

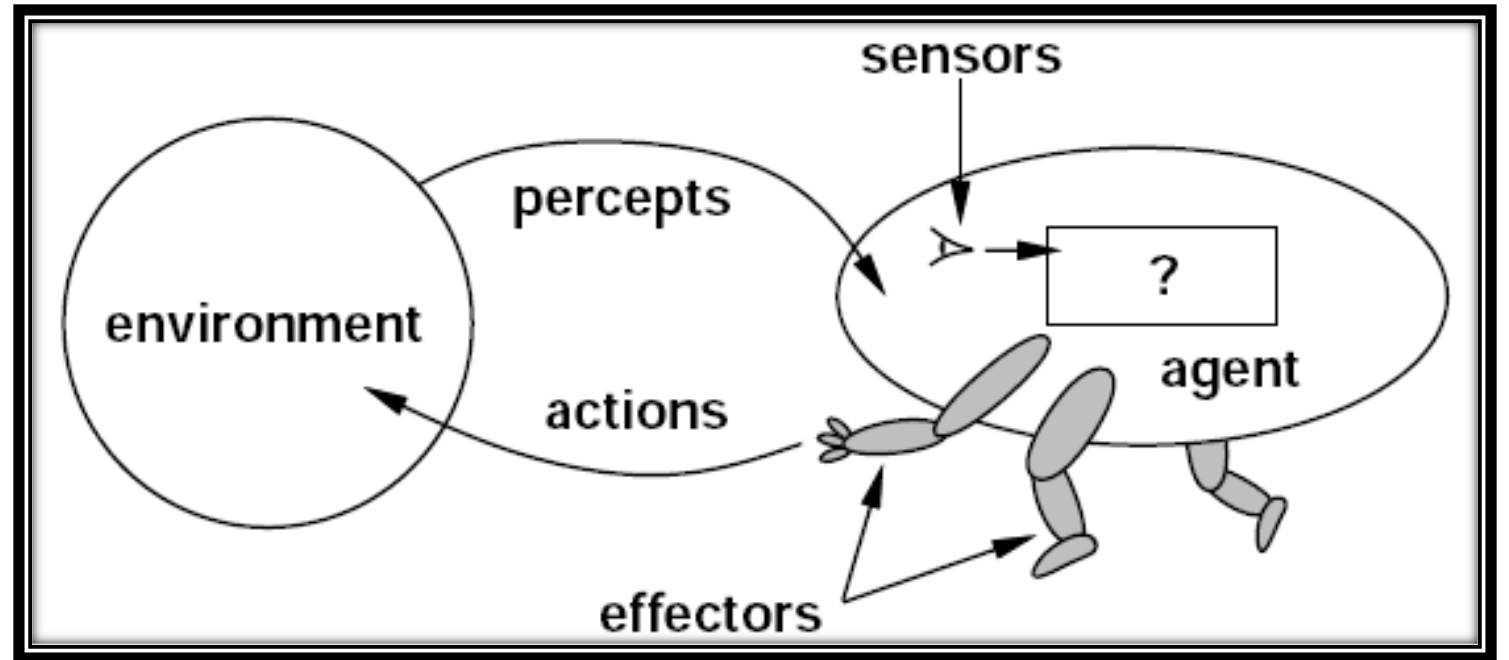
# Modelado basado en agentes para la simulación de fenómenos sociales

27/1/2025 SICSS

Marcelo Mendoza  
DCC UC, CENIA e IMFD

# ¿QUÉ ES UN AGENTE?

# Agentes

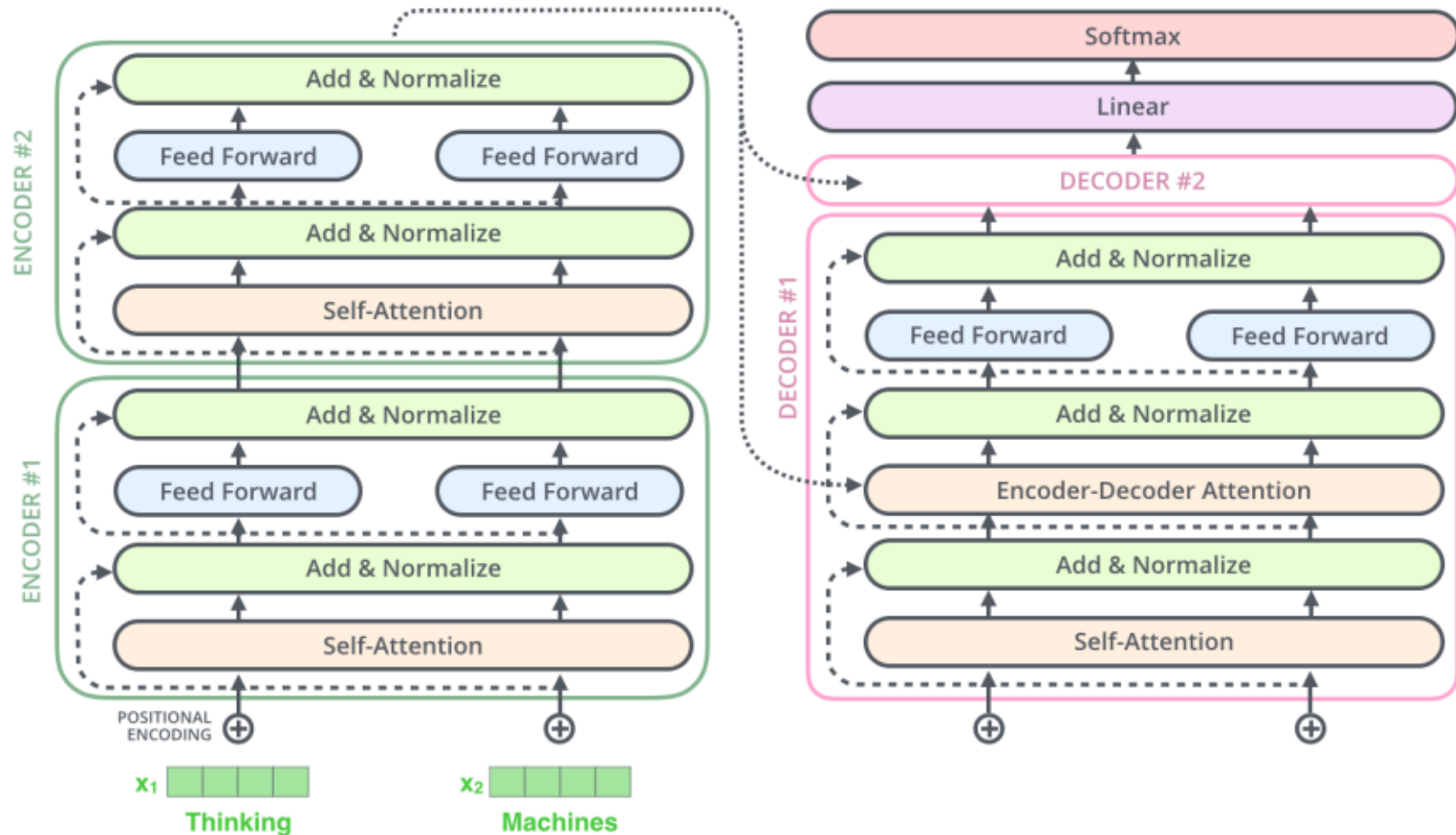


Un agente es una entidad que percibe el entorno, procesa información y toma acciones para lograr objetivos y metas específicas.

# LA ARQUITECTURA TRANSFORMER



## Stacked blocks, encoders y decoders del Transformer

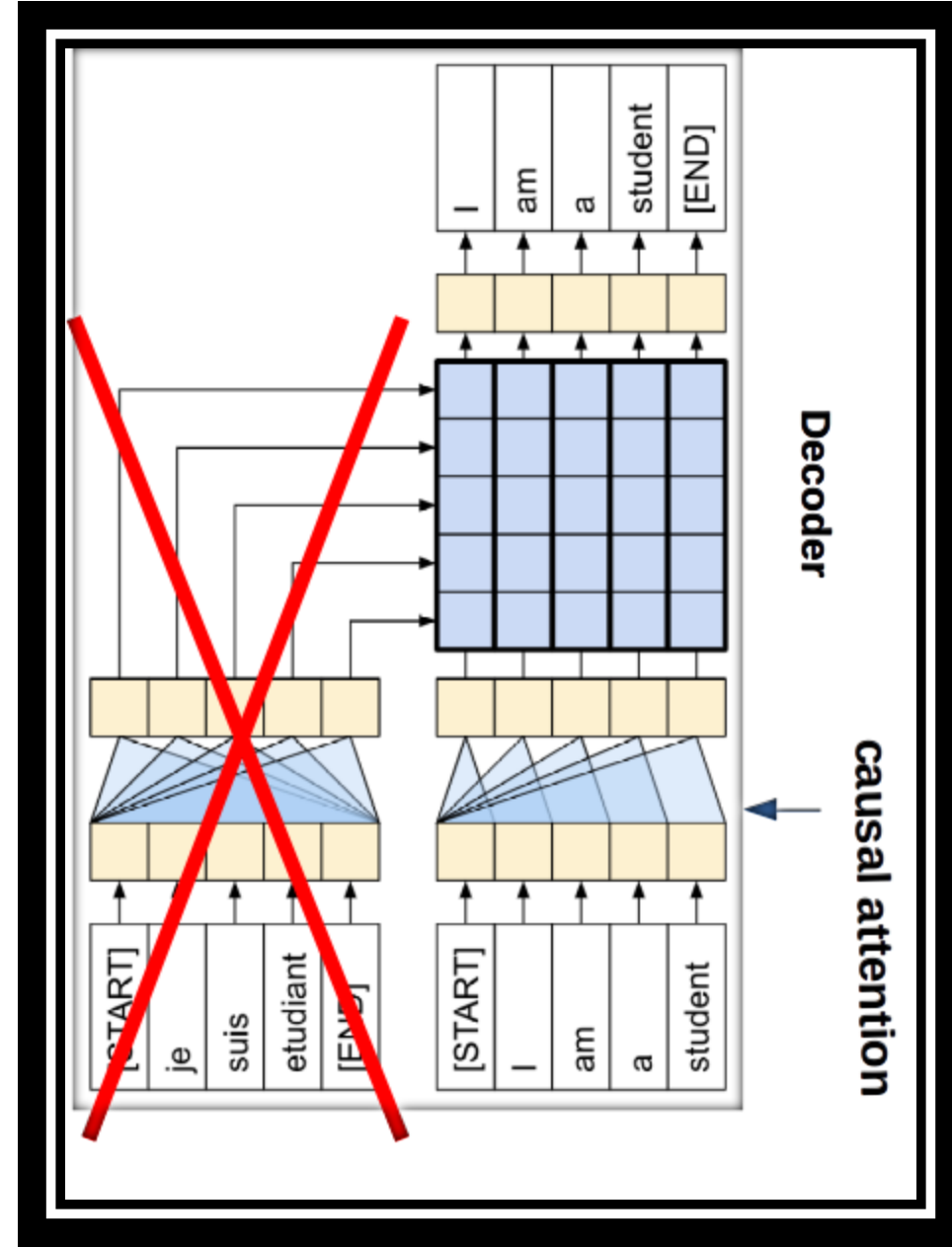




- **GENERATIVE  
PRETRAINED  
TRANSFORMER** -

# Transformer decoder

---



# Transformer decoder

Seq2seq

- Cada secuencia  $(m^1, \dots, m^n) \mapsto (y^1, \dots, y^n)$

se transforma en:

→ Símbolo separador

$$(w^1, \dots, w^{n+\eta+1}) = (m^1, \dots, m^n, \delta, y^1, \dots, y^n)$$

- Luego, el **decoder** resuelve la siguiente tarea:

secuencias largas

$$p(w^1, \dots, w^{n+\eta}) = \prod_{j=1}^{n+\eta} p(w^j | w^1, \dots, w^{j-1})$$

