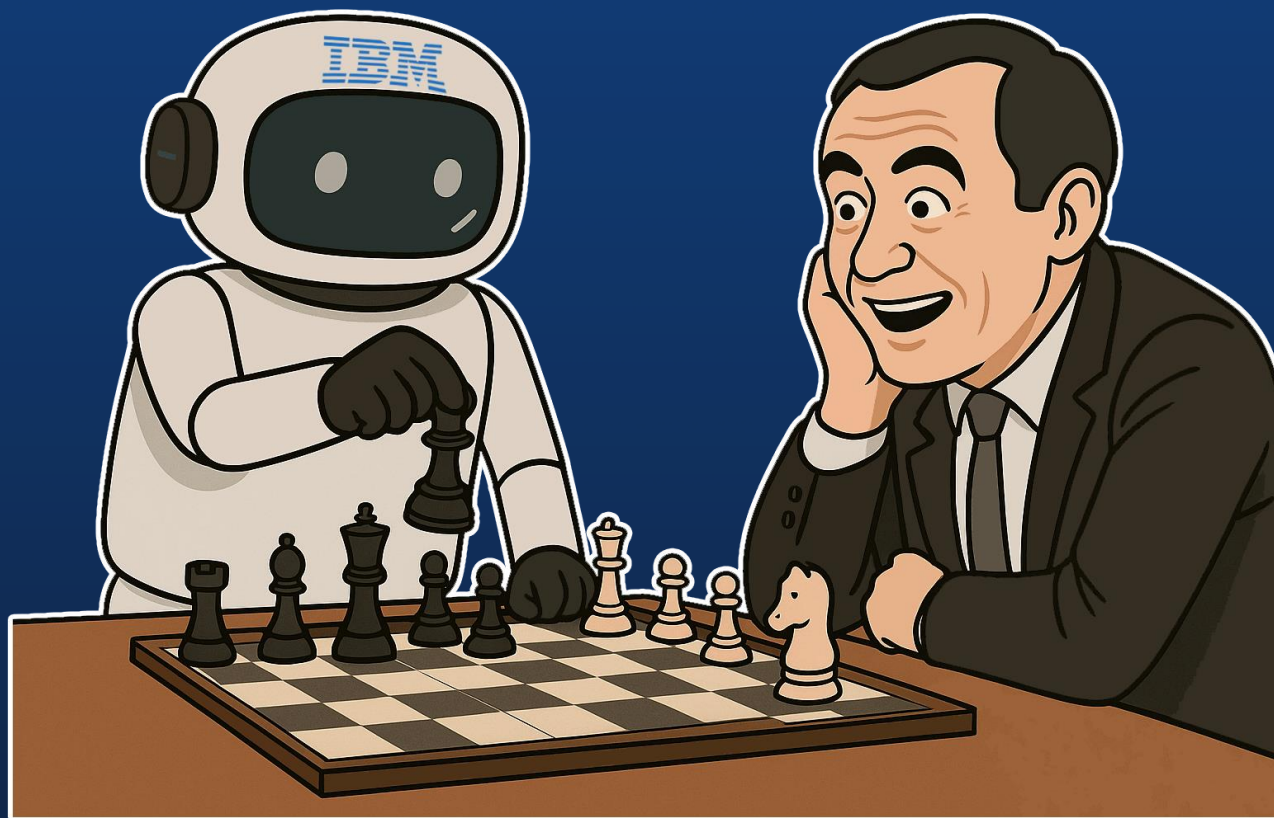


LLMs as RPG GAME MASTER

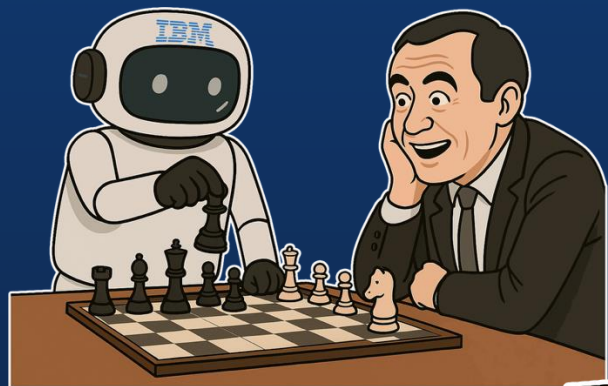
A Comparative Analysis on Performance
and Player Experience

