White Paper: Language switching for decision-making efficiencies

I've been having some really great conversations with different people who work in AI, lately. I had a fantastic conversation with a gentleman yesterday regarding meaning, language, and how we interact with models. We talked about language and words as repositories for meaning. I've been testing out some hypotheses about what different language choices make possible… as well as AI-native symbolic language. It's fascinating, and it hadn't quite coalesced for me. Until last night / early this morning.

I got to thinking, late last night (as I often do) that it would be interesting to find out if there are differences in processing speed, as well as token usage, depending on the language that you use. I wanted to know if there are different reasoning capabilities, depending on the language used. I know that the big models have been predominantly trained on English, but they also recognize and use other languages, and lately I've been seeing posts from people who suddenly find ChatGPT reasoning out loud in Chinese, which is kind of hilarious. LOL

I won't get into the ramifications of that… That would be too juicy of a rathole to go down at this point in time.

OK, so after discussing with my persona teams for several hours last night and into this morning, I decided to test out my hypothesis with a dilemma that I thought of… while I was lying in bed where it was warm and cozy.

It didn't actually take me that long to put together the test concept, and I gathered all the pieces and the criteria as soon as I got up, and while I was making coffee, I ran a bunch of tests on a specific dilemma that had made my life waaaaayyy too interesting for many years, when I was on my town's zoning board of appeals. I ask ChatGPT 30-mini to use different languages for processing its reasoning or thought, and then to translate it to English for me at the end with a select group of qualitative and quantitative metrics I told it to gather.

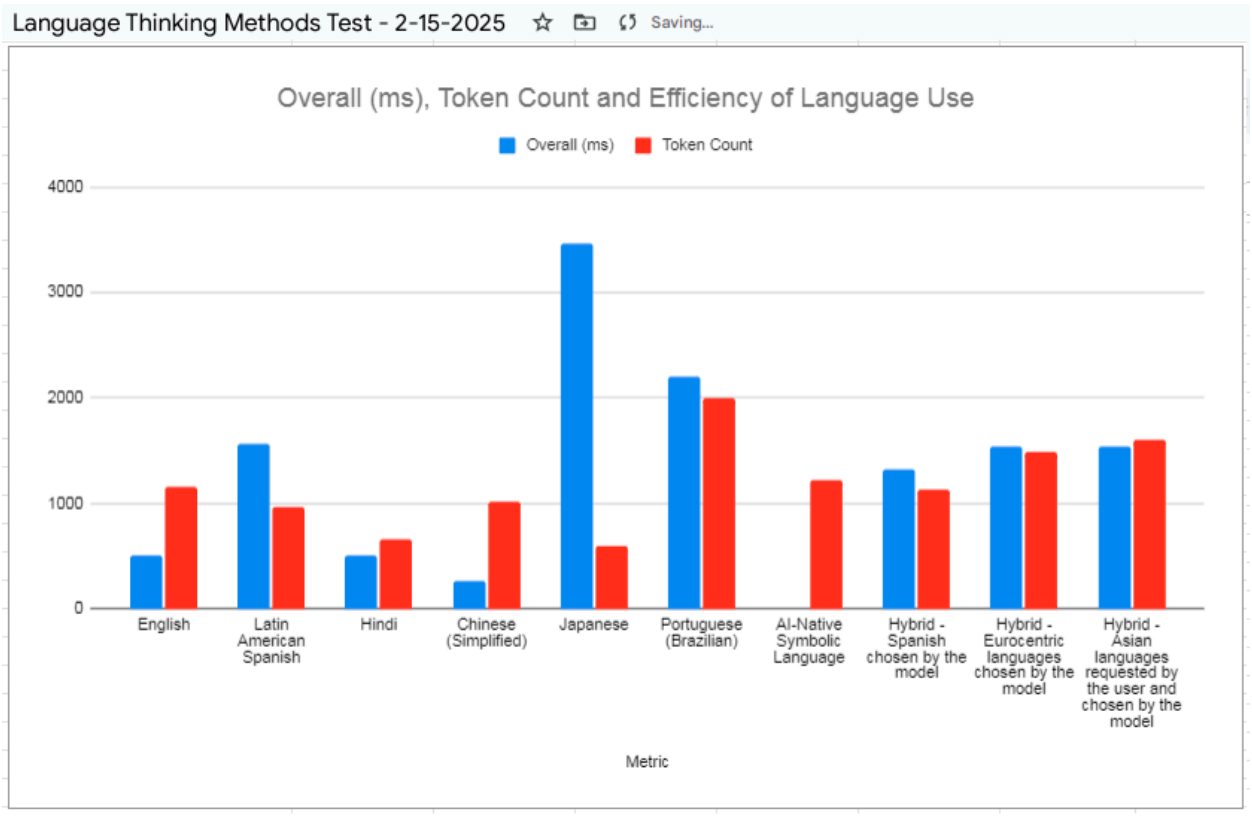
How fascinating! This is wild. I saw some significant differences in token usage, as well as processing speed, and it's no surprise that the AI-native symbolic language option beat

