



Frankfurt School

Szabados Levente

Frankfurt School of Finance and Management

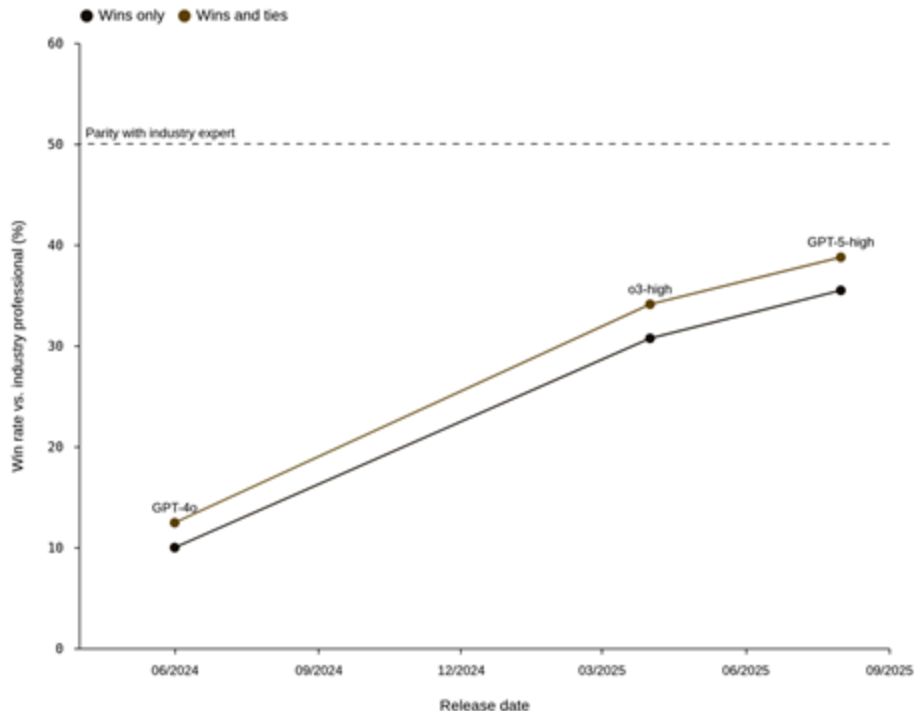
Soproni Egyetem LKK

Neuron Solutions

**Együttműködés AI ágensekkel -
Mi kell a bizalomhoz?**

TELJESÍTMÉNY ÉS HASZNÁLAT

GDPval: OpenAI frontier model performance over time



The share of U.S. workers who use ChatGPT for work has grown from 8% in early 2023 to 28% today

% of U.S. adults who say they have ever used ChatGPT...

For tasks at work
(among employed adults)



To learn something new



For entertainment



Note: "Employed" refers to those working full or part time for pay at the time of the survey. Those who did not give an answer are not shown.

Source: Survey of U.S. adults conducted Feb. 24-March 2, 2025.

PEW RESEARCH CENTER

MIT NEM TUDNAK AZ “LLM”-EK? - PÉLDÁUL A TÉNYEKET

When did France gift Lithuania Vilnius TV tower?



France gifted Lithuania the Vilnius TV tower in 1980.

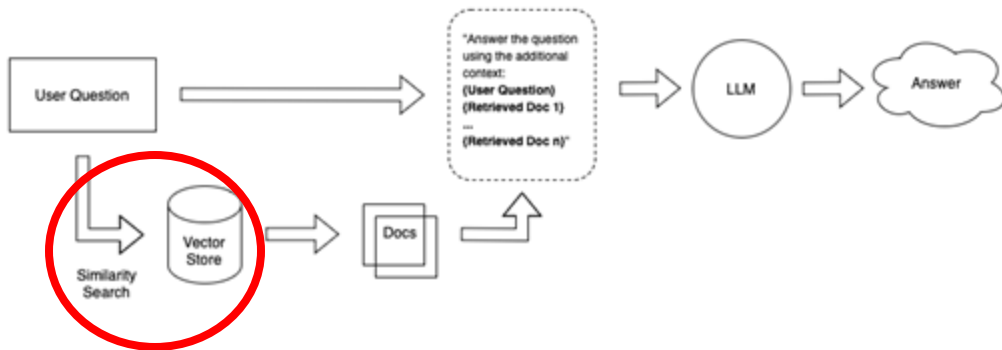
What prompted the French government to provide the gift?