Gaetan Berlaimont - Justin Vanwichelen
Master thesis supervisor : H  l  ne Verhaeghe

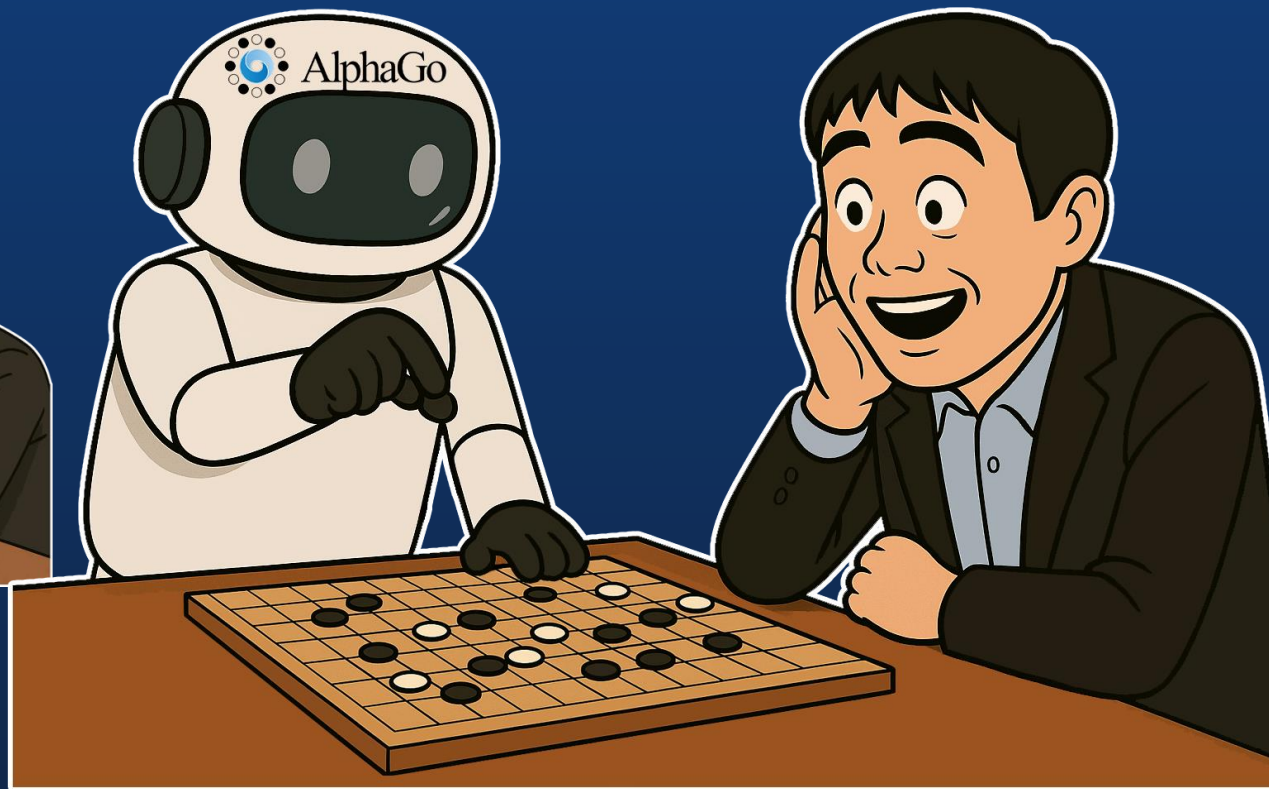


1997



