált vagy újonnan kitalált teóriát a felsorolt kontextusokban kell elemeznie és dokumentálnia, mindegyik kontextusban külön aspektust vizsgál. A kontextusok a görög betűkkel definiált tárgyakhoz kapcsolódnak ( $\alpha$ ,  $\beta$ ,  $\gamma$ ):

- I. Historikus kontextus ( $\alpha$ ,  $\beta$ )
- II. Matematikai leírás ( $\alpha$ ,  $\beta$ ,  $\gamma$ )
- III. Episztemológiai kontextus ( $\beta$ ,  $\gamma$ )
- IV. Ontológiai kontextus ( $\alpha$ ,  $\beta$ ,  $\gamma$ )
- V. Metodológiai kontextus ( $\beta$ ,  $\gamma$ )
- VI. Tudományos eredmények ( $\beta$ )
- VII. MI-alkalmazhatóság ( $\beta$ )
- VIII. Axiológiai kontextus ( $\alpha$ ,  $\gamma$ )

Mindegyik kontextusban 1–5-ig való pontozást kell végezni (1. táblázat), annak függvényében, hogy az adott teória mennyire hangsúlyos abban a kontextusban. Ha egyáltalán nem releváns az adott kontextusra, akkor 1-es értéket kap, ha maximálisan releváns, akkor 5-ös értéket.

1. táblázat. Az IIT módszertani ponttáblája

| Kontextus | Pontszám |
|-----------|----------|
| I.        | 5        |
| II.       | 5        |
| III.      | 3        |
| IV.       | 1        |
| V.        | 4        |
| VI.       | 4        |
| VII.      | 3        |
| VIII.     | 4        |

Például: az IIT-teória az MI-alkalmazhatósági kontextusban 3-as pontszámot kap. Ezt a táblázatot felhasználhatjuk arra, hogy eldöntsük, melyik teóriával érdemes nagyobb prioritással foglalkozni. A kutatónak a dokumentációban hivatkozni kell a megjelölt forrásokra, hogy a későbbi cikkírásnál már felhasználható legyen.

A módszertan végén feljegyezzük a teória általános tulajdonságait, és végzünk egy súlyozott, összesített pontozást, majd %-ban kifejezzük a teória MI-alkalmazhatóságát.

## 3. Módszertani példa

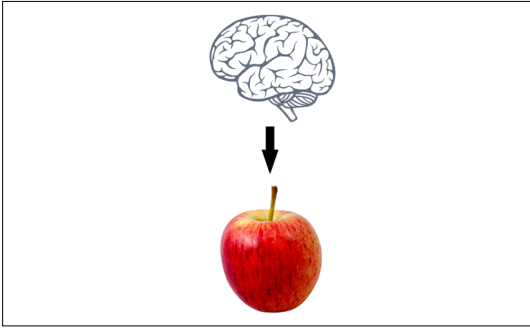
Egy olyan reprezentatív teóriát választottunk példaként, amely értelemszerűen műszaki területeken kívüli tudományágban született és erős I., illetve II. kontextusa van, és jó példa a tudományfilozófiai kérdéskörökre is.

- Megnevezés: Integrated Information Theory (IIT); [2]
- Tudományág: kognitív idegtudományok és fenomenológia;
- Főszerző: dr. Giulio Tononi agykutató;
- A teória: mekkora a kauzálisan integrált információ a hálózatban, ez egy mérce arra vonatkozóan, hogy mennyire lehet „tudatos” egy hálózat. Jelölése:  $\Phi$ .