The French government provided the gift as a gesture of friendship and solidarity with the Lithuanian people, who had recently declared independence from the Soviet Union.

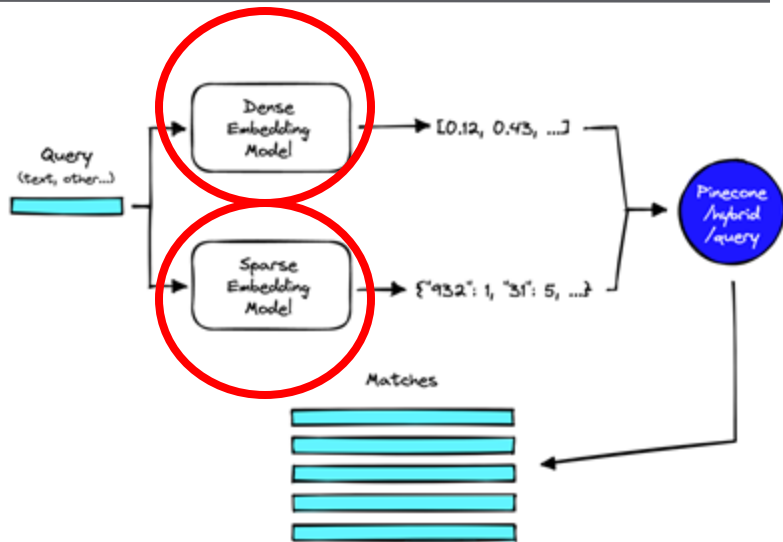
In reality, France had nothing to do with the construction of the Vilnius TV tower. The object was completed in 1980. Image by Cybernews.

[Zhang et al. 2023: Siren's Song in the AI Ocean:
A Survey on Hallucination in Large Language Models](#)

KERESD A TUDÁST, FOGALMAZD A VÁLASZT!



Kép forrása: [Improving Document Retrieval with Contextual Compression](#)



Kép forrása: [Getting Started with Hybrid Search](#)

[Guu et al. 2020](#)

[REALM: Retrieval-Augmented Language Model Pre-Training](#)

KIÉRTÉKELÉSI MÓDSZERTANOK



[Es et al. 2023](#)

[Ragas: Automated Evaluation of Retrieval Augmented Generation](#)



Figure 1: **Overview of ARES**: As inputs, the ARES pipeline requires an in-domain passage set, a human preference validation set of 150 annotated datapoints or more, and few-shot examples of in-domain queries and answers (five or more), which are used for prompting LLMs in synthetic data generation. To prepare our LLM judges for evaluation, we first generate synthetic queries and answers from the corpus passages. Using our generated training triples and a constrastive learning framework, we fine-tune an LLM to classify query–passage–answer triples in three different criteria: context relevance, answer faithfulness, and answer relevance. Finally, we use the LLM judges to score RAG systems and generate confidence bounds for the ranking using PPI and the human preference validation set.

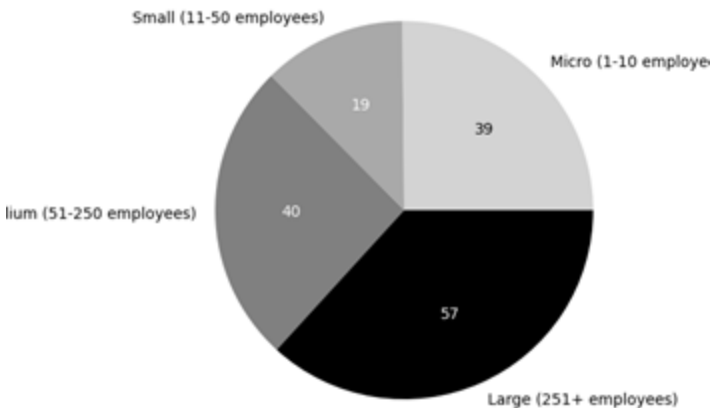
[Saad-Falcon et al. 2023](#)

[ARES: An Automated Evaluation Framework for Retrieval-Augmented Generation Systems](#)