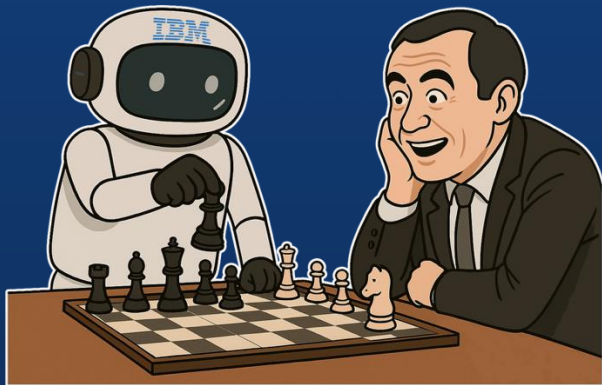


1997

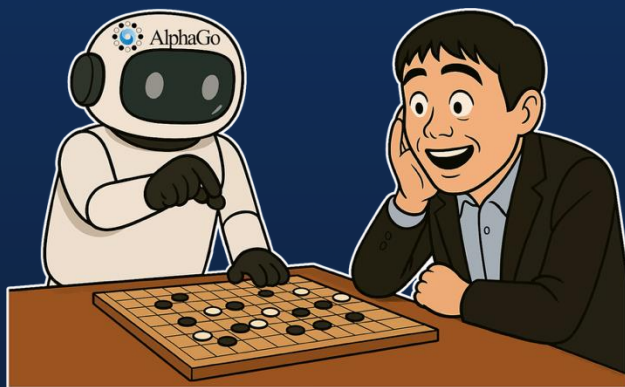


2016

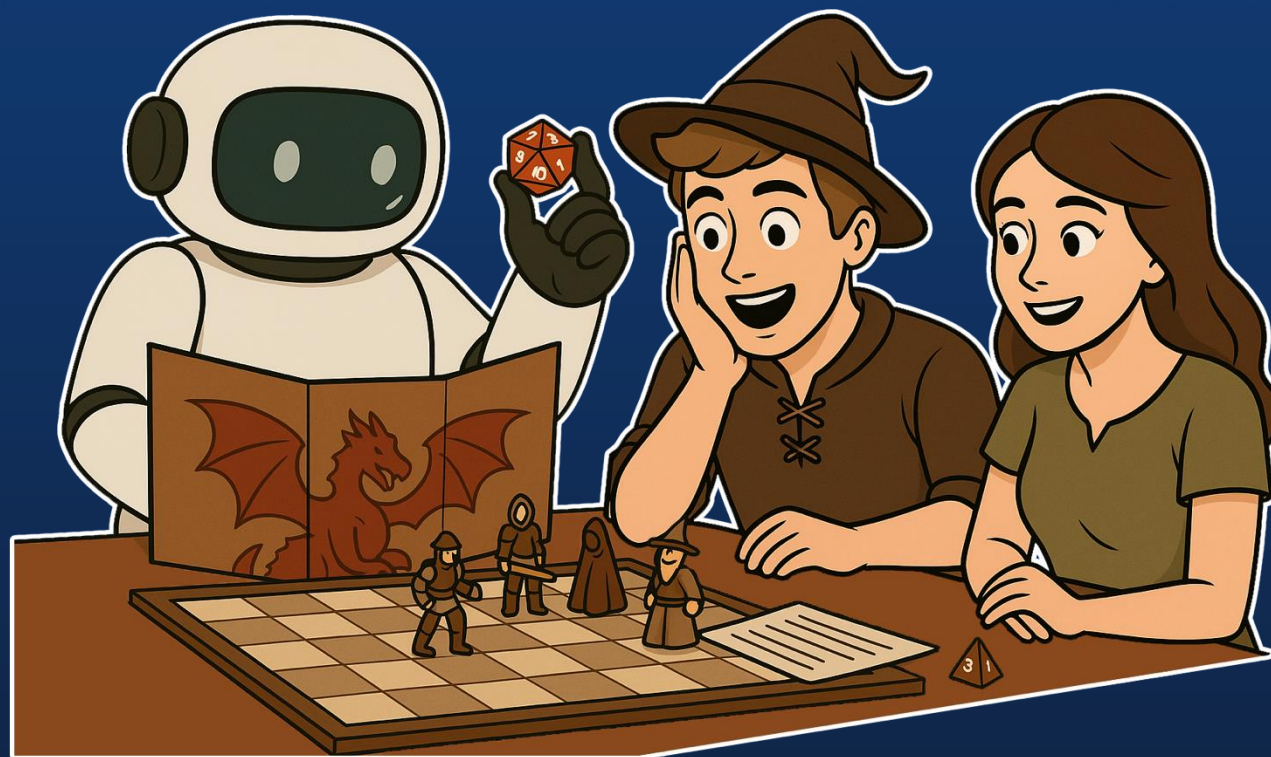




1997



2016



2025?