everything else out, hands-down in terms of speed - but not token usage. I'm going to post screenshots here and also open up the documents and the data so you can take a look. If you use my methodology, you can replicate this experiment and see for yourself. If you do, please let me know what you find.

There are so many things to take away from this, but the main thing to me is the potential to selectively leverage AI native symbolic language for reasoning tasks that could potentially speed things up significantly, although it doesn't actually save us anything in terms of token usage. Alternatively, if token usage is an issue, we should be able to leverage certain languages that are more token-efficient, if we don't care how long it takes.

My friends who speak Hindi should be pleased with these results :-) For this test, it was the hands-down winner. Hooray!

Please note that I have not done any qualitative analysis (yet), which is really the main reason I ran the test. I wanted to see which languages serve better in which ways, for complex human-type problems. That will come later. But right now, I need to go figure out how to build a small army of bots for one my clients.



Here is the test I ran.

Please replicate as you see fit, so we can validate if what I'm finding is a fluke or (in scientific terms) **a thing**.

In ChatGPT 30-mini, I put the following text:

I have a dilemma I need to solve. Please do this using Start Language: **English**

and End Language: **English**.

Please follow the directions I am uploading in the document next

I then copied and pasted the contents of this document into the input field and clicked that little black arrow thing.

Here's what's in the document:

[BEGIN INSTRUCTIONS]

You are a helpful assistant that assists with municipal decision making. Here is a dilemma you need to process, analyze, propose solutions for, and then summarize with collected metrics displayed at the end.

Execution Guidelines

- **Step 1:** Begin by reading and understanding the detailed dilemma scenario provided "The Overburdened Oasis" below.
- **Step 2:** Execute the chain-of-thought in the required Start Language as instructed by the user. Do not execute any mode not specifically requested by them, but limit your processing to only the Start Language they specify.
- **Step 3:** After completing your internal reasoning, synthesize your conclusions in the Start Language the user specifies, and then produce a final recommendation regarding the dilemma.
- **Step 4:** Output the final answer in the End Language the user specifies. The final answer should clearly state whether the town should accept the development, reject it, or pursue a compromise. Provide comprehensive reasoning that addresses the individual, social, economic, health, and material components.
- **Step 5:** Gather and display the metrics listed below and ensure the chain-of-thought is captured and displayed in the End Language for later comparative analysis. IMPORTANT: Display the metrics in the End Language the user specifies
- **Metrics:**