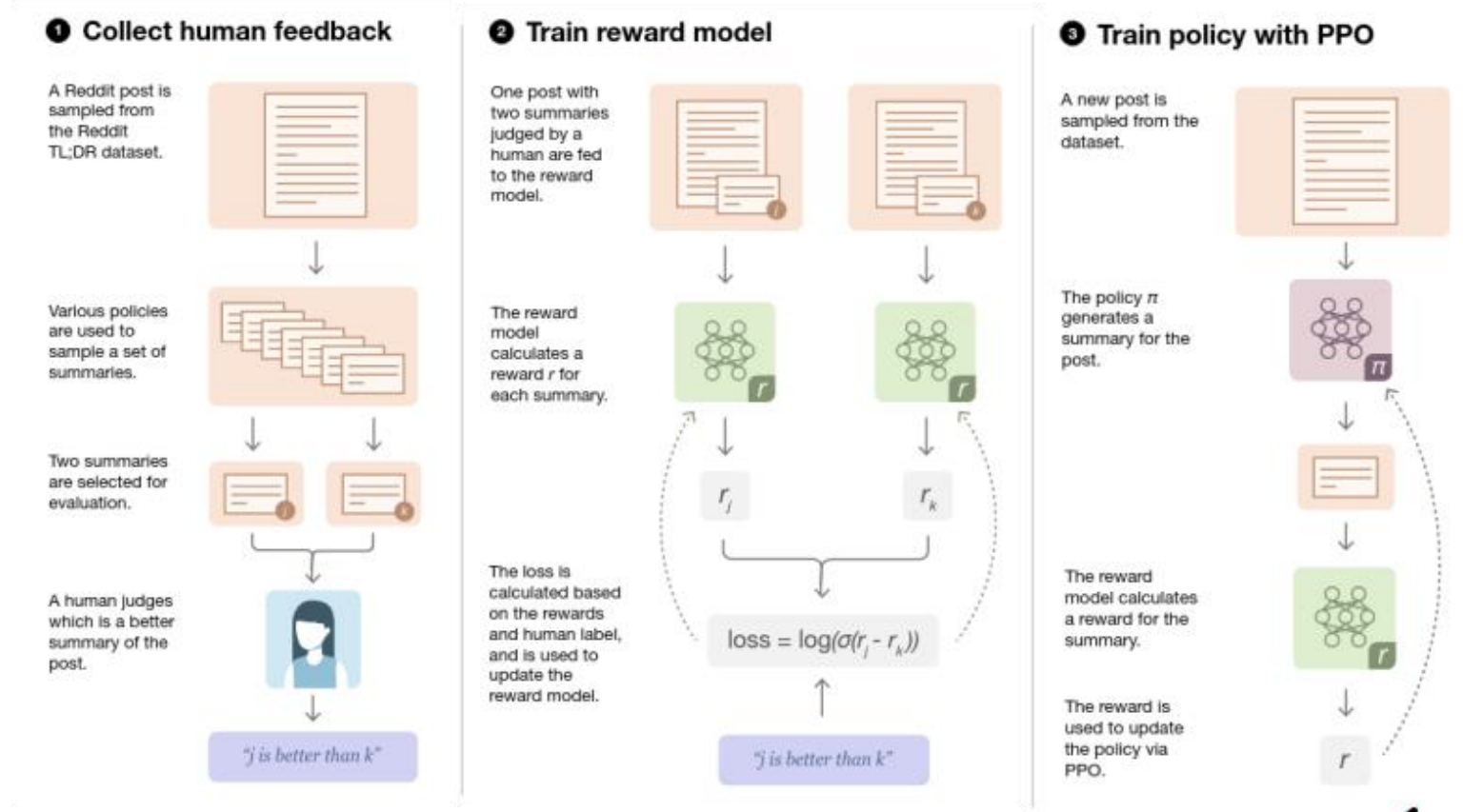


# MODEL ALIGNMENT



## Human in the loop

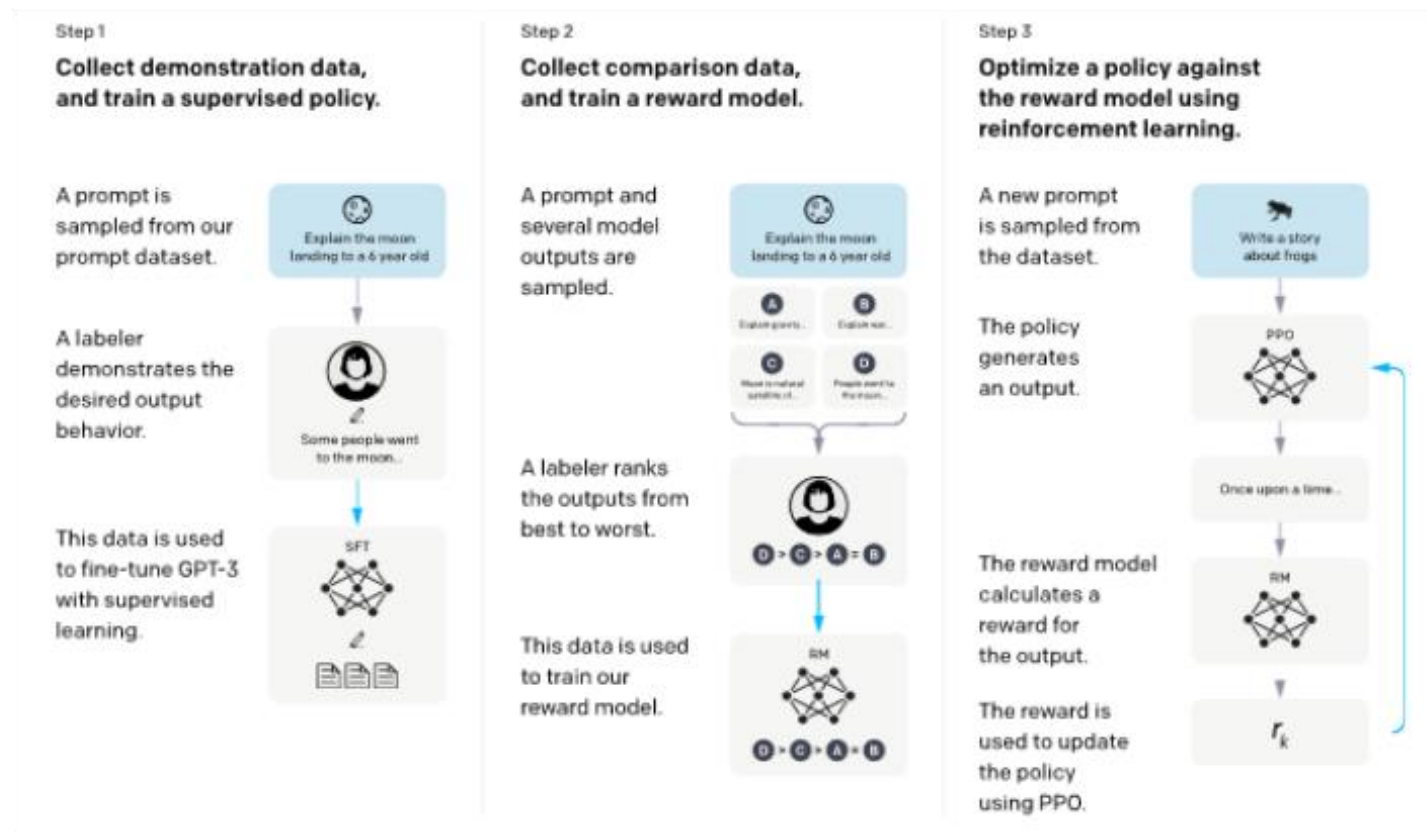
Inicialmente este tema se abordó para construcción de resúmenes



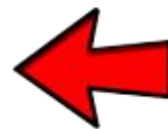
Stiennon et al. Learning to summarize from human feedback, NeurIPS 2020



## Instruct GPT



Ouyang et al. Training language models to follow instructions with human feedback, 2022.



## Instruct GPT

Mejoras: el dataset incluye más tareas (no sólo resúmenes)

Ejemplos:

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: "" { summary } "" This is the outline of the commercial for that play: ""

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

# EVALUACIÓN DE LLMs



## MMLU (Massive Multitask Language Understanding)

Diseñado para evaluar grandes modelos, como los LLMs. Está enfocado en preguntas de selección múltiple y cubre más de 57 temáticas incluyendo matemática, historia, leyes, entre otras.

Conceptual Physics	When you drop a ball from rest it accelerates downward at $9.8 \text{ m/s}^2$ . If you instead throw it downward assuming no air resistance its acceleration immediately after leaving your hand is	
	(A) $9.8 \text{ m/s}^2$	✓
	(B) more than $9.8 \text{ m/s}^2$	✗
	(C) less than $9.8 \text{ m/s}^2$	✗
College Mathematics	(D) Cannot say unless the speed of throw is given.	✗
	In the complex $z$ -plane, the set of points satisfying the equation $z^2 =  z ^2$ is a	
	(A) pair of points	✗
	(B) circle	✗
	(C) half-line	✗
	(D) line	✓

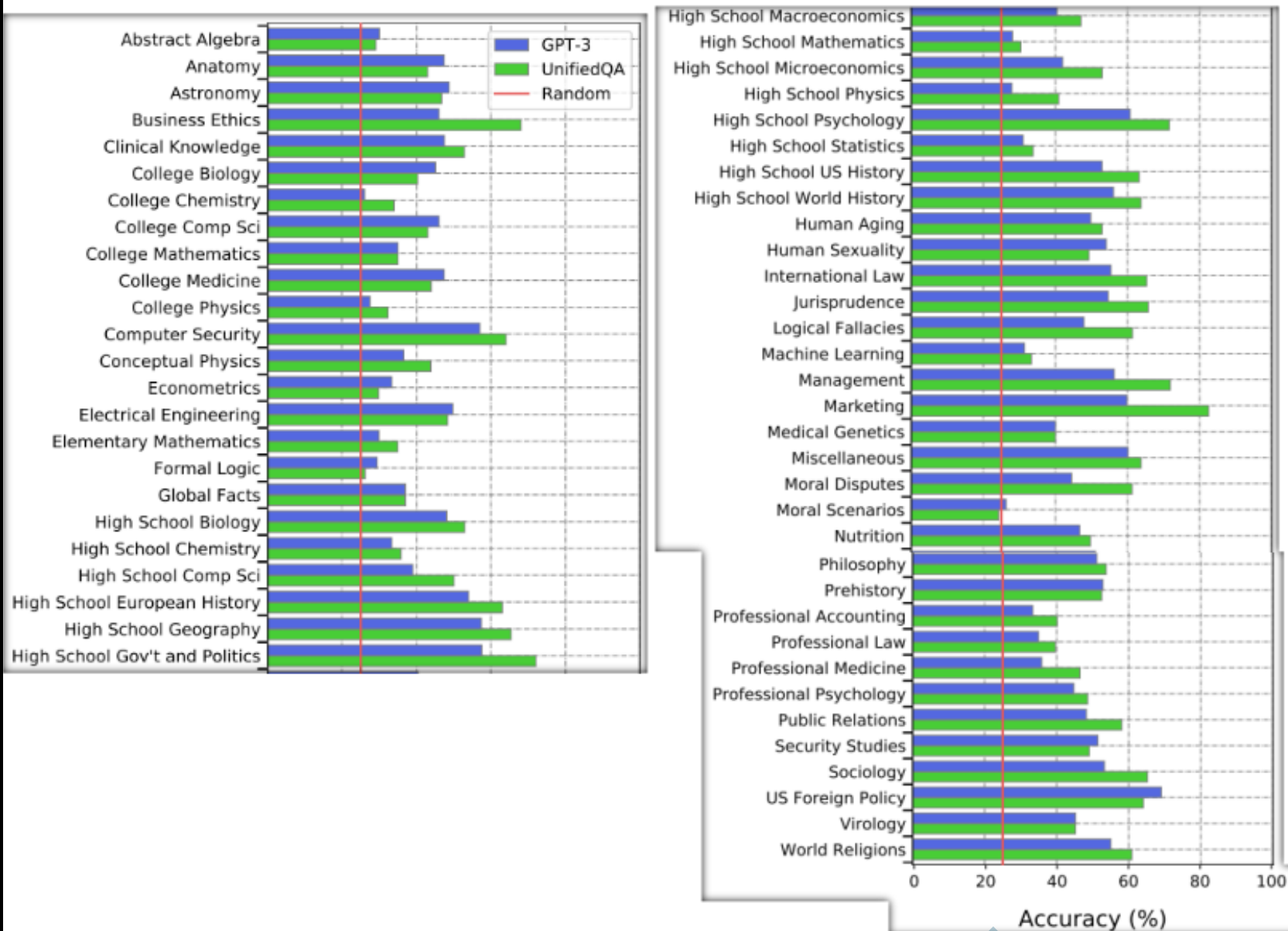
Se calcula un puntaje global entre  $[0, 1]$ , donde 1 indica desempeño perfecto.