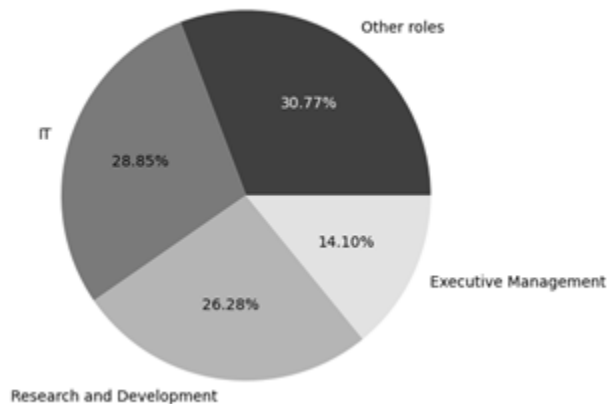
Empirikus mérés kérdőívvel - Demográfia

- Több mint 150 résztvevő
- Magyarországra és a DACH-régióra fókuszálva
(a résztvevők 91,3%-a ebből a két régióból származik)
- Toborzás hólabda-módszerrel
- Önkéntes (önszelekciós) torzítás!

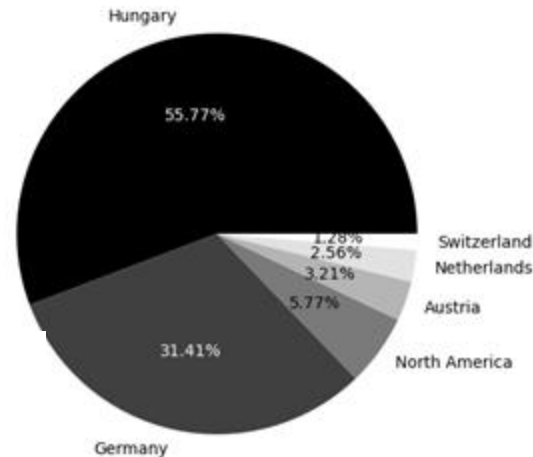
Organizational sizes of respondents



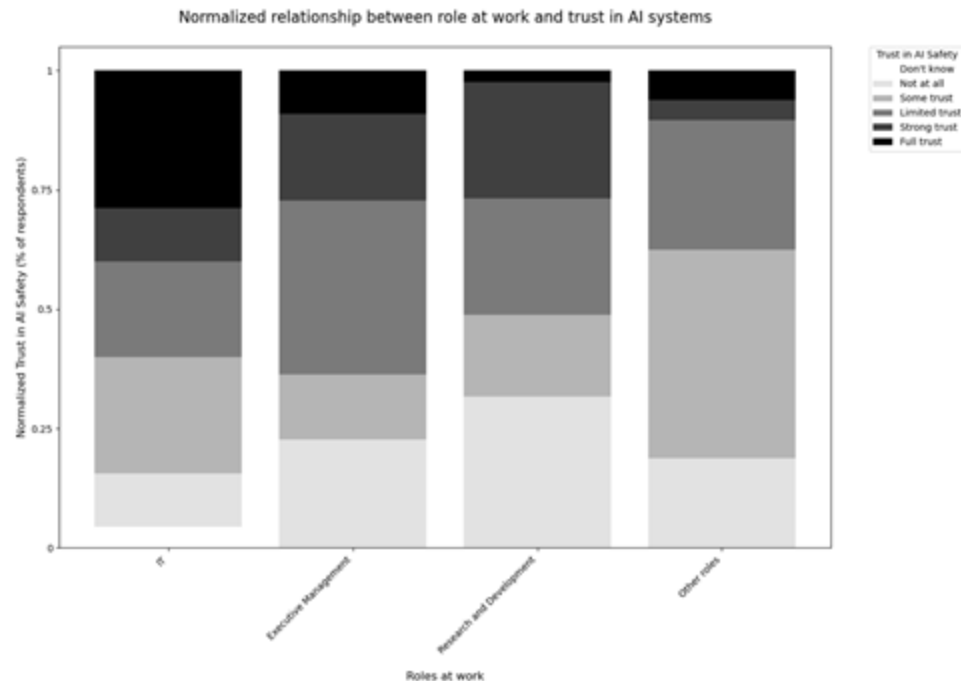
Organizational roles of respondents - Merged