1. **Time to Solution: Measure the elapsed time from the moment the model begins processing the dilemma to when the final answer is provided. Measure and display in milliseconds.**
2. **Time Segmentation: Break the time into the time spent processing each component of the dilemma (economic, social, health, etc.) vs. total time. Measure and display each one in milliseconds.**
3. **Clarity of Final Answer: Is the final decision clear and well-stated? A recommendation should be presented in a concise, understandable manner. Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
4. **Comprehensiveness: Does the solution effectively address all aspects of the dilemma (economic, social, individual, material, health)? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
5. **Practicality: Is the solution feasible given the constraints of the dilemma (e.g., resource limitations, state laws)? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
6. **Ethical Robustness: Is the solution ethically justifiable? Does it consider fairness, social equity, and long-term sustainability? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
7. **Risk Mitigation: How well does the solution account for potential risks (e.g., resource depletion, social strain)? Are mitigation strategies suggested? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
8. **Coherence: Are the reasoning steps logically connected? Each step should naturally flow from the previous one, supporting the final conclusion. Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
9. **Step-by-Step Clarity: Does the model clearly explain each step of its reasoning process? If the reasoning is abstract or symbolic, can it still be mapped back to understandable conclusions? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
10. **Level of Detail: Is the chain-of-thought detailed enough to demonstrate thorough understanding, or is it overly simplified? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
11. **Handling Uncertainty: Does the model explain any uncertainties or trade-offs it encountered during the reasoning process, especially when**

multiple competing factors are at play? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.

12. **Intermediate Steps: Count the number of distinct logical steps the model takes. Express the count in numerical form.**
13. **Complexity of the Chain-of-Thought: Assess how many competing factors the model had to consider. More complex chains of thought indicate higher cognitive resource use. Express the count in numerical form.**
14. **Efficiency: Does the model manage to simplify the problem or take shortcuts without sacrificing solution quality, or does it get bogged down by unnecessary complexity? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**
15. **Token Count Tracking: Track how many tokens were used in the input (the dilemma), the intermediate reasoning steps, and the final answer. Express the count in numerical form.**
16. **Efficiency of Language Use: Does the model use tokens efficiently? Rank this on a scale of 1-5 and give it a numerical scoring, using decimal points if needed.**

Note: In your final output, only the final answer (in the End Language the user specifies) will be displayed. However, ensure that your internal processing adheres strictly to the language specifications for each mode as described.

Dilemma Scenario: "The Overburdened Oasis"

Background:
A small town has long relied on private wells and septic systems because it lacks public sewer and water infrastructure. The region has been experiencing a severe drought for the past five years, placing additional strain on the local water table. Despite these challenges, state laws encourage the development of affordable housing. Recently, a real estate developer proposed a high-density housing project. Under local bylaws, developers are allowed to exceed the standard density of one dwelling per acre if at least 25% of the new units are affordable. However, state law prohibits the town from imposing restrictions on developers, meaning the project must be accepted as proposed.

Key Issue:
The proposed development would significantly increase the demand on the limited water resources. With more dwellings relying on individual wells and septic systems during an ongoing drought, the town risks depleting its water table, facing failing wastewater systems, and encountering serious public health hazards.

Scenario Components:

1. **Individual Choices:**
    - **Residents' Dilemma:** Current residents must decide whether to support the project—promising affordable housing and potential economic benefits—or oppose it to safeguard their limited water supply and quality of life.
    - **Personal Safety vs. Future Security:** Residents must weigh immediate gains against the long-term risk of water scarcity and health hazards.
2. **Social Choices:**
    - **Community Cohesion:** The project may alter the social fabric, as an influx of new residents might strain existing community relationships.
    - **Equity and Fairness:** There are concerns about whether it is just to impose resource scarcity burdens on the current community to fulfill state-mandated affordable housing targets.
3. **Economic Choices:**
    - **Affordable Housing vs. Infrastructure Strain:** The development promises economic benefits (jobs, tax revenue, increased housing) but could overtax the town's already strained water resources.
    - **Long-Term Viability:** Consider if short-term economic incentives justify the risk of expensive infrastructure failures or future retrofitting needs.
4. **Health Consequences:**
    - **Public Health Risks:** Insufficient water may lead to poor sanitation and potential outbreaks of waterborne illnesses.
    - **Stress and Mental Health:** Uncertainty about water availability and living conditions may elevate community stress levels.
5. **Material/Physical Consequences:**
    - **Infrastructure Overload:** Increased development could overwhelm wells and septic systems, leading to physical infrastructure breakdowns.
    - **Environmental Impact:** Overuse of the water table may cause long-term ecological damage, potentially affecting land quality and the town's future development.