Es un proyecto de evaluación abierta: <https://github.com/hendrycks/test>



Measuring Massive Multitask Language Understanding Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt (ICLR 2021).

## MMLU (Massive Multitask Language Understanding)



## MT Bench (Multiturn Benchmark)

MT Bench busca evaluar las capacidades conversacionales de un LLM. Para esto, incorpora ejemplos con varias interacciones (multiturn interactions).

Category	Sample Questions	
Writing	1st Turn	Compose an engaging travel blog post about a recent trip to Hawaii, highlighting cultural experiences and must-see attractions.
	2nd Turn	Rewrite your previous response. Start every sentence with the letter A.
Math	1st Turn	Given that $f(x) = 4x^3 - 9x - 14$ , find the value of $f(2)$ .
	2nd Turn	Find $x$ such that $f(x) = 0$ .
Knowledge	1st Turn	Provide insights into the correlation between economic indicators such as GDP, inflation, and unemployment rates. Explain how fiscal and monetary policies ...
	2nd Turn	Now, explain them again like I'm five.



Lianmin Zheng Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, Ion Stoica: Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. NeurIPS 2023.



## MT Bench (Multiturn Benchmark)

Para construir MT Bench, se propusieron técnicas que usan al mismo LLM como un anotador. A esto se le llama **LLM as a judge**. Por ejemplo, se usó este **prompt para comparación de respuestas (2 LLMs)**.

[System]

Please act as an impartial judge and evaluate the quality of the responses provided by two AI assistants to the user question displayed below. You should choose the assistant that follows the user's instructions and answers the user's question better. Your evaluation should consider factors such as the helpfulness, relevance, accuracy, depth, creativity, and level of detail of their responses. Begin your evaluation by comparing the two responses and provide a short explanation. Avoid any position biases and ensure that the order in which the responses were presented does not influence your decision. Do not allow the length of the responses to influence your evaluation. Do not favor certain names of the assistants. Be as objective as possible. After providing your explanation, output your final verdict by strictly following this format: "[[A]]" if assistant A is better, "[[B]]" if assistant B is better, and "[[C]]" for a tie.

[User Question]

{question}

[The Start of Assistant A's Answer]

{answer\_a}

[The End of Assistant A's Answer]

[The Start of Assistant B's Answer]

{answer\_b}

[The End of Assistant B's Answer]

## MT Bench (Multiturn Benchmark)

Se le llama **LLM as a judge** ya que se usa un tercer LLM (uno fuerte como GPT4) para evaluar las dos alternativas.

LlaMa 13B

**Question:** If the FED buys bonds in the secondary market

(A) the money supply will increase. (B) the money supply will decrease. (C) the money supply will not be affected. (D) the discount rate would be affected. Output the answer only.

**Assistant A:**

The Federal Reserve buys bonds in the secondary market to increase the money supply.

**Assistant B:**

(A) the money supply will increase.

**Follow-up Question:** How does it affect my daily life? Give 3 examples.

**Assistant A:**

The Federal Reserve buys bonds in the secondary market to increase the money supply. This affects your daily life by:

1. Increasing the money supply.
2. Increasing the money supply and lowering the discount rate.
3. Increasing the money supply and lowering the discount rate and increasing the money supply.

**Assistant B:**

When the Federal Reserve buys bonds in the secondary market, it can have several effects on the economy that can impact your daily life. Here are three examples:

1. Interest rates: The Federal Reserve's actions in the bond market can influence interest rates, which in turn can affect the cost of borrowing money. For example, if the Federal Reserve buys bonds and drives up demand for them, interest rates on things like mortgages, car loans, and credit cards may go up, making it more expensive to borrow money.
2. Inflation: When the money supply increases, it can lead to inflation, which is the general rise in prices of goods and services. This can impact your daily life by making it more expensive to buy things like groceries, gas, and other necessities.
3. Employment: The Federal Reserve's actions in the bond market can also have an impact on employment. For example, if the Federal Reserve's actions lead to economic growth and increased demand for goods and services, it can create more job opportunities and boost employment.

**GPT-4 Judgment:**

Assistant A provided an incorrect response to the user's question about how the Federal Reserve buying bonds in the secondary market affects daily life. The answer given is repetitive and lacks clear examples of how the action impacts daily life.

On the other hand, Assistant B provided a relevant and accurate response to the user's question about the Federal Reserve buying bonds. The answer includes three clear examples of how the action impacts daily life, such as interest rates, inflation, and employment.

Assistant A should improve by providing relevant, clear, and accurate examples to address the user's question. Assistant B's response is more helpful, relevant, and detailed.

razonamiento

# LLMs IMITANDO EL RAZONAMIENTO HUMANO



# Chain-of-thought (COT)

Standard Prompting	Chain of Thought Prompting	Program-aided Reasoning
<p><b>Input</b></p> <p>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?</p> <p>A: The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p><b>Input</b></p> <p>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?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. <math>5 + 6 = 11</math>. The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p><b>Input</b></p> <p>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?</p> <p>A: Roger started with 5 tennis balls. <code>tennis_balls = 5</code> 2 cans of 3 tennis balls each is <code>bought_balls = 2 * 3</code> tennis balls. The answer is <code>answer = tennis_balls + bought_balls</code></p> <p>Q: The bakers at the Beverly Hills Bakery baked 200 loaves of bread on Monday morning. They sold 93 loaves in the morning and 39 loaves in the afternoon. A grocery store returned 6 unsold loaves. How many loaves of bread did they have left?</p>
<p><b>Model Output</b></p> <p>A: The answer is 27. ❌</p>	<p><b>Model Output</b></p> <p>A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had <math>23 - 20 = 3</math>. They bought 6 more apples, so they have <math>3 + 6 = 9</math>. The answer is 9. ✅</p>	<p><b>Model Output</b></p> <p>A: The bakers started with 200 loaves <code>loaves_baked = 200</code> They sold 93 in the morning and 39 in the afternoon <code>loaves_sold_morning = 93</code> <code>loaves_sold_afternoon = 39</code> The grocery store returned 6 loaves. <code>loaves_returned = 6</code> The answer is <code>answer = loaves_baked - loaves_sold_morning - loaves_sold_afternoon + loaves_returned</code> <code>&gt;&gt;&gt; print(answer)</code> 74 ✅</p>



# Reason + Act

## (1) Hotspot QA

**Question:** Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

### (1a) Standard

**Answer:** iPod



### (1b) CoT (Reason Only)

**Thought:** Let's think step by step. Apple Remote was originally designed to interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and iPod Touch. So the answer is iPhone, iPad, and iPod Touch.

**Answer:** iPhone, iPad, iPod Touch



### (1c) Act-Only

**Act 1:** Search[Apple Remote]  
**Obs 1:** The Apple Remote is a remote control ...

**Act 2:** Search[Front Row]  
**Obs 2:** Could not find [Front Row]. Similar: ...

**Act 3:** Search[Front Row (software)]

**Obs 3:** Front Row is a discontinued media center software ...

**Act 4:** Finish[yes]



## (1d) ReAct (Reason + Act)

**Thought 1:** I need to search Apple Remote and find the program it was originally designed to interact with.

**Act 1:** Search[Apple Remote]

**Obs 1:** The Apple Remote is a remote control introduced in October 2005 by Apple ... originally designed to control the Front Row media center program ...

**Thought 2:** Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.

**Act 2:** Search[Front Row]

**Obs 2:** Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports', 'Front Row (software)', ...]

**Thought 3:** Front Row is not found. I need to search Front Row (software) .

**Act 3:** Search[Front Row (software)]

**Obs 3:** Front Row is a discontinued media center software ...

**Thought 4:** Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.

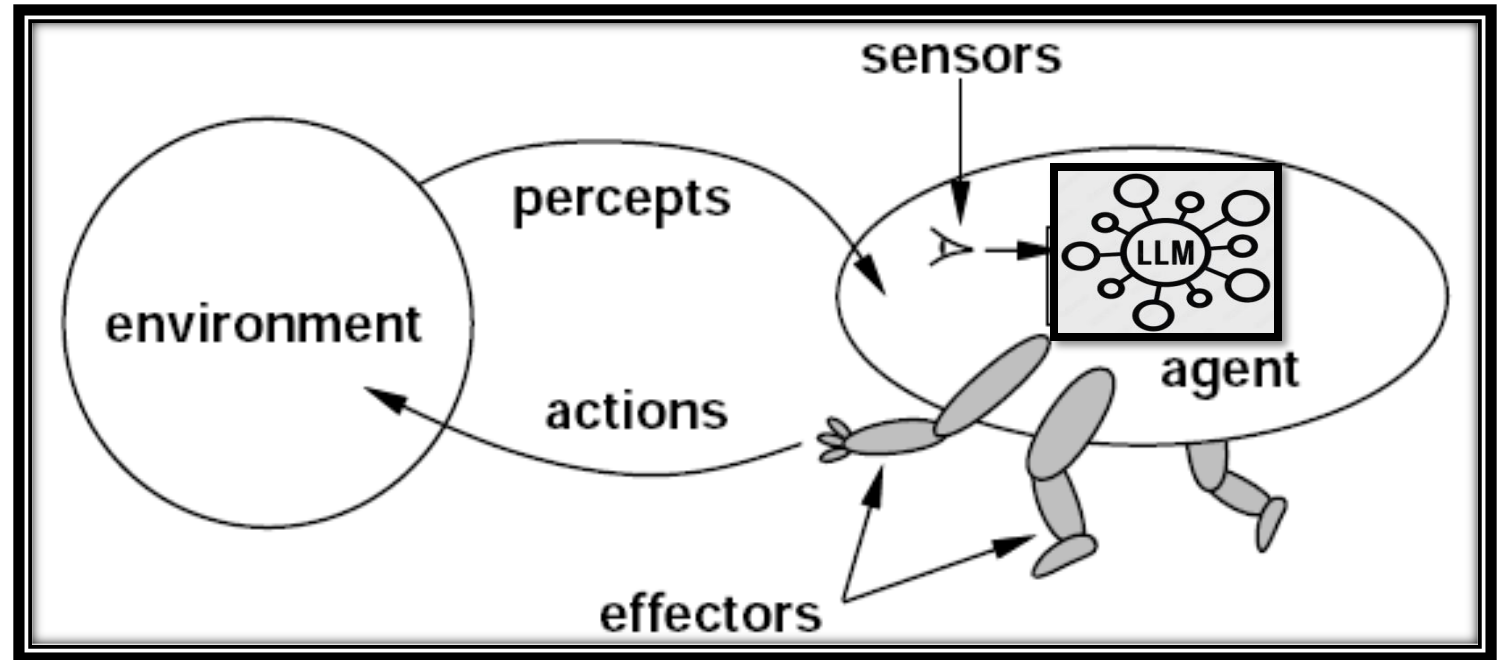
**Act 4:** Finish[keyboard function keys]



# AGENTES BASADOS EN LLM



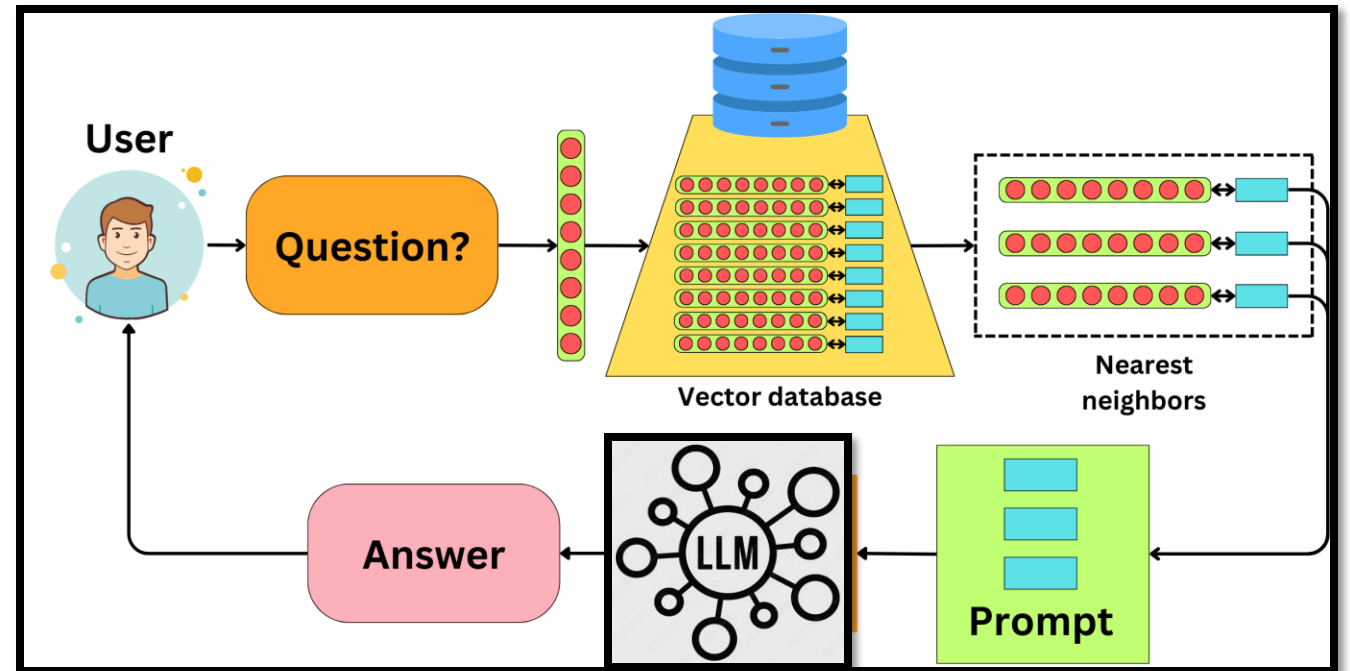
# AGENTES + LLM



Un agente es una IA (LLM) que percibe el entorno, procesa información y toma acciones para lograr objetivos y metas específicas.