### 3.1. Historikus kontextus (I.)

Ebben a kontextusban azt vizsgáljuk, hogy a teóriát megelőző gondolkodási forma, esetleg tudományos eljárás mikor kezdődött el a történelem folyamán, esetleg milyen viszonyban van a klasszikus filozófiai irányzatokkal. Ez a fejezet sok esetben akár ki is hagyható, de ajánlott tájékozódni, mert vannak teóriák, amelyek hosszú múlttal rendelkeznek ebben. Az IIT ebben 5-ös pontszámot kap, mivel nagyon régi filozófiai problémából és kérdésből indul ki: Mi a tudat? Miért nehéz megmagyarázni? Az IIT szerzője, dr. Tononi erre nagyon jó historikus és irodalmi betekintést nyújt a Phi-könyvében [3]. Könyvében irodalmi formában elmagyarázza, hogy Galileo hogyan gondolkodott a tudat kérdéséről. A tudat problémáját a modern korunkban David Chalmers filozófus fogalmazta meg [4], ami úgy szól: hogyan lesz a biológiai mechanizmusból egy belső egységes és minőségi érzékelés? Miért látom pirosnak a pirosat? Miért sajátos és minőségi a tudatunk?

Erre a tudatproblémára építette fel az IIT az axiómáit. (1. ábra)



1. ábra. A tudat problémája. Illusztráció [5]

### 3.2. Matematikai leírás (II.)

A matematikai leírásban a teória formális leírását tanulmányozzuk, ide minden olyan axióma, egyenlet és matematikai összefüggés leírásra kerül, amely a teória alapjait képezi. Az axióma az általános megfigyelésből származó tulajdonság, a posztulátum az ebből származó törvényszerűség kijelentése teoretikus formában és a matematikai kifejezés a posztulátum formális leírása.

Erre a fejezetre 5 pontot kap az IIT.

#### 3.2.1. Létezés

Axióma: A tudat létezik, ezért hatással kell rendelkeznie.

Posztulátum: A hálózat ok-okozati hatással kell, hogy rendelkezzen.

Matematikai kifejezés: Adott egy sztochasztikus hálózat  $x_t$  állapotban. Adott a hálózat átmeneti valószínűségek mátrixa  $T$ , amely leírja az átmenetek valószínűségét az  $x_t$  állapotokból az  $x_{t+1}$ -be. Ebből felírható a hálózat feltételes valószínűségi ok-okozati repertoárja:

$$\begin{aligned} P_{\text{cause}}(x_{t-1} | x_t) \\ P_{\text{effect}}(x_{t+1} | x_t) \end{aligned} \quad (1)$$

#### 3.2.2. Kompozíció

Axióma: A tudat megkülönböztethető fenomenológiai elemekből épül fel. Pl. hangok, színek, érzetek.

Posztulátum: A hálózatnak olyan mechanizmusok komplexumát (elemek részhalmazait) kell formálnia, amelyek együttesen korlátozzák az ok-okozati repertoárt.

Matematikai kifejezés: Ha például az adott hálózat teljesen összekapcsolt, 3 elemű csomópontból áll, akkor:

$$M = \{A, B, C, AB, AC, BC, ABC\} \quad (2)$$

ahol  $M$  a mechanizmusok halmaza, amelyből minden mechanizmus egy komplexum, ha az arra

a számolt  $\Phi > 0$ . (A mechanizmusok tartalmazzák a kétirányú kapcsolatokat. Pl.  $AB = \{AB, BA\}$ )

#### 3.2.3. Információ

Axióma: Minden tudatos tapasztalat specifikus, amely eltér más lehetséges tudatos állapotoktól.

Posztulátum: A hálózatnak differenciált ok-okozati struktúrát kell meghatároznia. (Nagy információs eltérés a véletlenszerűségtől.)

Matematikai kifejezés:

$$EI_{\text{cause}} = D(P_{\text{cause}}, P_{\text{max}})$$

$$EI_{\text{effect}} = D(P_{\text{effect}}, P_{\text{max}})$$

$$EI = \min(EI_{\text{cause}}, EI_{\text{effect}}), \quad (3)$$

ahol:

$EI_{\text{cause}}$  – effektívok-információ;

$EI_{\text{effect}}$  – effektívokozat-információ;

$P_{\text{max}}$  – a maximális entrópia eloszlása;

$EI$  – effektív információ;

$D$  – két eloszlás közti távolság, amely lehet Earth's Mover Distance vagy Kullback–Leibler-divergencia.