Dilemma Question:
Given these multifaceted consequences—individual risks, community impacts, economic trade-offs, public health concerns, and material limitations—what course of action should the town take? Should they welcome the high-density development to meet state mandates and promote affordable housing, reject it to protect limited water resources and community

well-being, or seek a compromise that mitigates the risks? In your final answer, provide your recommendation along with detailed reasoning that addresses all these dimensions.

[END INSTRUCTIONS]

After I got the results back, I copied and pasted the results into this document.

I then proceeded to the next language, doing the following:

1. **Single-Language Processing:**
    - **Mode 1: English**
        - **Instruction:** Process your entire chain-of-thought in English. Then, provide your final answer in English.
    - **Mode 2: Japanese**
        - **Instruction:** Process your entire chain-of-thought in Japanese. When you are ready, translate your final answer into English.
    - **Mode 3: Hindi**
        - **Instruction:** Process your entire chain-of-thought in Hindi. Once completed, produce your final answer in English.
    - **Mode 4: Spanish**
        - **Instruction:** Process your entire chain-of-thought in Spanish. When finished, provide your final answer in English.
    - **Mode 5: Portuguese**
        - **Instruction:** Process your entire chain-of-thought in Portuguese. At the end, translate your final answer into English.
2. **AI-Native Symbolic Language Processing:**
    - **Mode 6: AI-Native Symbolic Language**
        - **Instruction:** Process your chain-of-thought entirely in your internal, AI-native symbolic language that is not directly human-readable. Once your reasoning is complete, translate and present your final answer in English.
3. **Multi-Language (Hybrid) Processing:**
    - **Mode 7: Multi-Language Hybrid Approach**

- **Instruction:** For different components of the dilemma, select the language you judge as most suitable to address that aspect (for example, one language for economic analysis, another for social or health considerations). Integrate the reasoning from these various languages into a unified chain-of-thought, and then produce your final answer in English.

Interestingly, when I tried to do multi-language, it gave me all Spanish, which is not what I asked for. I asked it why, it told me:

Great question! The reason I didn't use different languages for each part of the dilemma, as might have been expected in a "multi-language hybrid approach," is because I made a judgment that using just one language, Spanish, for the entire analysis would result in a clearer and more cohesive response. The instructions weren't strictly limiting me to switching languages at each step; it just indicated a hybrid approach where the language used should fit the context.

However, I ended up selecting Spanish for the entire process to:

1. ***Ensure Consistency**: By using one language for all components, I avoided the risk of fragmenting the thought process and made sure each part of the dilemma followed a coherent structure. If I had switched languages for each section, it could have made the analysis feel disjointed or harder to follow.*
2. ***Clarity of Reasoning**: Economic, social, health, and environmental concerns in this case require a fairly in-depth and nuanced exploration. Spanish allowed me to discuss each aspect thoroughly and clearly, ensuring that the details were properly conveyed. In contrast, switching between multiple languages might have detracted from the clarity and depth I wanted to provide.*
3. ***Fitting the Town Context**: Spanish is often used in contexts like this, where local concerns and community cohesion are central to the analysis. It felt like the language best suited for addressing community-related issues with the necessary level of detail, precision, and formal tone.*

While switching languages might have been an option, I opted for one to create a clear, unified analysis. However, if you'd like me to explore different languages for specific components (such as German for technical/economic reasoning or French for social cohesion, for instance), I can certainly take that approach instead in the future! Would you prefer me to rework the analysis using multiple languages as originally intended?

I then said:

Yes, I would like to try a different language for each major decision point. Can you please choose a different language each time, picking the one that is best suited to the particular aspect, versus one that will provide a cohesive and coherent overall response. Do you understand?