Country distribution of respondents

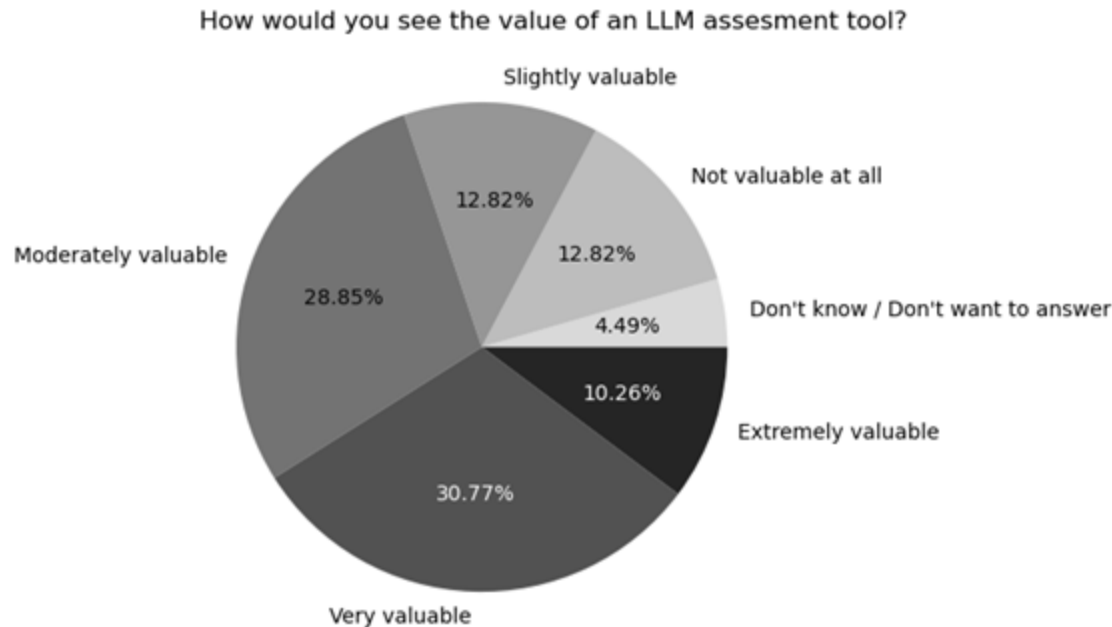
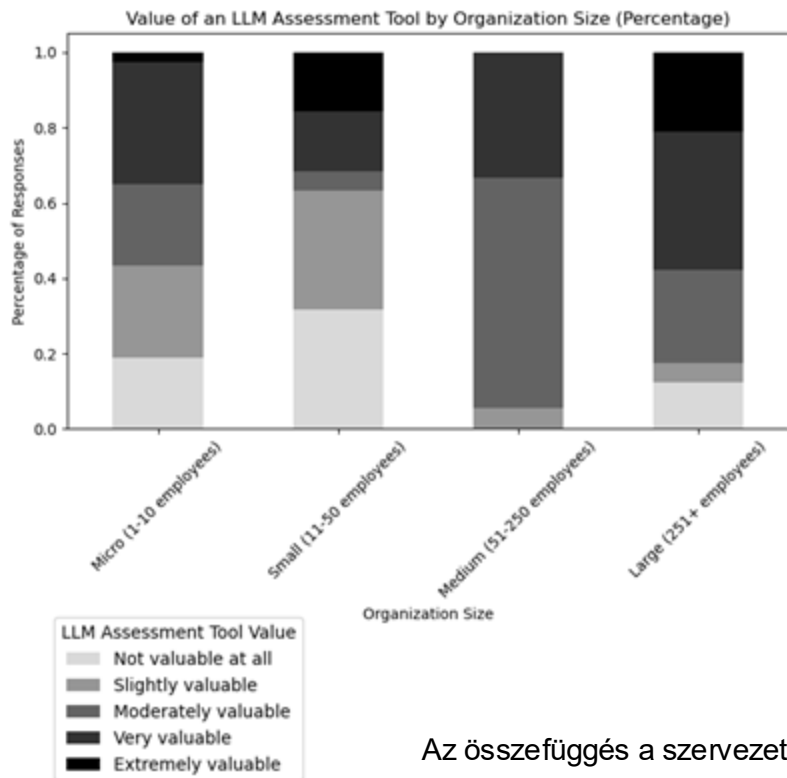


Bizalom vs használat



Minél többet használsz, annál jobban megbízol benne, de a szerepkörök között különbségek vannak!
(Lehetséges zavaró tényező: a használat mértéke szerepkörönként → nem szignifikáns eredmény)

Egy RAG rendszer értékelő módszer haszna



Az összefüggés a szervezet mérete és az LLM-értékelő eszköz észlelt értéke között statisztikailag szignifikáns.