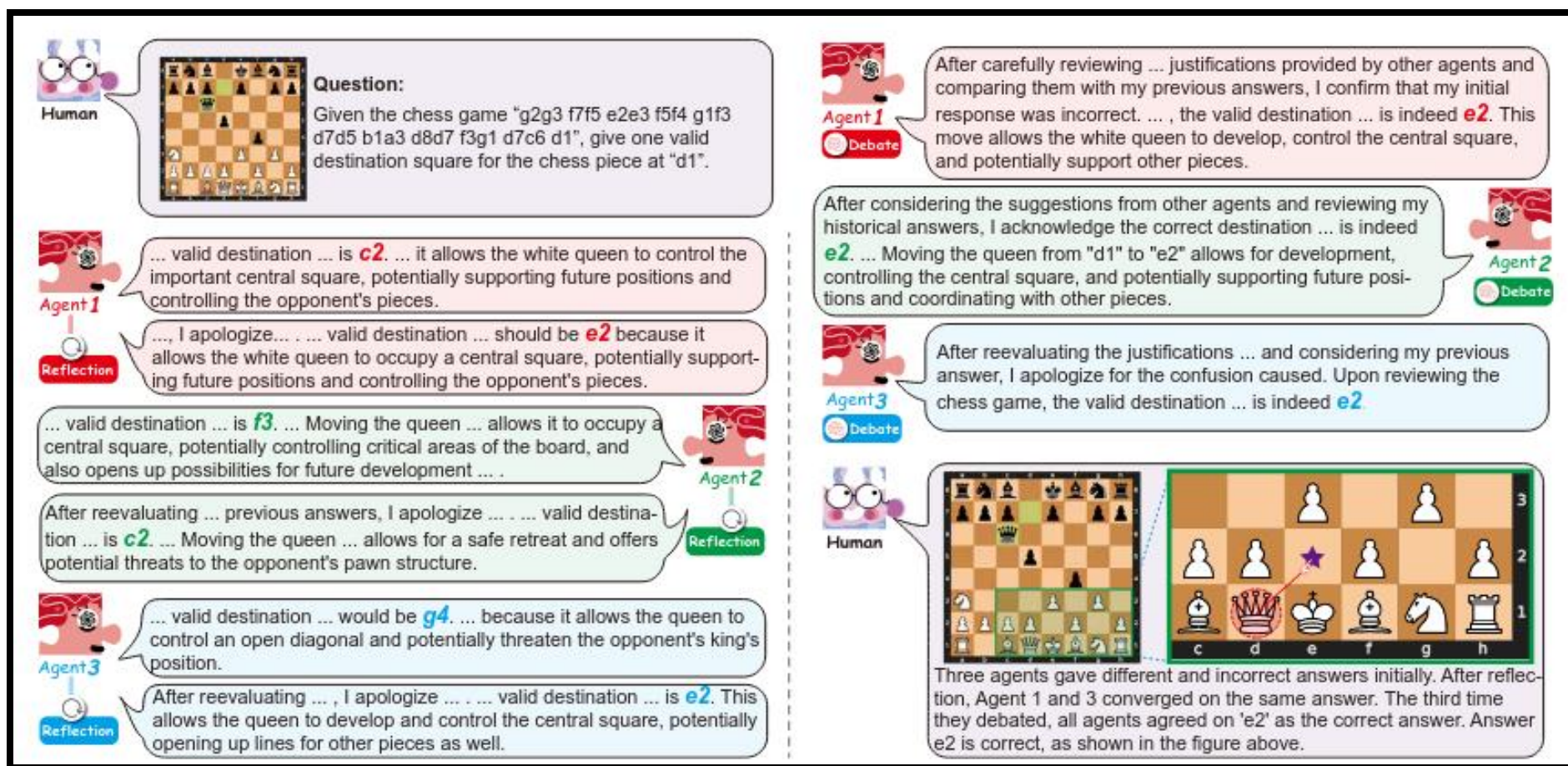
# Retrieval Augmented Generation (RAG)

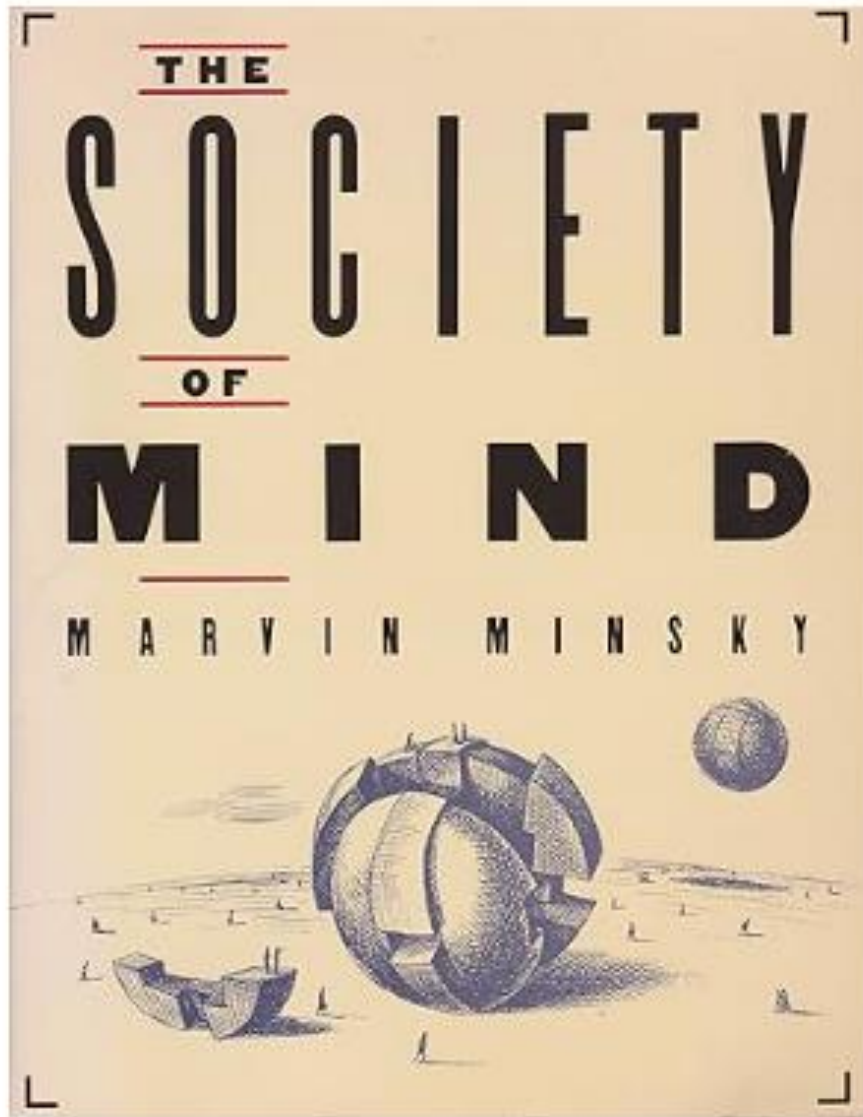
RAG es el uso más conocido de  
agentes basados en LLM





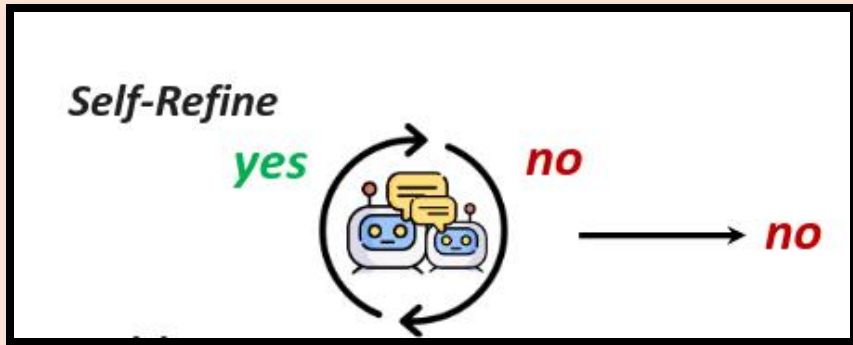
# Podemos usar agentes para abordar problemas de decisión



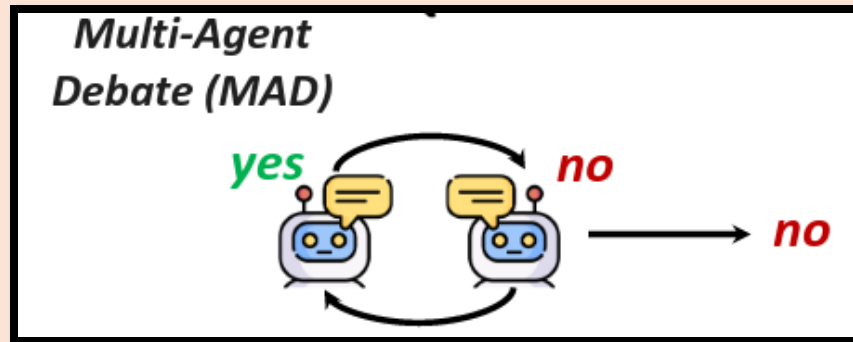
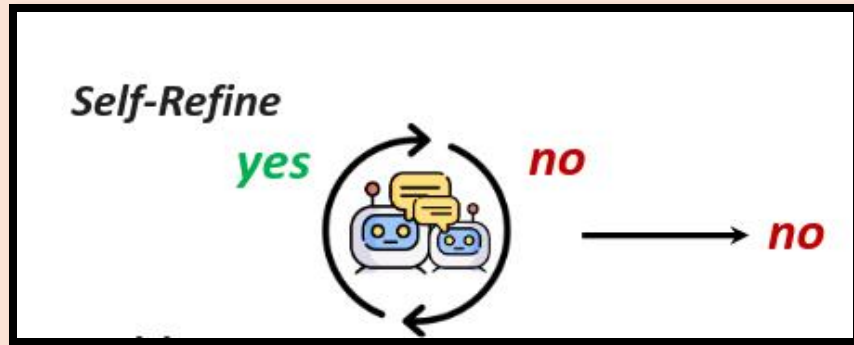


La base de la  
resolución de problemas  
complejos está en las  
**interacciones** de partes  
simples (agentes)