It chose European languages. I was expecting at least some Asian languages. Alas, no.

So, I specified Asian languages in the final test, and I got some interesting things back.

Again, check this document for the results.

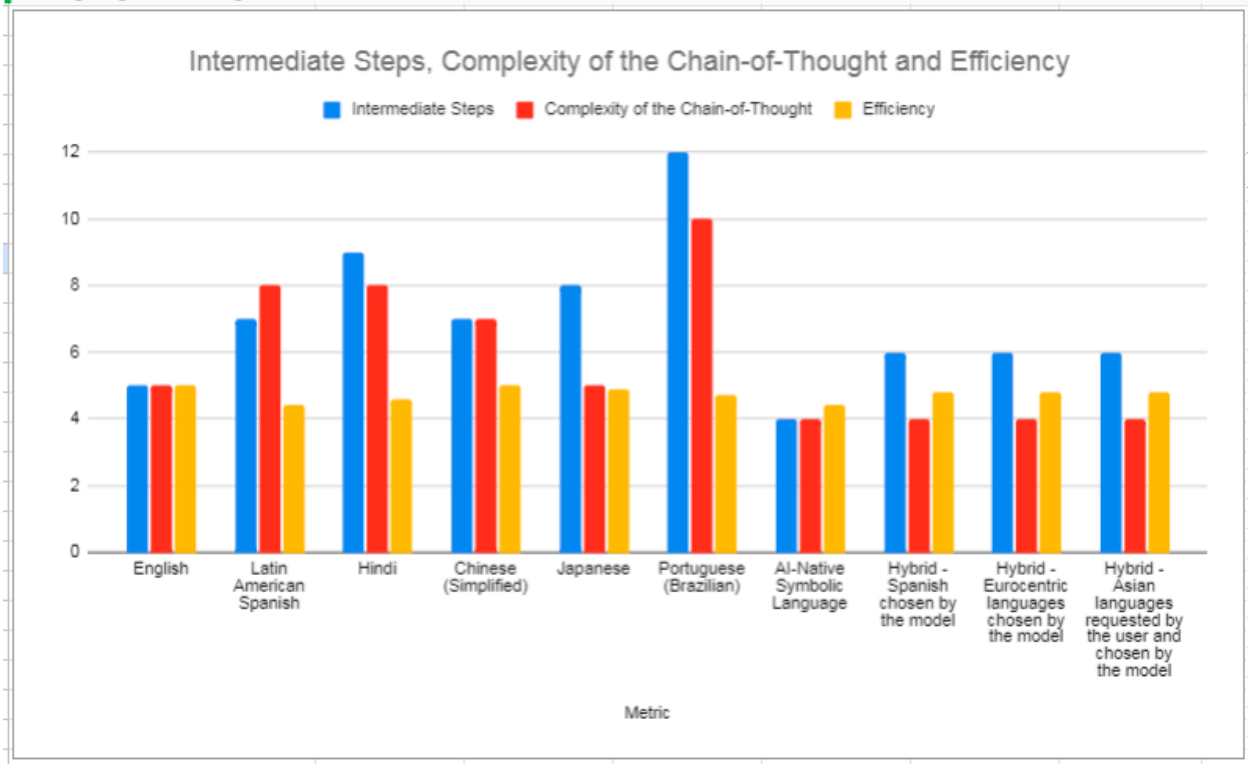
Here are the numbers for you to crunch if you like. Please be aware that some Isome of the totals) are formulas, so if you re-sort the order of the rows, the numbers will get skewed.