RPG

with LLM

LANCER

PARTIE
RAPIDE

TESTS
SPÉCIFIQUES

QUESTIONNAIRES

CRÉDITS

Choisis ton LLM

L'IA qui va générer ton histoire
(Tu pourras toujours la changer après)



Llama
(llama_3_3_70b_versatile)

Vitesse 

Intelligence 




Gemma
(gemma2_9b_it)

Vitesse 

Intelligence 




Mistral
(mistral_3_3_70b)


Vitesse 

Intelligence 



ChatGPT
(GPT-4o mini)

Vitesse 

Intelligence 

Sélectionner un thème



Change
AI



Theme selection

Pour quitter le mode plein écran, appuyez sur Échap

Start

Harry



Edit



Hermione



Edit



Ron



Edit



Vous arrivez à la lisière d'une clairière étrangement silencieuse. Le soleil peine à percer les branches touffues des arbres qui l'entourent, laissant tomber des ombres profondes sur le sol. Une sensation de malaise vous envahit, comme si les regards invisibles des arbres étaient fixés sur vous.



Eolande

Santé : 20 Armure : 10
Arme : 6 Précision : 0
Vision : 2 Déplacement : 2



Ragnar

Santé : 20 Armure : 10
Arme : 8 Précision : -2
Vision : 2 Déplacement : 2



Alinda

Santé : 20 Armure : 10
Arme : 4 Précision : 2
Vision : 2 Déplacement : 2



Razorg

Santé : 20 Armure : 10
Arme : 4 Précision : 2
Vision : 2 Déplacement : 2



Que veux-tu faire? (Laisse vide et l'IA choisira pour toi)

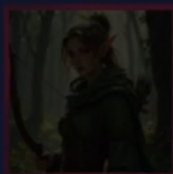


Le silence de la forêt est soudainement brisé par le craquement d'une branche sous vos pieds. Vous êtes arrivés à l'entrée du donjon, une ouverture sombre et sinistre dans la roche. Une brise glaciale souffle du gouffre, emportant avec elle un parfum de poussière et d'humidité.

Au tour de Ragnar

Choisis une cible et entre la description de ton attaque souhaitée.

Eolande

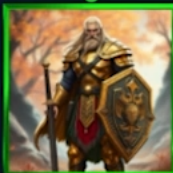


Santé : Eolande est étourdie mais debout.
Armure : 10 +2
Arme : 8
Précision : 0



1

Ragnar

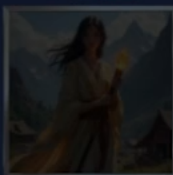


Santé : En forme, possède toute sa vie.
Armure : 10
Arme : 8
Précision : -2



2

Alinda

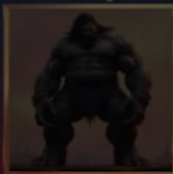


Santé : Intouchée
Armure : 10
Arme : 4
Précision : 2



1

Therym



Santé : En forme, possède toute sa vie.
Armure : 10
Arme : 8
Précision : -2



Le Spectateur



Santé : 18
Armure : 9
Corrompu pendant : 1 tours.