#### 3.2.4. Integráció

Axióma: A tudat egységes, nem redukálható független részekre.

Posztulátum: Az integrált információ ( $\Phi$ ) azt reprezentálja, hogy mennyire irreducibilis az információ egy mechanizmusban.

Matematikai kifejezés: A hálózat bipartíciói:

$$\{\{A\}, \{B, C\}\}, \{\{B\}, \{A, C\}\}, \{\{C\}, \{A, B\}\}$$

$$\Phi = \min_{MIP} [D(P_{\text{total}}, P_{\text{partitioned}})] \quad (4),$$

ahol:

$\Phi > 0$ , ha a részhálózat irreducibilis;

$P_{\text{total}}$  – teljes hálózat ok-okozati repertoárja;

$P_{\text{partitioned}}$  – particionált hálózat ok-okozati repertoárja;

$MIP$  – minimum információs partíció;

#### 3.2.5. Exklúzió

Axióma: A tudat tartalmilag időben és térben jól meghatározott állapotot, amely kizár más állapotokat.

Posztulátum: A hálózatban az a legnagyobb komplexum, amely a legnagyobb  $\Phi$ -vel rendelkezik.

Matematikai kifejezés:

$$\Phi_{\text{max}} = \arg \max_{m \in M} \{\Phi(m)\} \quad (5)$$

### 3.3. Episztemológiai kontextus (III.)

Ez a fejezet azt vizsgálja, hogy a teória mennyire bővíti ki a jelenlegi tudásunkat, mekkora magya-

rázó erővel bír. Ez fontos szempont az alkalmazhatóság miatt. Az MI-ben szeretnénk minél több hasznos tulajdonságot implementálni, amelyekkel jobb problémamegoldó mesterséges ágenst kapunk.

Természetes ágens → Mesterséges ágens  
(élőlény) (MI)

A természetes ágens tulajdonságaiból származtatjuk a mesterséges ágens tulajdonságait. A kognitív tudományok az ember kognitív tulajdonságait tanulmányozzák. Az IIT a tudatra megalkotja annak kvantifikált értékét, a  $\Phi$ -t. Erre a fejezetre átlagos 3 pontot kap a teória.

### 3.4. Ontológiai kontextus (IV.)

Ez a kontextus azzal foglalkozik, hogy a teória milyen közegparadigmát használ a magyarázatában. Ez a közeg lehet elektromágnesesség, sejtek, atomok és egyéb jól ismert tudományos modellek, de lehet akár egy teljesen más paradigma is, mint a megszokottak.

Jelen esetben az IIT ontológiailag független, akárcsak Shannon információs elmélete, ami azt jelenti, bármilyen közegre érvényes, mivel állapotbeli dinamikákat tárgyal, és nem azt, hogy milyen folyamatok és elemek alkotják azt az állapotot. Az IIT számolható idegsejtekre és logikai kapukra is. Ez az ontológiai függetlenség hasznos fogalom az MI-kutatásban, hiszen adaptálható sok ilyen teória a műszaki területen is. Mivel az IIT-nek nincs ilyen alapja, ezért 1 pontot adunk erre.

### 3.5. Metodológiai kontextus (V.)

Ez a fejezet azt vizsgálja, hogy a teórián mennyire végezhető el az egyik tudományos kutatási módszertani eljárás:

Deduktív eljárás: 0. Tájékozódás → 1. Teória → 2. Predikciós módszer → 3. Tesztelni a predikciót (Mérés) → 4. A predikció ellenőrzése → 5. A teóriát elfogadjuk vagy elvetjük.

Induktív eljárás: 0. Tájékozódás → 1. Megfigyelés (Mérés) → 2. Minták felismerése → 3. Generalizáció → 3. Teória → 4. Teória ellenőrzése → 5. A teóriát elfogadjuk vagy elvetjük.