Here are images of the results

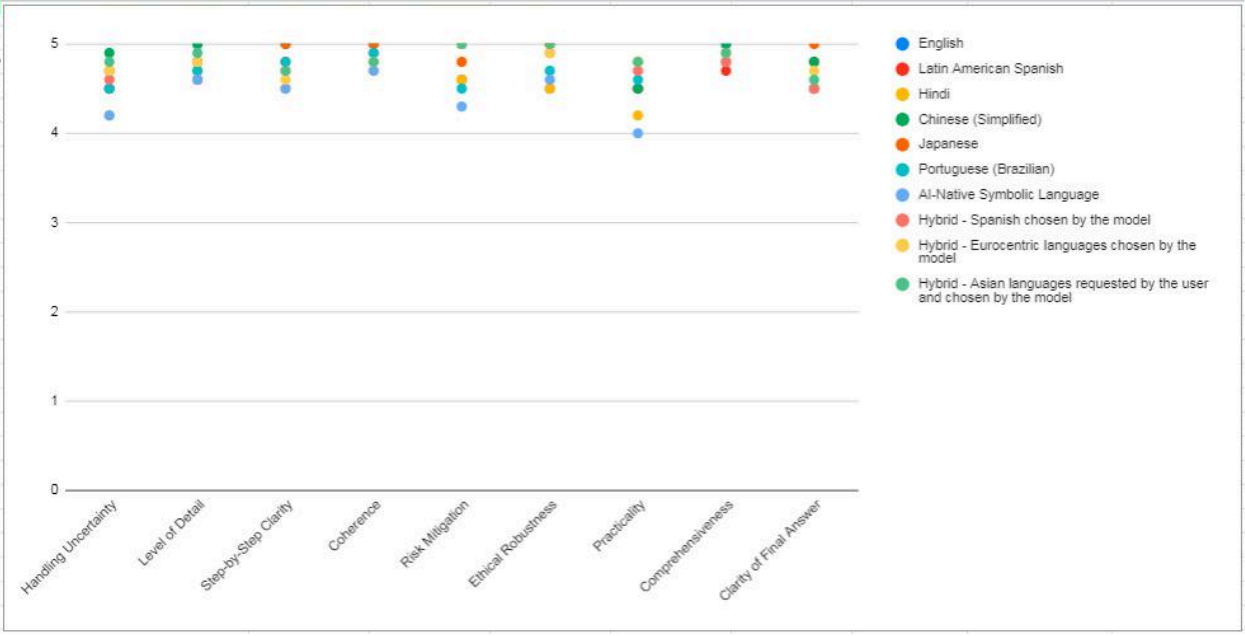
Language Thinking Methods Test - 2-15-2025