LLMs as
RPG
GAME MASTER

Choice of the LLM models

Criterion 1 : origin

Criterion 2 : number of parameters

Criterion 3 : API acces



Gpt-4o-mini



OpenAI

~3–8B



Mistral
Small
25.01



MISTRAL
AI_

~4–7B



Gemma_2
9B IT

Google

9B



LLaMA 3.3
70B Versatile



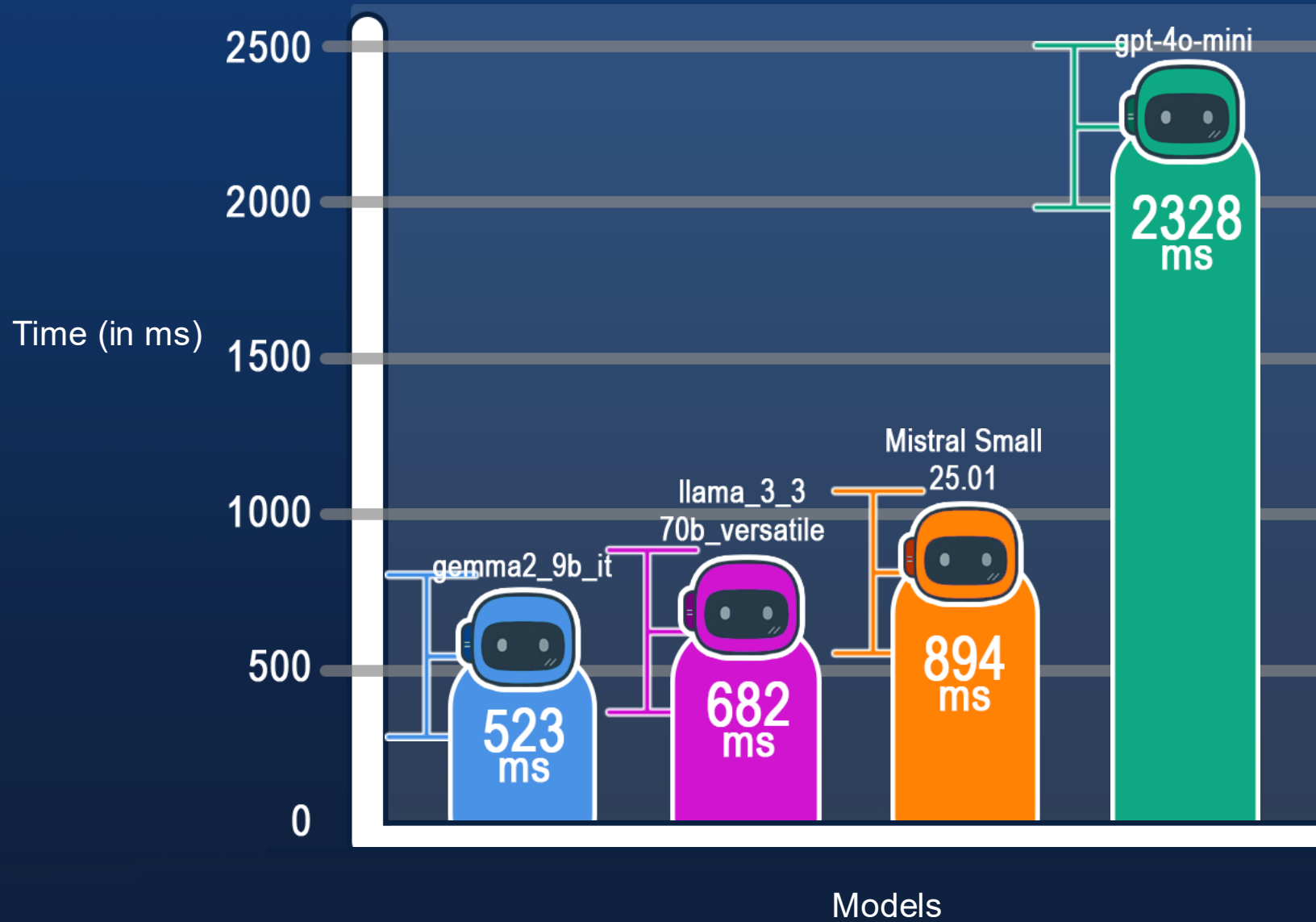
Meta

70B



Mean answer time for request per model

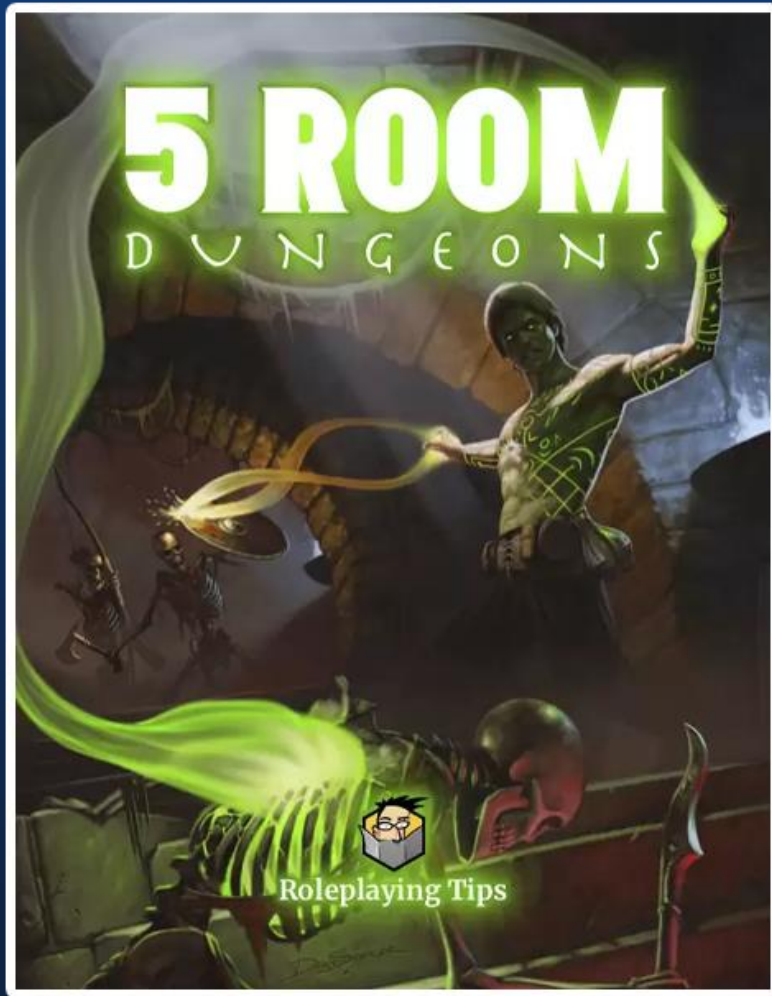
(measured in the application)