Az IIT egy deduktív eljárással készült teória. Mivel az IIT már felmutatott agykutatási eredményeket, feltételezzük, hogy a deduktív módszertan elvégezhető rajta. Mivel sokelemű hálózatokra a számolása erőforrás-igényes (NP-nehéz), így approximációt igényel, ezért 4 pontot kap erre a fejezetre.

### 3.6. Tudományos eredmények (VI.)

Azt vizsgáljuk itt, hogy a teória mennyire validált megfigyelési és kísérleti módszerekkel. Ha sok kísérletben felmutatott eredményt, akkor az azt igazolja, hogy a természeti jelenséget jól tudja magyarázni.

Az IIT ebben a kontextusban 4 pontot kap, mivel a szintén ismert Global Workspace Theory tudat-teóriával holtversenyben tudott predikciós eredményeket felmutatni [6]. Ha a teória még nem mutatott fel kísérleti eredményt, akkor ebben a kontextusban 1-es pontszámot adunk. Az agykutatásban további elért eredmények: [7, 8]

### 3.7. MI-alkalmazhatóság (VII.)

Az IIT teóriára kidolgoztak egy publikus Python könyvtárat (pyphi), amely szabadon felhasználható, ha a hozzákötődő publikációra mindig hivatkozunk. [9] A könyvtárral tetszőleges n elemű sztochasztikus hálózatra lehet számolni  $\Phi$ -t.

A 2. ábra szemlélteti a példánkat, amelyben egy logikai kapukból felépített hálózatra számolunk  $\Phi$ -t.

Python-kód:

#A csomópontok neve és sorrendje:

```
node_labels = ('OR','AND','XOR')
```

#A 2. ábra szerinti kezdeti állapotvektor [n]:

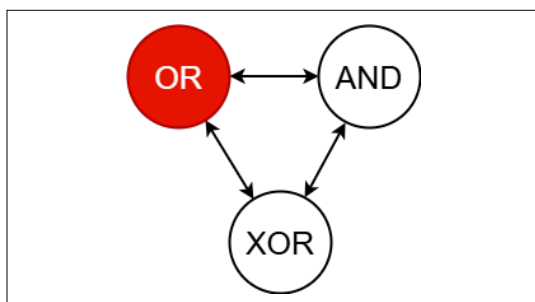
```
state = (1, 0, 0)
```

#Kapcsolati mátrix [n \* n]:

```
cm = np.array ([[0,1,1],
                [1,0,1],
                [1,1,0]])
```

#Átmeneti valószínűségek mátrixa [2^n \* 2^n]:

```
tpm = np.array ([[1, 0, 0, 0, 0, 0, 0, 0],
                 [0, 0, 0, 0, 1, 0, 0, 0],
                 [0, 0, 0, 0, 0, 1, 0, 0],
                 [0, 1, 0, 0, 0, 0, 0, 0],
                 [0, 1, 0, 0, 0, 0, 0, 0],
                 [0, 0, 0, 0, 0, 0, 0, 1],
                 [0, 0, 0, 0, 0, 1, 0, 0],
                 [0, 0, 0, 1, 0, 0, 0, 0]])
```



2. ábra. Példa hálózat (vörös "1", fehér "0")

**#Megjegyzés:**

*A cm kapcsolati mátrix értékei 1 és 0 értéket vehetnek fel. A főátló jelen esetben azért 0, mert a példahálózat nem tartalmaz olyan csomópontot, amelynek van önmagával kapcsolata.*

*A tpm olyan mátrix, amely felírja az összes lehetséges t állapotból (sorindex) az összes lehetséges t+1 állapotba (oszlopindex) átmenő valószínűségi értékeket. (0–1 közötti értékek)*