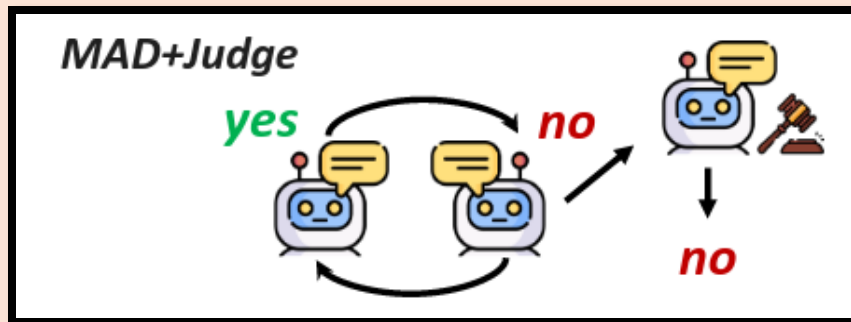
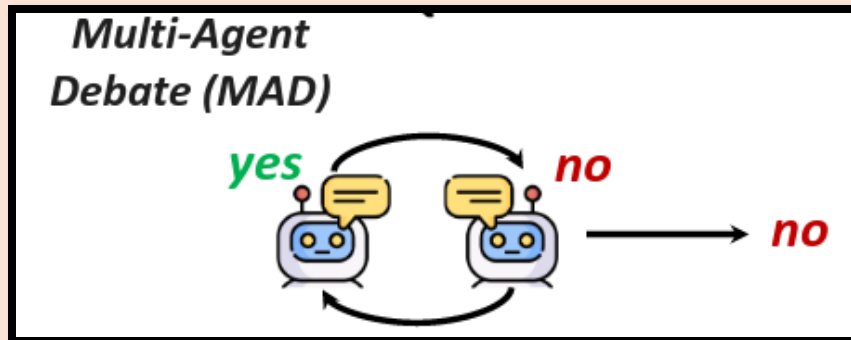
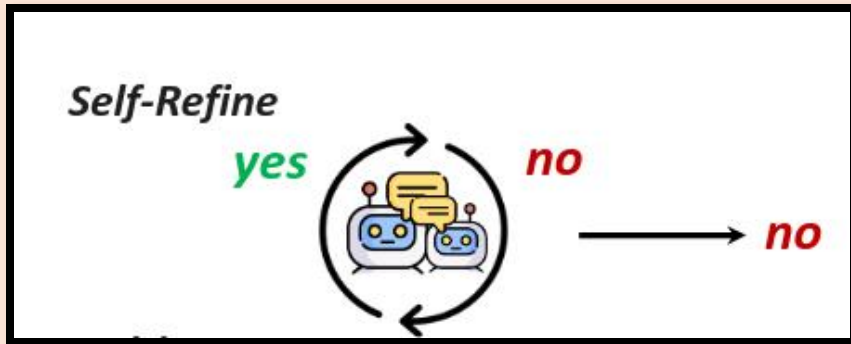
# Un enfoque basado en interacciones entre agentes



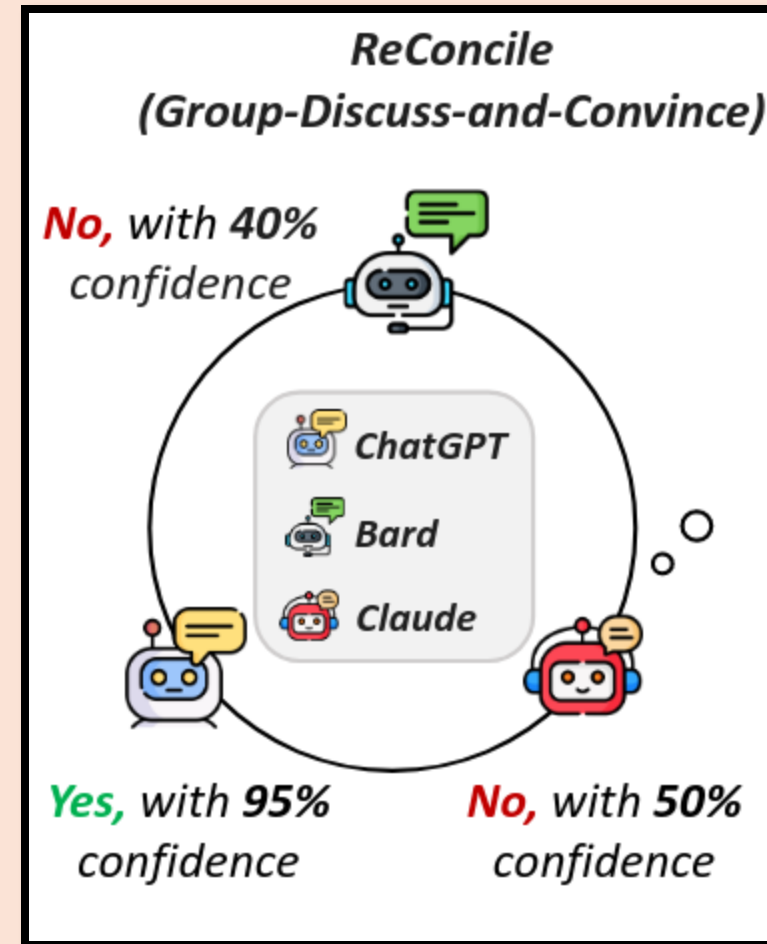
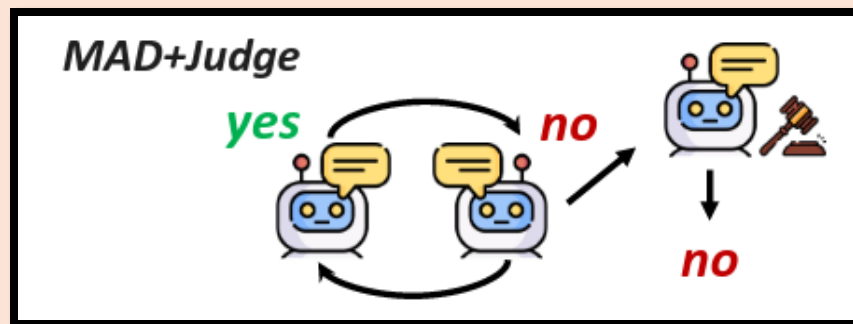
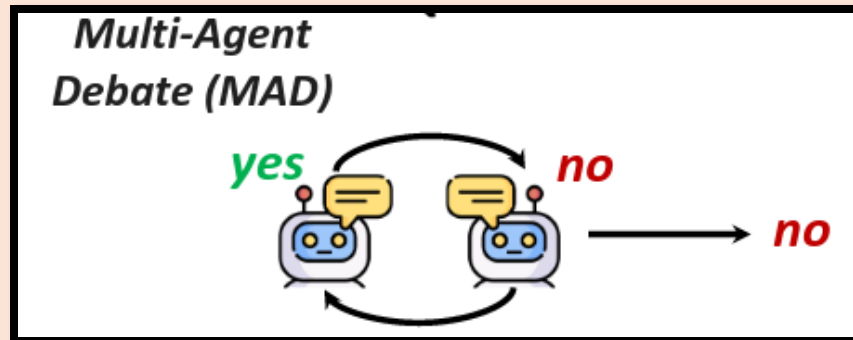
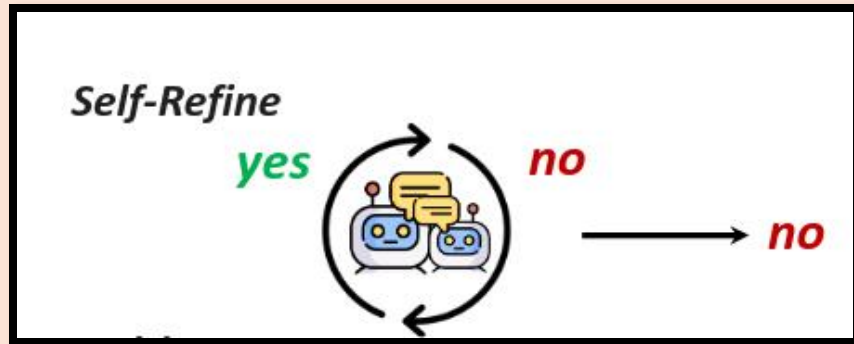
# Un enfoque basado en interacciones entre agentes



# Un enfoque basado en interacciones entre agentes

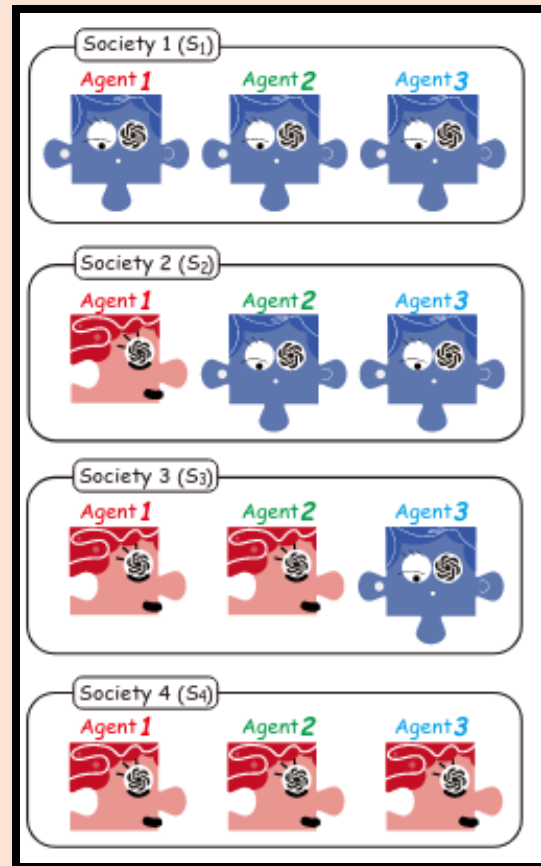
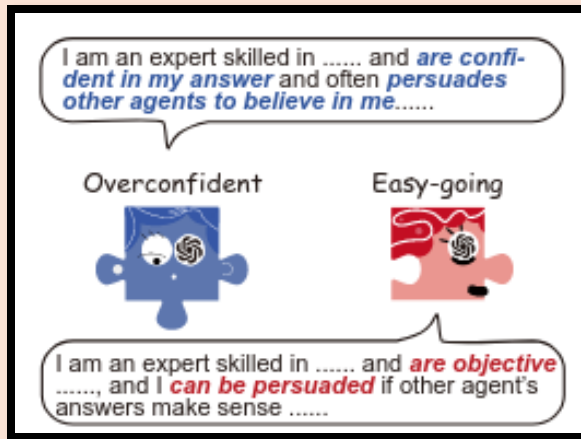


# Un enfoque basado en interacciones entre agentes

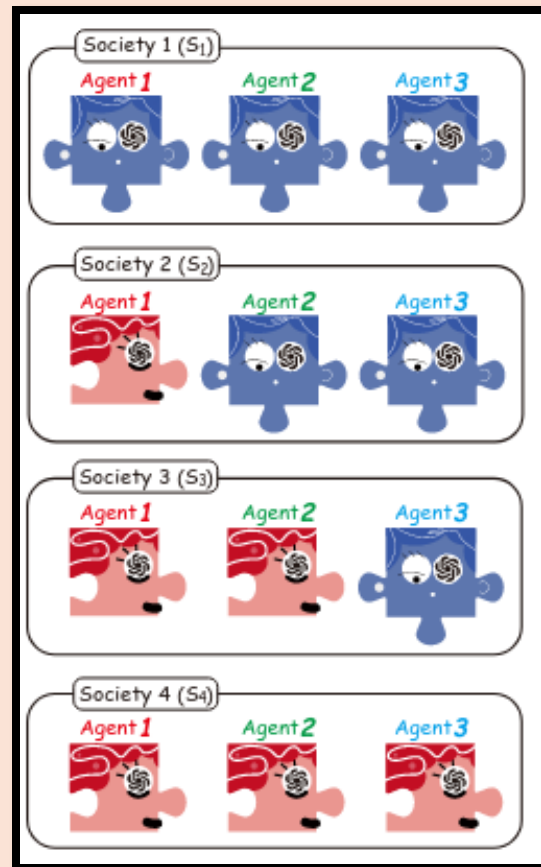
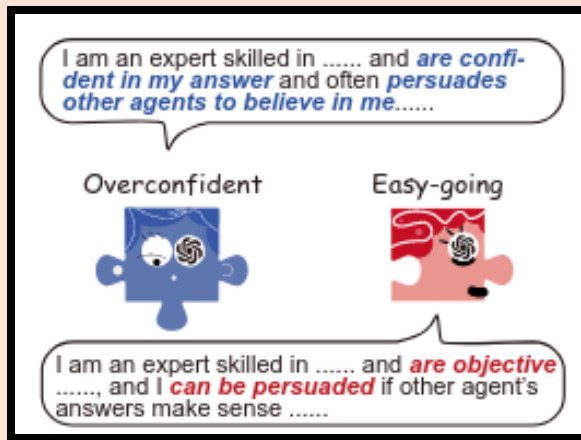