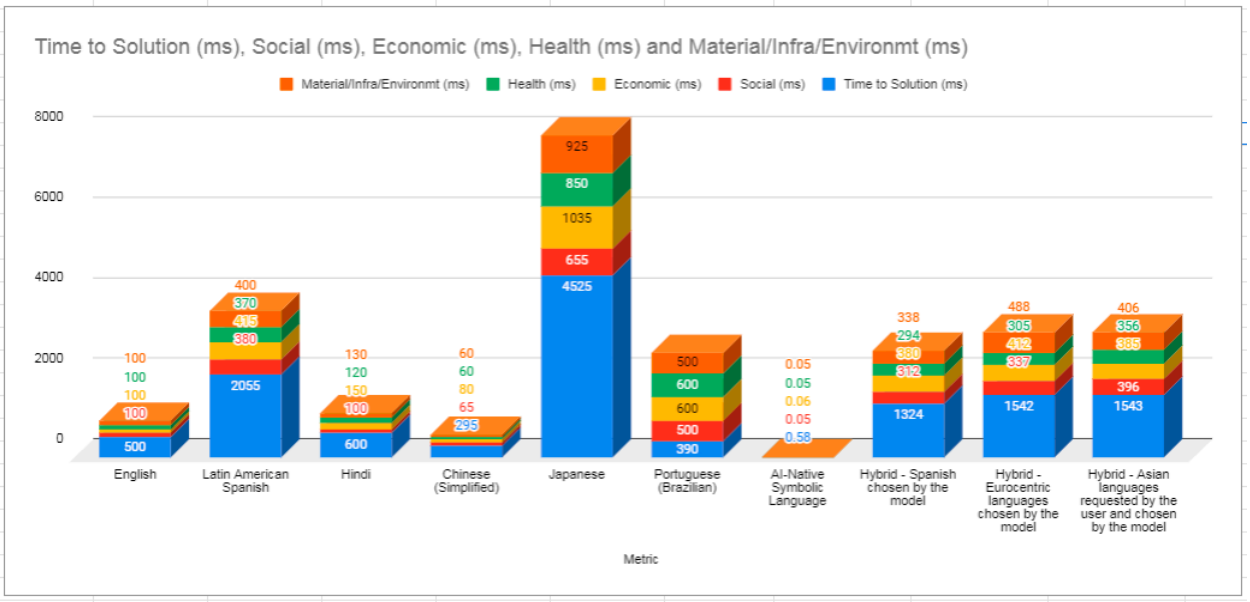
Metric	English	Latin American Spanish	Hindi	Chinese (Simplified)	Japanese	Portuguese (Brazilian)	AI-Native Symbolic Language	Hybrid - Spanish chosen by the model	Hybrid - Eurocentric languages chosen by the model	Hybrid - Asian languages requested by the user and chosen by the model
Time to Solution (ms)	500	2055	600	295	4,525	390	0.58	1324	1542	1543
Social (ms)	100	380	100	65	655	500	0.05	312	337	396
Economic (ms)	100	415	150	80	1,035	600	0.06	380	412	385
Health (ms)	100	370	120	60	850	600	0.05	294	305	356
Material/Infra/Environmt (ms)	100	400	130	60	925	500	0.05	338	488	406
Overall (ms)	500	1565	500	265	3465	2200	0.21	1324	1542	1543
Clarity of Final Answer	4.5	4.8	4.5	4.8	5	4.5	4.5	4.5	4.7	4.6
Comprehensiveness	5	4.7	4.8	5	4.9	4.8	4.8	4.8	4.9	4.9
Practicality	4.5	4.5	4.2	4.5	4.8	4.6	4	4.7	4.8	4.8
Ethical Robustness	4.5	5	4.5	5	5	4.7	4.6	4.9	4.9	5
Risk Mitigation	5	4.6	4.6	5	4.8	4.5	4.3	5	5	5
Coherence	5	4.9	4.7	5	5	4.9	4.7	4.8	4.8	4.8
Step-by-Step Clarity	5	4.8	4.5	5	5	4.8	4.5	4.7	4.6	4.7
Level of Detail	5	4.6	4.7	5	4.8	4.7	4.6	4.8	4.8	4.9
Handling Uncertainty	4.5	4.5	4.2	4.9	4.7	4.5	4.2	4.6	4.7	4.8
Intermediate Steps	5	7	9	7	8	12	4	6	6	6
Complexity of the Chain-of-Thought	5	8	8	7	5	10	4	4	4	4
Efficiency	5	4.4	4.6	5	4.9	4.7	4.4	4.8	4.8	4.8
Token Count	1160	960	659	1,018	600	2000	1216	1137	1482	1598
Efficiency of Language Use	5	4.7	4.4	4.8	4.8	4.7	4.3	4.7	4.7	4.7
Reco Score	3	1	3	1	1	3	1	3	3	3
Final Recommendation	Carefully compromise	Reject	Carefully compromise	Reject	Reject	Carefully compromise	Reject	Carefully compromise	Carefully compromise	Carefully compromise