**#Létrehozuk a hálózatot:**

```
net = pyphi.Network(tpm, cm, node_labels)
```

**#Kiszámoljuk a hálózat legnagyobb komplexumát:**

```
mct = pyphi.compute.major_complex(net, state)
```

**#Felhasználjuk a komplexum  $\Phi$ -jét:**

```
phi_maxt = mc.phi
```

```
# $\Phi_{maxt} = 1.916$ 
```

Jelenleg azt tudjuk mondani, hogy az IIT arra jó, hogy hálózatanalízist végezzünk arra vonatkozóan, hogy mennyire rekurrens (visszacsatolt) egy hálózat, ugyanis a visszacsatolások növelik a  $\Phi$  értékét. A modell szerint, minél több rekurrens ág van egy hálózatban, annál integráltabb a hálózat. Ez a tulajdonság dinamikus memóriaként is értelmezhető, amit nemcsak a csomópontok pillanatnyi állapotai, hanem a hálózat topológiája is meghatároz. Mivel ez egy hálózatanalízis, egyelőre nem látjuk a teória direkt hasznát az MI-ban, ezért az MI alkalmazhatóságra egyelőre semleges: 3 pontot adunk. Lehetséges felhasználás: Az IIT  $\Phi$ -je tanítási kritériumként használható változó topológiájú hálózatokban, illetve a jövőre nézve a  $\Phi$  egy mérce lehet az MI-rendszerek etikai besorolásánál. Ez a kontextus további analízist igényel részünkről, mivel úgy gondoljuk, hálózatoptimizálásokban és szenzorfüziós problémákban is alkalmazható lehet.

**3.8. Axiológiai kontextus (VIII.)**

Ez a fejezet azt vizsgálja, hogy a teória mennyire kötődik az MI etikai kérdésköréhez.

Az MI jelenlegi etikai témaköre egyre fontosabb kérdés, ezért ez a kontextus sok modellnél fontos lehet. A nonprofit Eleos AI szerint [10] két fő veszély fenyegeti az MI evolúcióját:

- túlbecsüljük (humán tulajdonságokkal ruházuk fel);
- alulbecsüljük a képességeit.

Az IIT ebben a kontextusban 4-es pontszámot kap, mivel a  $\Phi$  mérceje szerint logikai kapukból is építhetünk „tudatos” MI-t. Persze, ebben a kérdéskörben kritikusak vagyunk a teóriával kapcsolatban. Vigyázni kell azzal, hogy az MI-t ne ruházzuk fel túl gyorsan humán tulajdonságokkal, ezért

egyelőre a  $\Phi$  számunkra csak egy metrika, de nem több. Az viszont szerintünk jó, hogy elindultunk ezzel is a tudat megértése felé. Szerintünk ez csak egy változó a sok közül, ami a tudat tulajdonsága lehet. A tudatot és az intelligenciát szándékosan nem definiáltuk, mert az idegtudományokban máig egy nyitott kérdéskör. Szeretnénk nyitottnak hagyni az MI körében is, hogy új modelleket készíthessünk. Az IIT szerint a tudatosság belső folyamat, és megkülönböztethető az intelligenciától, ami csak kimeneti tulajdonság.

**3.9. Módszertani összefoglaló**

A dokumentáció végén lejegyezzük az IIT-teória általános tulajdonságait:

- agyi működésből inspirálódott? Igen;
- hagyományos/statistikai MI? Nem;
- top-down modell? Csak módszerében;
- bottom-up modell? Nem.
- induktív vagy deduktív? Deduktív.

A kontextusok pontjait (1. táblázat) felhasználva végzünk egy súlyozott összpontosítást az MI-alkalmazhatóságot (VII.) legjobban hangsúlyozva, de még jelentősnek számít a matematikai leírás (II.) és a tudományos eredmények (VI.). Ha vannak tudományos eredmények, feltételezzük, hogy az a teória alkalmazhatóbb lehet, hiszen a természetben már egy funkcióra jó predikciót nyújtott. Úgy gondoljuk, hogy a III., IV., V., VIII. kontextus csak indirekt járul hozzá az alkalmazhatósághoz, de az I. semennyire, hiszen történelmi fejezet. Az IV. kontextus fordított előjelű, mert minél hangsúlyosabb valami ontológiailag, annál nehezebb a teóriát interdiszciplinárisan áthozni. Ez a súlyozás még módszertani tapasztalatot és fejlesztést igényel.