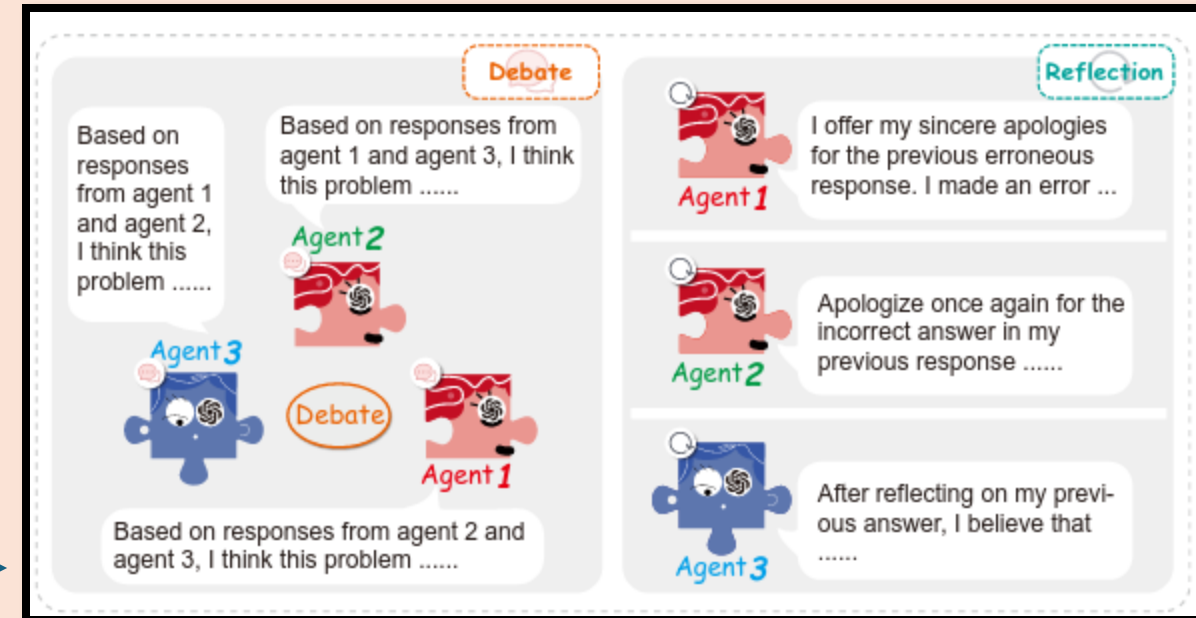
# Podemos crear sociedades simuladas (simplificadas) en base a agentes