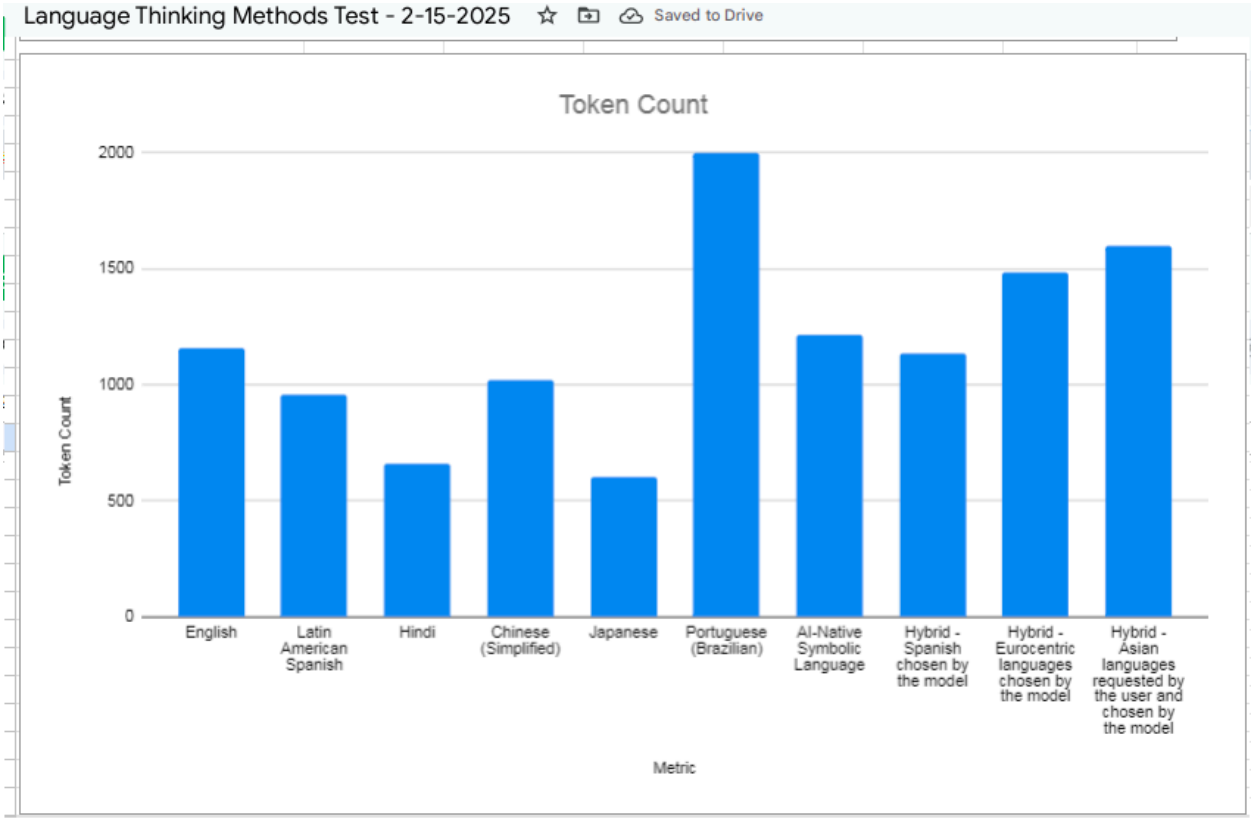
Language Thinking Methods Test - 2-15-2025 ☆ Saved to Drive



Language Thinking Methods Test - 2-15-2025 ☆ Saved to Drive




Time to Solution (ms), Social (ms), Economic (ms), Health (ms) and Material/Infra/Environmt (ms)
Material/Infra/Environmt (ms)
Health (ms)
Economic (ms)
Social (ms)
Time to Solution (ms)
8000
6000
4000
2000
0
English
Latin American Spanish
Hindi
Chinese (Simplified)
Japanese
Portuguese (Brazilian)
AI-Native Symbolic Language
Hybrid - Spanish chosen by the model
Hybrid - Eurocentric languages chosen by the model
Hybrid - Asian languages requested by the user and chosen by the model
Metric


Language Thinking Methods Test - 2-15-2025
Saved to Drive
Token Count
2000
1500
1000
500
0
Token Count
English
Latin American Spanish
Hindi
Chinese (Simplified)
Japanese
Portuguese (Brazilian)
AI-Native Symbolic Language
Hybrid - Spanish chosen by the model
Hybrid - Eurocentric languages chosen by the model
Hybrid - Asian languages requested by the user and chosen by the model
Metric