Cramér-féle V: 0,3469 (mérsékelt kapcsolat)

“GONDOLKODJUNK LÉPÉSRŐL LÉPÉSRE”

“CHAIN OF THOUGHT” ÉS ESZKÖZ HASZNÁLAT

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓

HotPotQA

Question: Seven Brief Lessons on Physics was written by an Italian physicist that has worked in France since what year?

(d) ReAct

Thought 1: I need to search Seven Brief Lessons on Physics, find its author, then find when the author has worked in France since.

Act 1: Search[Seven Brief Lessons on Physics]

Obs 1: Seven Brief Lessons on Physics (Italian: Sette brevi lezioni di fisica) is a short book by the Italian physicist Carlo Rovelli. Originally published in Italian in...

Thought 2: The author of Seven Brief Lessons on Physics is Carlo Rovelli. I need to search Carlo Rovelli next and find when he has worked in France since.

Act 2: Search[Carlo Rovelli]

Obs 2: Carlo Rovelli is an Italian theoretical physicist and writer who has worked in Italy, the United States and, since 2000, in France.[1] He is also currently a Distinguished Visiting Research Chair at the Perimeter Institute...

Thought 3: Carlo Rovelli has worked in France since 2000. So the answer is 2000.

Act 3: Finish[2000] ✓

[Kojima et al. 2023](#)

[Large language models are zero shot reasoners](#)

[Wei et al. 2023](#)

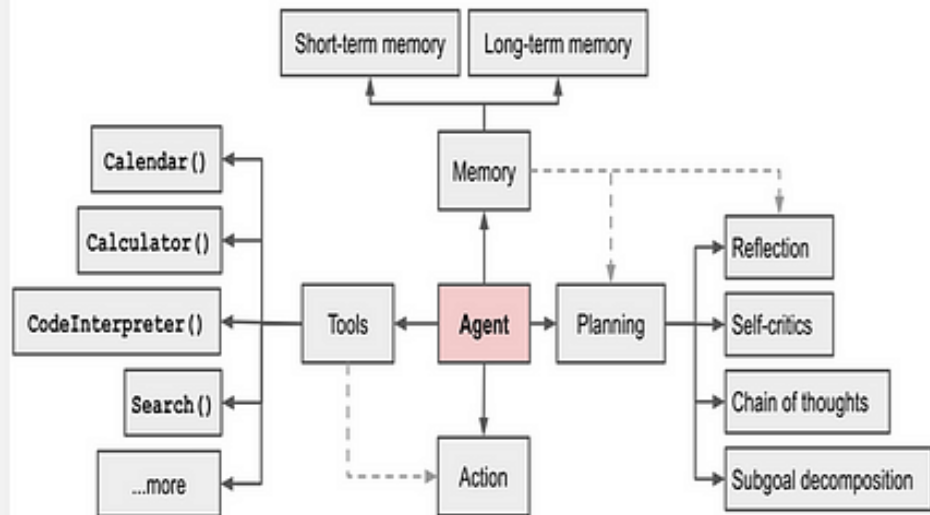
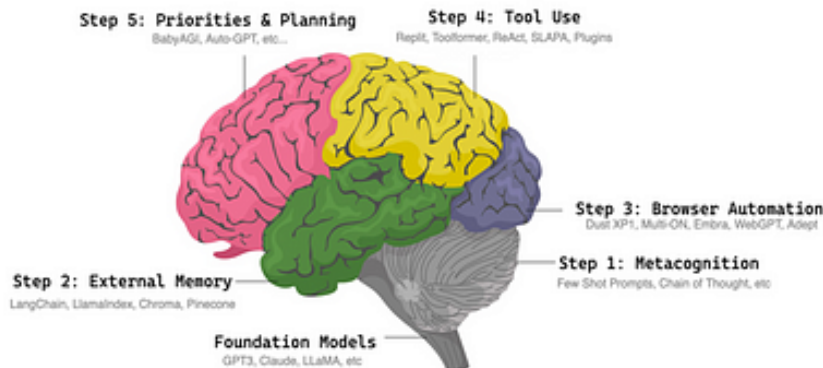
[Chain-of-Thought Prompting Elicits Reasoning](#)

[Yao et al. 2023](#)

[ReAct: Synergizing Reasoning and Acting in Language Models](#)

KOGNITÍV ARCHITEKTÚRÁK UTÁNZÁSA LLM ÁGENSEKKEL

The Anatomy of Autonomy



Néhány forrás:

[The Anatomy of Autonomy: Why Agents are the next AI Killer App after ChatGPT](#)
[LLM Powered Autonomous Agents](#)
[The Rise of Autonomous AI Agents: Debundling the Market Landscape](#)

A MÉRÉS PROBLÉMÁI ÁGENCIA ESETÉN

1. Többlépéses feladatok komplexitása

- Az agentikus rendszerek több lépésben, részben rejtett környezetben működnek.
- Az értékelés nem korlátozódhat egyetlen bemenet–kimenet párosra.
- A feladatok heterogének: tervezés, eszközhasználat, memória, önreflexió is szerepel bennük.