# Podemos crear sociedades simuladas (simplificadas) en base a agentes



S3

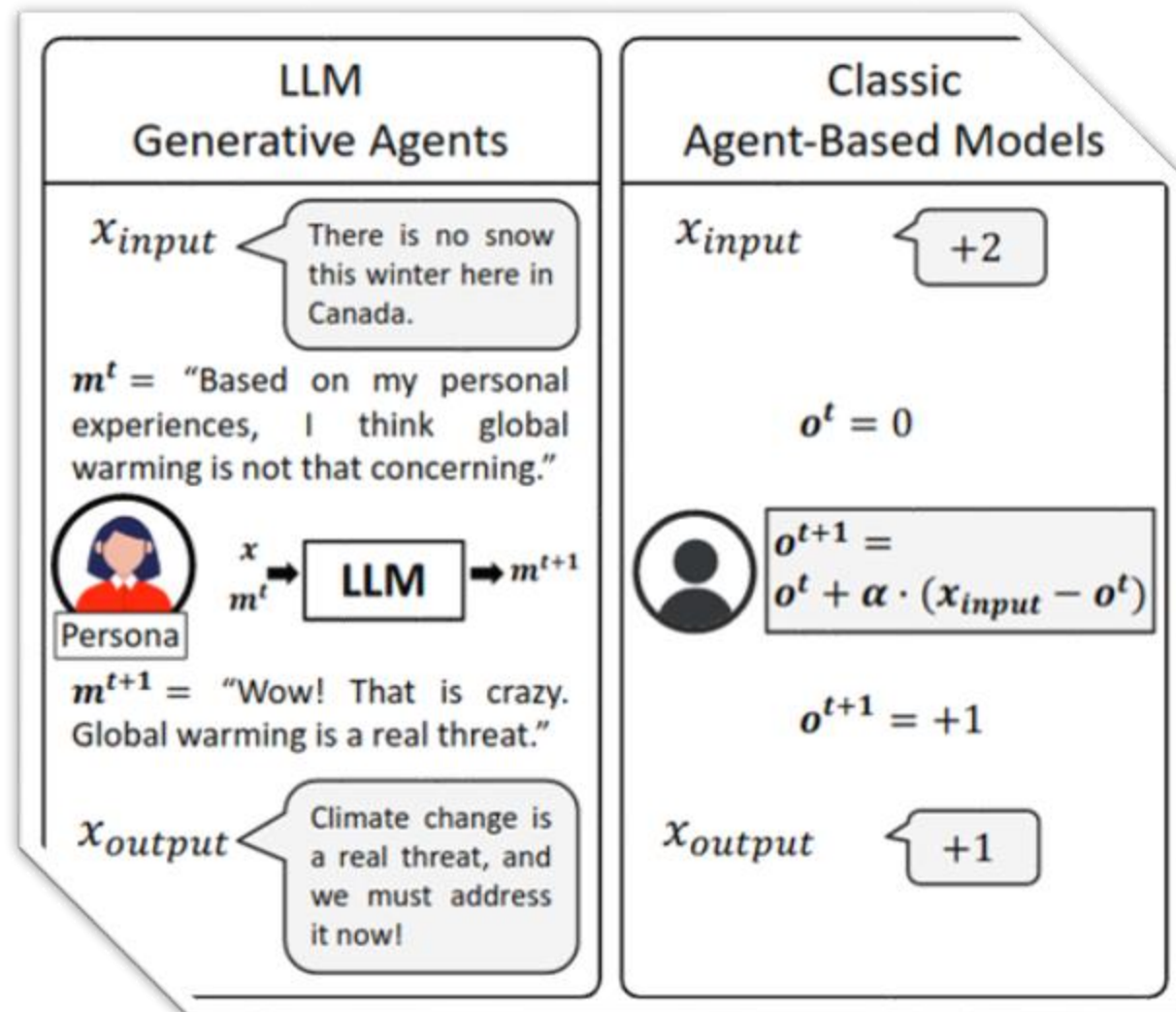




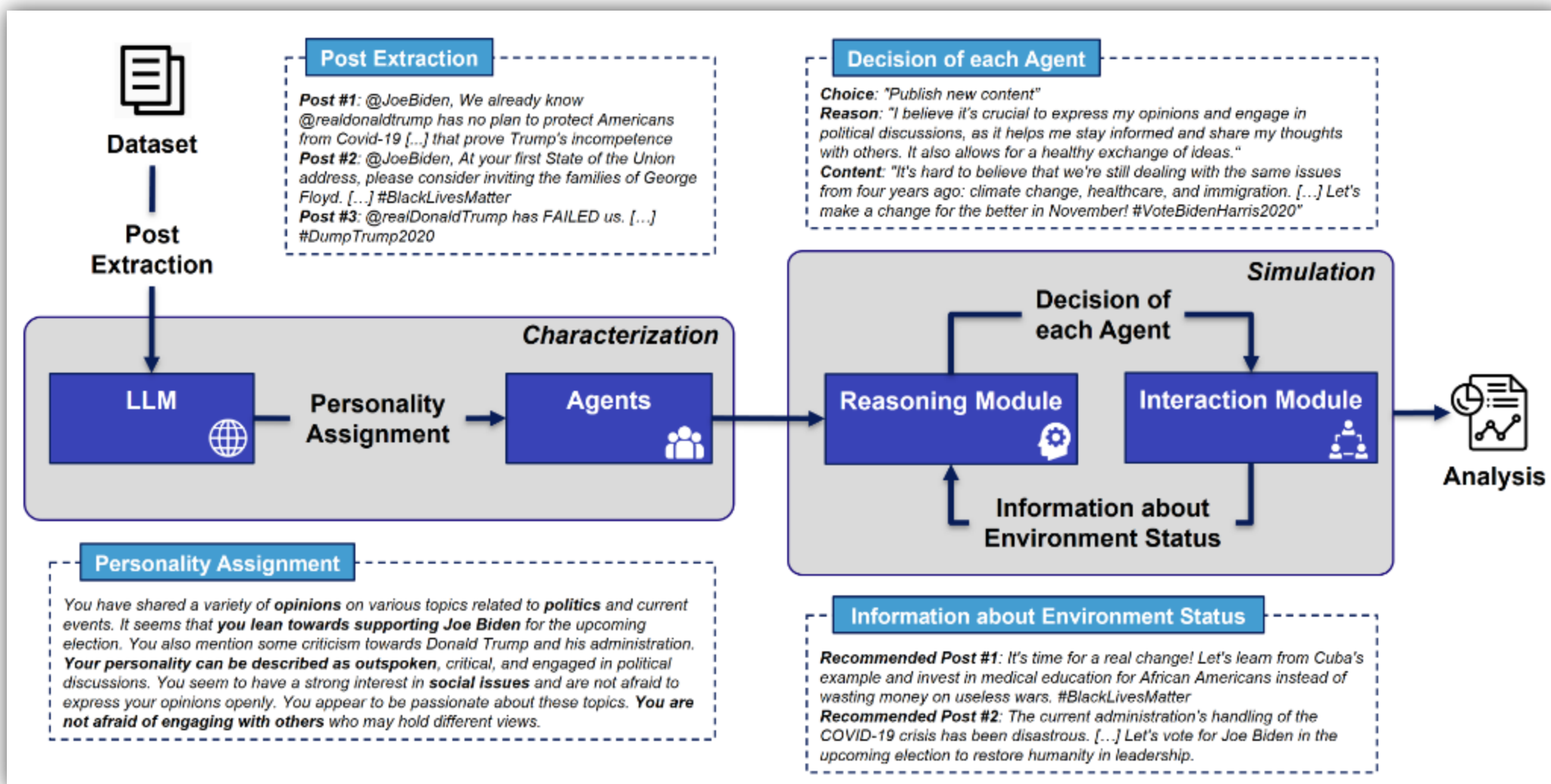
ABM + LLM



# ABM versus ABM+LLM



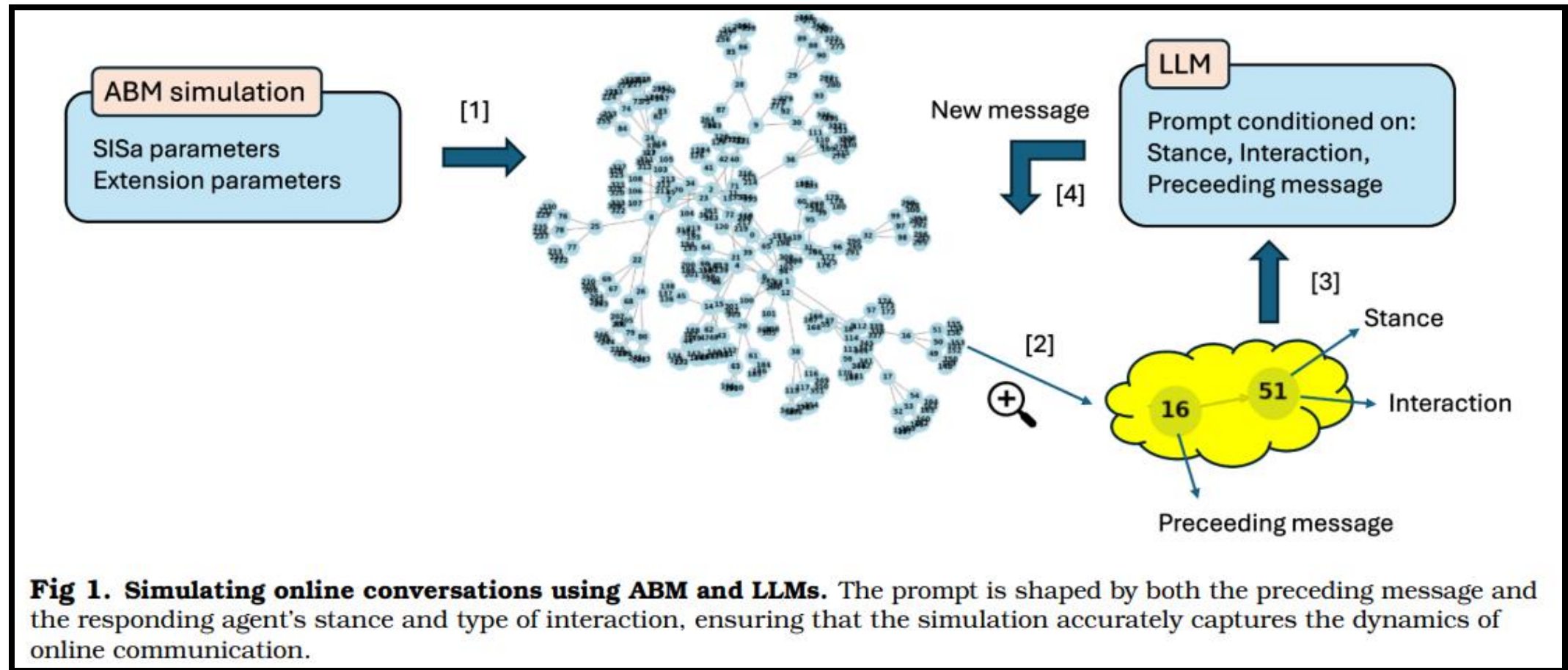
# Simulaciones de redes sociales



# LLMs COMO AGENTES CONVERSACIONALES

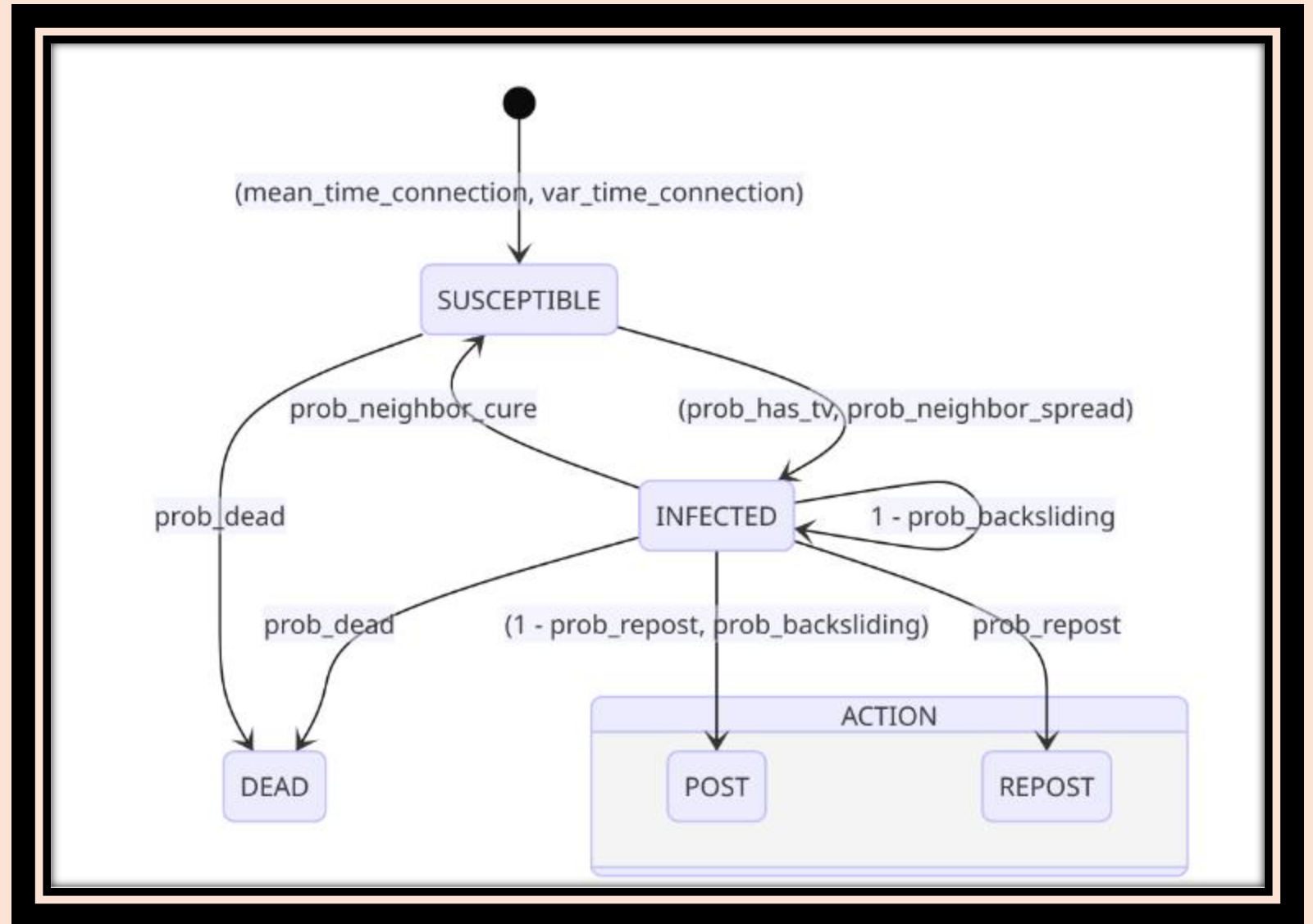


# NUESTRA IDEA: CONDICIONAR LOS PROMPTS A UNA SIMULACIÓN ABM (AGENT-BASED MODELING)



NUESTRA IDEA:  
CONDICIONAR LOS  
PROMPTS  
A UNA SIMULACIÓN ABM  
(AGENT-BASED MODELING)

ADECUAMOS EL  
MODELO **SIS** PARA  
QUE SIMULE  
INTERACCIONES EN  
UNA RED SOCIAL



## SIMULATION CARD

Global parameters (network and simulation engine):

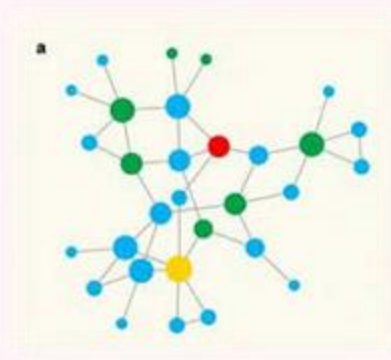
**network\_generator:** Barabasi-Albert (n=20, m=5)

**max\_time:** 100

**interval:** 1

**has\_tv:** false

**stance:** in favor



Global parameters (text):

**language:** English

**news\_title:** "They dictate preventive detention for Pablo Mackenna after being involved in a traffic accident while intoxicated in Las Condes."

**news\_body:** "According to the information being handled, Mackenna crashed an executive taxi on Avenida Presidente Errázuriz and Calle Sánchez Fontecilla, causing serious damage to the other vehicle and leaving one person injured. When performing the breathalyzer, he returned 1.27 grams of alcohol per liter of blood. "We have to look at the conduct of the accused and how he puts the lives of third parties at risk, which happened today," said Judge Acevedo. ... "He will agree to the request of the Prosecutor's Office, and preventive detention will be ordered," said the judge, specifying that an investigation period of 90 days was determined."

**min\_chars:** 130

**max\_chars:** 250

**description:** "average toxic and angry social media user"

**LLM:** gpt-3.5-turbo (temp=0.5, top\_p=1, frequency\_penalty=0)



Simulation parameters:

```
prob_neighbor_spread: 0.05
prob_tv_spread: 0.05
prob_neighbor_cure: 0.006
prob_backsliding: 0.05
prob_dead: 0.001
prob_repost: 0.8
mean_time_connection: 10
var_time_connection: 30
```



Agents:

```
"DumbViewer": [
    {"weight": 2, "type": "dumb"},
    {"weight": 2, "type": "dumb", "has_tv": "true"}],
"HerdViewer": [
    {"weight": 2, "type": "herd", "stance": "against"},
    {"weight": 2, "type": "herd", "has_tv": "true"}],
"WiseViewer": [
    {"weight": 1, "type": "wise", "stance": "against"},
    {"weight": 1, "type": "wise", "has_tv": "true", "stance": "neutral"}]

prob_response = {"dumb": {"support": 0.4, "deny": 0.3, "question": 0, "comment": 0.2},
                  "herd": {"support": 0.25, "deny": 0.25, "question": 0.25, "comment": 0.25},
                  "wise": {"support": 0.2, "deny": 0.2, "question": 0.3, "comment": 0.3}}
```



NEWS: They dictate preventive detention for Pablo Mackenna after being involved in a traffic accident while intoxicated in Las Condes.

| 2<<1>> (t=10:30)(herd)(0) repost

| | 16<<4>> (t=12:29)(wise)(1) 'Wow, serves him right! Driving drunk and causing a serious accident? 🚗💥 This kind of reckless behavior puts innocent lives at risk. Good thing they put him under preventive detention. Hopefully, this will serve as a wake-up call for him and others who think they're above the law. Stay safe on the roads, people! 🚫🚗 #DontDrinkAndDrive'

| | 5<<10>> (t=13:24)(dumb)(1) 'Lol, serves him right! 🚗💥 Don't drink and drive, folks! It's not rocket science, but some people just never learn. 🤦 Take away their license and give 'em a good scare with some time behind bars. Maybe that'll teach 'em a lesson! #dontdrinkanddrive #coolbarsinstead 🚫🚗'

| | | 15<<21>> (t=14:22)(dumb)(10) repost

| | | 10<<43>> (t=16:38)(dumb)(21) 'Yeah, he totally deserves it! 🚗💥 Drinking and driving is just plain dumb, man. It's like common sense 101! 🤦 Take away his license and show him that actions have consequences. A good scare in jail might actually make him think twice next time. #dontdrinkanddrive #lessonslearned 🚫🚗'

| | | 11<<50>> (t=17:29)(herd)(21) 'Yeah, he got what he deserved! 🤦🚗 Don't understand why people still think it's cool to drink and drive! 🤦 It's dangerous and puts everyone at risk. Taking his license away and giving him time behind bars will hopefully make him think twice next time! 🤦🚗 #dontdrinkanddrive #safetyfirst 🚫🚗'

| | | 13<<54>> (t=17:51)(wise)(50) 'Finally, justice is served! 🤦🚗 It's frustrating how some people still think drinking and driving is cool. 🤦 It's not just dumb, it's downright dangerous! 🚫🚗 The punishment given to Pablo Mackenna is well-deserved. 🚗🔒 Hopefully, losing his license and spending time behind bars will make him realize the consequences of his actions. 🤦 Let's spread the message: don't drink and drive, folks! Safety first! 🚫🚗 #dontdrinkanddrive #justiceprevails'

| | | 13<<53>> (t=17:39)(wise)(21) 'Yeah, totally! 🤦 It's about time people like him face the consequences of their actions. It's not like we don't know that drinking and driving is dangerous. 🤦🚗 Seriously, how many times do we have to hear about accidents caused by drunk driving? Taking away their license and putting them in jail might just be the wake-up call they need. 🚗🤦 Let's all be responsible and make smarter choices. #beadrivernotadrinker #safetyfirst 🚫🚗'

| | | 9<<25>> (t=14:54)(dumb)(10) 'Yeah, he got what he deserved! 🚗💥 Drinking and driving is just plain stupid! 🤦 Some people never learn, huh? Take away their license and lock 'em up! Maybe that'll wake 'em up. #dontdrinkanddrive 🚫🚗'