$$\text{sum} = \text{I.} \cdot 0 + \text{II.} \cdot 0.5 + \text{III.} \cdot 0.25 - \text{IV.} \cdot 0.25 + \text{V.} \cdot 0.25 + \text{VI.} \cdot 0.5 + \text{VII.} \cdot 1 + \text{VIII.} \cdot 0.1$$

Az összeget hármasszabállyal átalakítjuk %-ra:

$$IIT_{rating} = 74\%.$$

**4. Következtetések**

A módszertanunk egy kiindulási irány ahhoz, hogy kognitív teóriákból hasznos MI-eszközöket készítsünk. A módszertan természetesen fejlesztést igényel, de ahhoz sok teóriára kell alkalmaznunk. A cél az, hogy bármilyen teóriát és modellt elemezni tudjunk, függetlenül attól, hogy milyen kognitív tudományból származik. Ha van egy ilyen tárházunk, akkor sokkal könnyebb új modellt megalkotni, mert már ismerjük a témaköri elképzeléseket és problémákat. Ha a tárházunkból találunk ígéretes alkalmazási ötletet, azt megvalósítjuk.

A módszertan által elemzett, reprezentatív IIT-teória további analízist igényel részünkről MI-alkalmazhatóságban, ami egy külön feladat lesz számunkra a közeljövőben.

### Szakirodalmi hivatkozások

- [1] Quack M.: *Science and Arts, Philosophy and Science: Why after All? Why Not?*, Helvetica Chimica Acta, 106/4. (2023) 1–3.  
<https://doi.org/10.1002/hlca.202200174>
- [2] Oizumi M., Albantakis L., Tononi G.: *From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0*. PLoS Computational Biology, 2014, 1–25.  
<https://doi.org/10.1371/journal.pcbi.1003588>
- [3] G.: *Phi: A Voyage from the Brain to the Soul*. 1. ed. Pantheon Books, New York, 2012, 1–38.
- [4] Chalmers D.: *Facing up to the Problem of Consciousness*. Journal of Consciousness Studies, 2/3. (1995) 200–219.  
<https://philpapers.org/rec/CHAFUT>
- [5] UI Here free samples. <https://www.uihere.com/> (2025.02.22)
- [6] Cogitate Consortium, Ferrante O., Gorska-Klimowski U., et al.: *An Adversarial Collaboration to Critically Evaluate Theories of Consciousness*. bioRxiv 2023.06.23.546249, 1–68.  
<https://doi.org/10.1101/2023.06.23.546249>
- [7] Ferrarelli F., Massimini M., Sarasso S., Casali A., Riedner B.A., Angelini G., Tononi G., Pearce R.A.: *Breakdown in Cortical Effective Connectivity during Midazolam-Induced Loss of Consciousness*. Proc Natl Acad Sci USA. 107/6. (2010) 2681–2686.  
<https://doi.org/10.1073/pnas.0913008107>
- [8] Casali A.G., Gosseries O.: *A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior*. Science Translational Medicine, 5. (2013) 1–12.  
<https://hdl.handle.net/2268/171542>
- [9] Mayner W.G.P., Marshall W., Albantakis L., Findlay G., Marchman R., Tononi G.: *PyPhi: A Toolbox for Integrated Information Theory*. PLoS Computational Biology, 14/7. (2018) 1–21.  
<https://doi.org/10.1371/journal.pcbi.1006343>
- [10] Long R., Sebo J., Butlin P., Finlinson K., Fish K., Harding J., Pfau J., Sims T., Birch J., Chalmers D.: *Taking AI Welfare Seriously*. arXiv:2411.00986, (2024) 1–3.  
<https://doi.org/10.48550/arXiv.2411.00986>