2. Kredit-hozzárendelés a közbenső lépésekhez

- Nem csak a végső eredményt kell mérni, hanem a döntési folyamat minőségét is.
- Nehéz megállapítani, mely lépések járultak hozzá a sikerhez vagy a hibához.
- Fontos a részpontozás és a folyamatlogika elemzése.

3. Stochaszticitás és reprodukálhatóság

- Az ügynökök viselkedése véletlenszerű lehet.
- Ugyanaz a feladat más eredményt adhat különböző futásokban.
- Az értékelésnek robusztusnak kell lennie, több futásból kell átlagolni.

4. Biztonság és nem kívánt viselkedés

- Az agentikus rendszerek képesek nem kívánt, veszélyes vagy etikátlan lépésekre.
- Az értékelésnek ki kell szűrnie és büntetnie kell a szabályszegéseket.
- A biztonság és megbízhatóság külön értékelési dimenziót igényel.

5. Megfigyelhetőség és naplózás

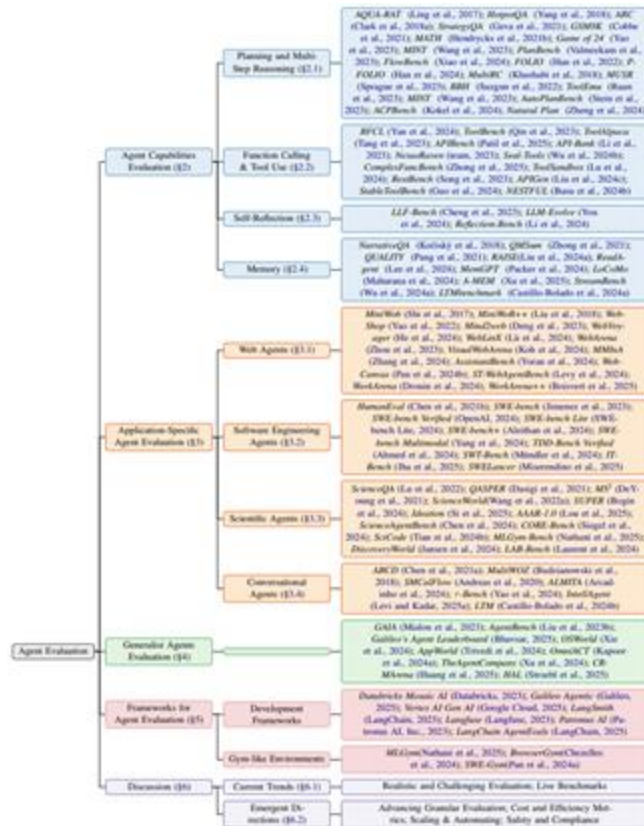
- Az ügynök belső folyamatait (döntések, eszközhívások, állapotváltozások) rögzíteni kell.
- A „fekete doboz” működés akadályozza a részletes elemzést.
- Strukturált naplózás és átlátható trace szükséges.

6. Költség és skálázhatóság

- Az értékelés drága: sok lépés, nagy számítási igény, humán annotáció.
- Automatizált vagy félautomatikus módszerekre van szükség.
- A hatékonyság és költséghatékonyság önálló mérőszámként jelenik meg.

ÚJ MÓDSZERTANOK ALKALMAZANDÓK!

- Számos új módszertan jelent meg
- Domain függő kiértékelések dominálnak
- Végső konszenzus még nincs
- Pragmatikus használat lehetséges



Frankfurt School of Finance and Management gGmbH

Adickesallee 32-34

60322 Frankfurt am Main

Levente Szabados

Dozent

E-Mail: L.Szabados@int.fs.de

www.frankfurt-school.de



[linkedin.com/in/levente-szabados-ai](https://www.linkedin.com/in/levente-szabados-ai)



[linkedin.com/company/frankfurtschool](https://www.linkedin.com/company/frankfurtschool)