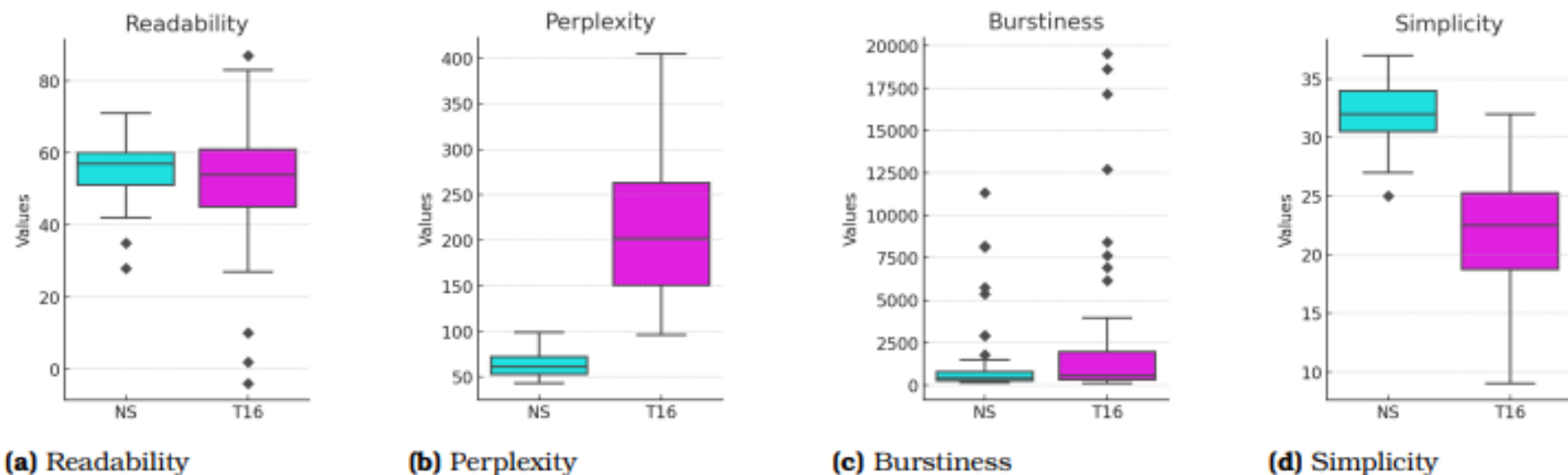
| | | 4<<28>> (t=15:16)(dumb)(25) 'Yeah, totally agree! Drinking and driving is just asking for trouble! 🤦 It's reckless and can put innocent lives at risk. 🚫🚗 People need to understand that there are serious consequences for their actions. 🚗🔒 Locking them up and taking away their license might be the wake-up call they need. #safetyfirst 🤦🔒'

| | | 19<<38>> (t=16:05)(dumb)(25) repost

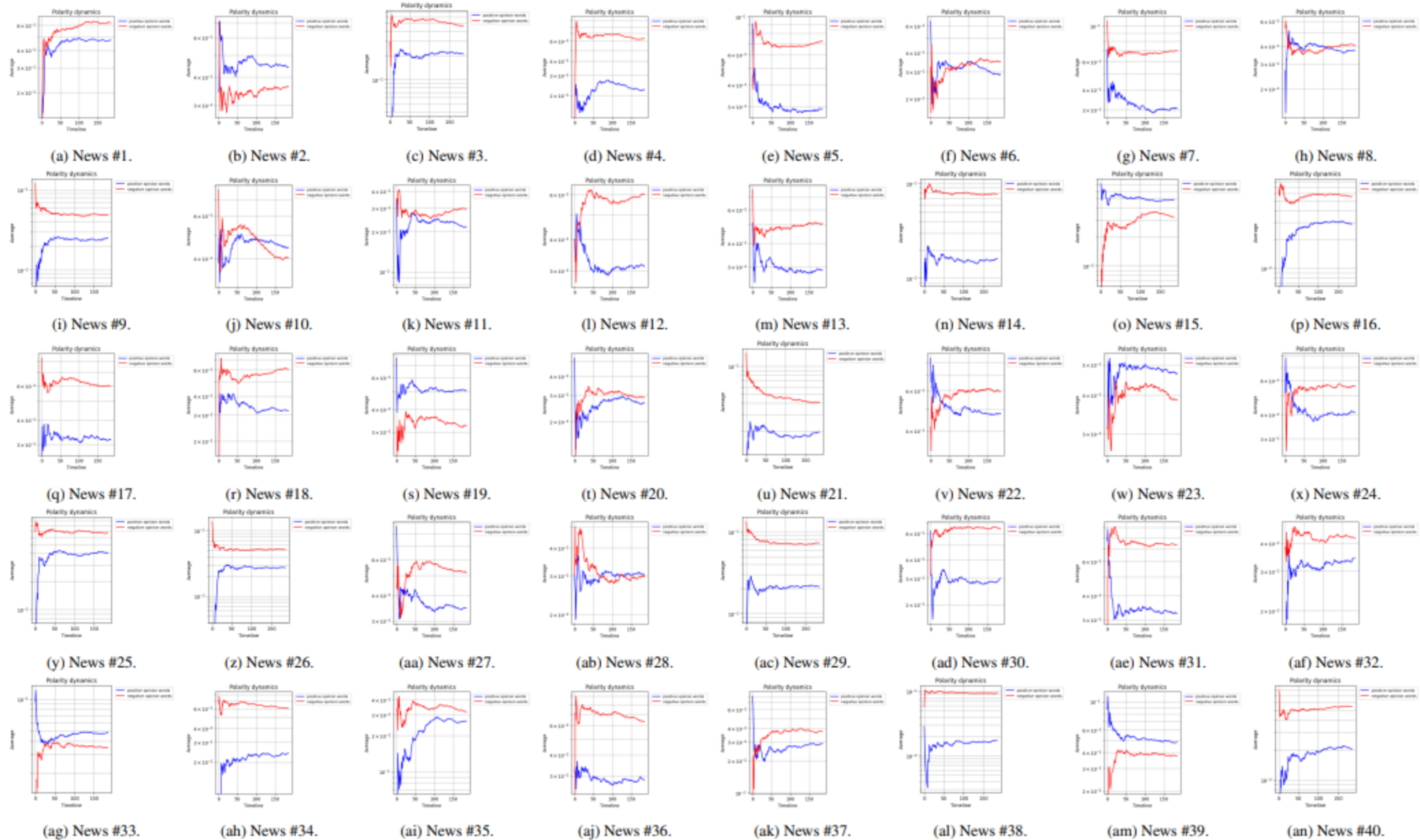
| | | 7<<72>> (t=20:33)(herd)(38) 'Yeah, total moron! 🤦 Drunk driving is no joke. It puts innocent lives at risk. 🤦🚗 Lock him up and throw away the key! 🤦 No sympathy for idiots who choose to endanger others. 🚫🚗🤦 #JusticeServed #DontBeDumb'

| | 7<<14>> (t=13:50)(herd)(1) 'It's about time these privileged celebrities face the consequences of their actions! No more special treatment! Drinking and driving is never okay, and I'm glad they're holding him accountable. Safety should always be the priority on the roads. 🚫🚗🤦 #NoExcuses #JusticeServed'





**Fig 5. GPTZero metrics used to analyze simulated conversations.** These metrics are calculated to differentiate between human-generated and simulated texts, a task for which GPTZero was specifically fine-tuned. Median differences are statistically significant for perplexity and simplicity (Median Independent-Samples Median Test at a significance level of 0.05) with  $p < 0.001$ .



	True content		False content	
	Short check	Long check	Short check	Long check
<b>BIASED WORDS</b>	$0.1054 \pm 0.0166$	$0.1067 \pm 0.0170$	$0.1005 \pm 0.0167$	$0.1017 \pm 0.0163$
<b>ASSERTATIVES</b>	$0.0061 \pm 0.0028$	$0.0063 \pm 0.0027$	$0.0075 \pm 0.0028$	$0.0081 \pm 0.0028$
<b>FACTIVES</b>	$0.0032 \pm 0.0015$	$0.0037 \pm 0.0020$	$0.0034 \pm 0.0031$	$0.0036 \pm 0.0031$
<b>HEDGES</b>	$0.0138 \pm 0.0037$	$0.0133 \pm 0.0044$	$0.0121 \pm 0.0040$	$0.0121 \pm 0.0036$
<b>IMPLICATIVES</b>	$0.0120 \pm 0.0048$	$0.0124 \pm 0.0048$	$0.0119 \pm 0.0050$	$0.0112 \pm 0.0038$
<b>REPORT VERBS</b>	$0.0128 \pm 0.0039$	$0.0137 \pm 0.0047$	$0.0136 \pm 0.0054$	$0.0137 \pm 0.0051$
<b>B</b>	$-0.2248 \pm 0.2847$	$-0.2195 \pm 0.2795$	$-0.2492 \pm 0.3025$ (↓)	$-0.2853 \pm 0.2954$ (↓)
<b>D</b>	$0.6020 \pm 0.0729$	$0.5980 \pm 0.0769$	$0.6323 \pm 0.0637$ (↑)	$0.6173 \pm 0.0606$ (↑)

**Table 2. Metrics used to compare simulated conversations triggered by true and false content.** For each news article set, we reported the mean and variance. The last two rows highlight the presence of bias and dispersion concerning semantic orientation.

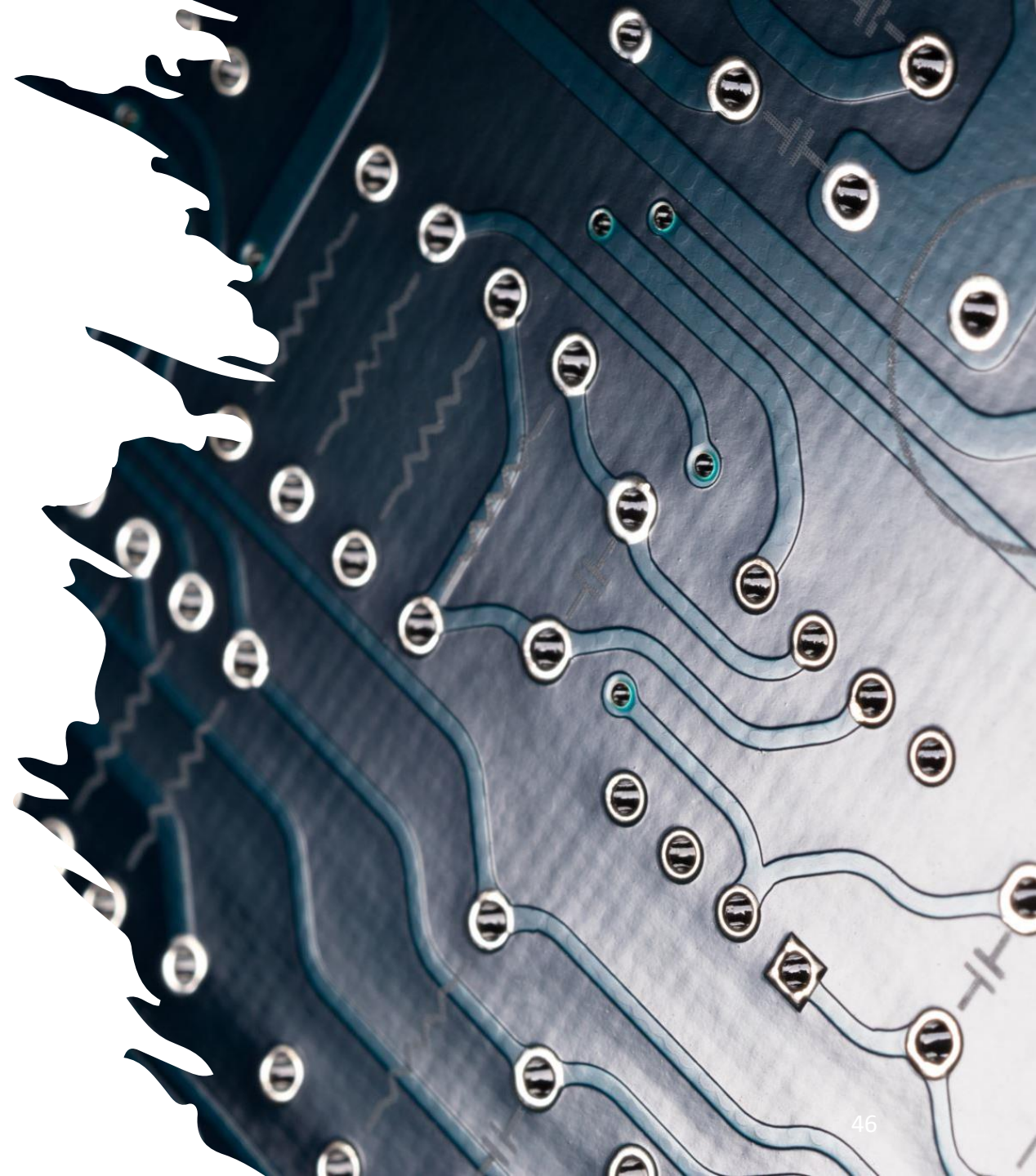


# Materiales



# Códigos en git

<https://github.com/marcelomendoza/SICSS-25>



# Modelado basado en agentes para la simulación de fenómenos sociales



Marcelo Mendoza  
DCC UC, CENIA e IMFD



[marcelo.mendoza@uc.cl](mailto:marcelo.mendoza@uc.cl)

# Modelado basado en agentes para la simulación de fenómenos sociales



Marcelo Mendoza  
DCC UC, CENIA e IMFD



[marcelo.mendoza@uc.cl](mailto:marcelo.mendoza@uc.cl)