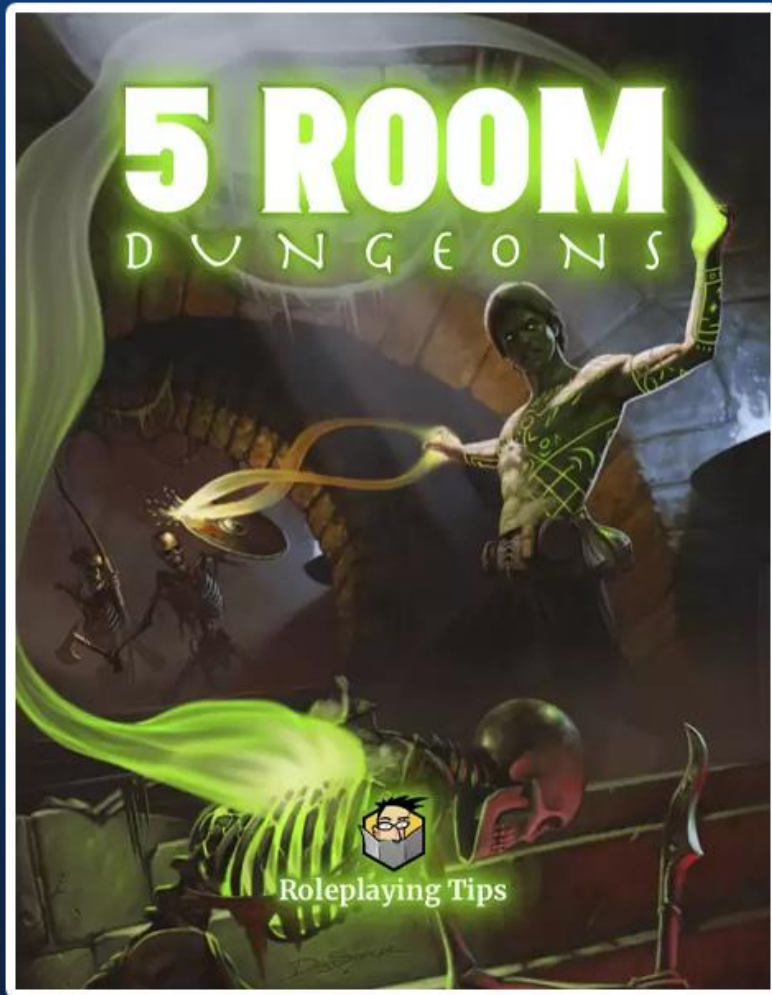
The application



The application



The application



Room 1: Entry

Room 2: puzzle

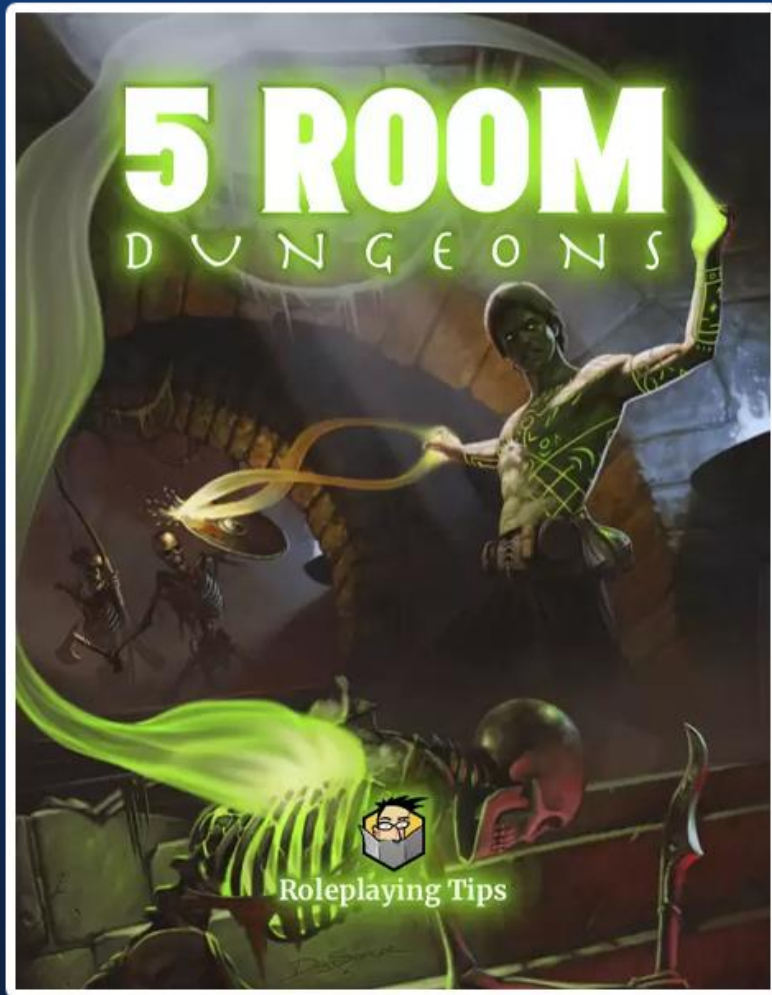
Room 3: exploration

Room 4: boss

Room 5: reward



The application



Room 1: Entry

Room 3: exploration

Room 4: boss



1. Choice of LLM



Meta AI

Llama

(llama_3_3_70b versatile)

Speed _____
Intelligence _____

Pick

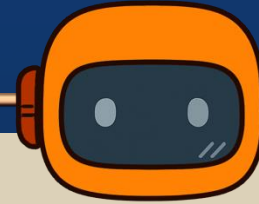


Gemma

(gemma2_9b_it)

Speed _____
Intelligence _____

Pick



Mistral

(Mistral Small 25.01)

Speed _____
Intelligence _____

Pick



ChatGPT

(gpt-4o mini)

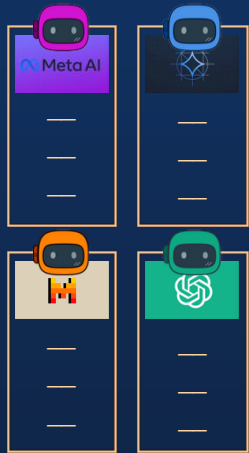
Speed _____
Intelligence _____

Pick



The application

1. Choice of LLM



2. Choice of theme (or create it)

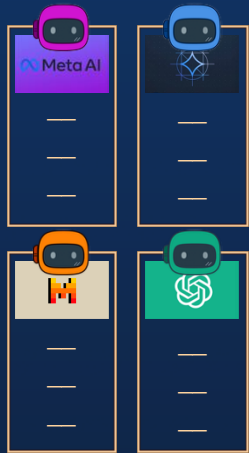


The application

3. Course of the adventure

2 distinct types of game phases

1. Choice of LLM

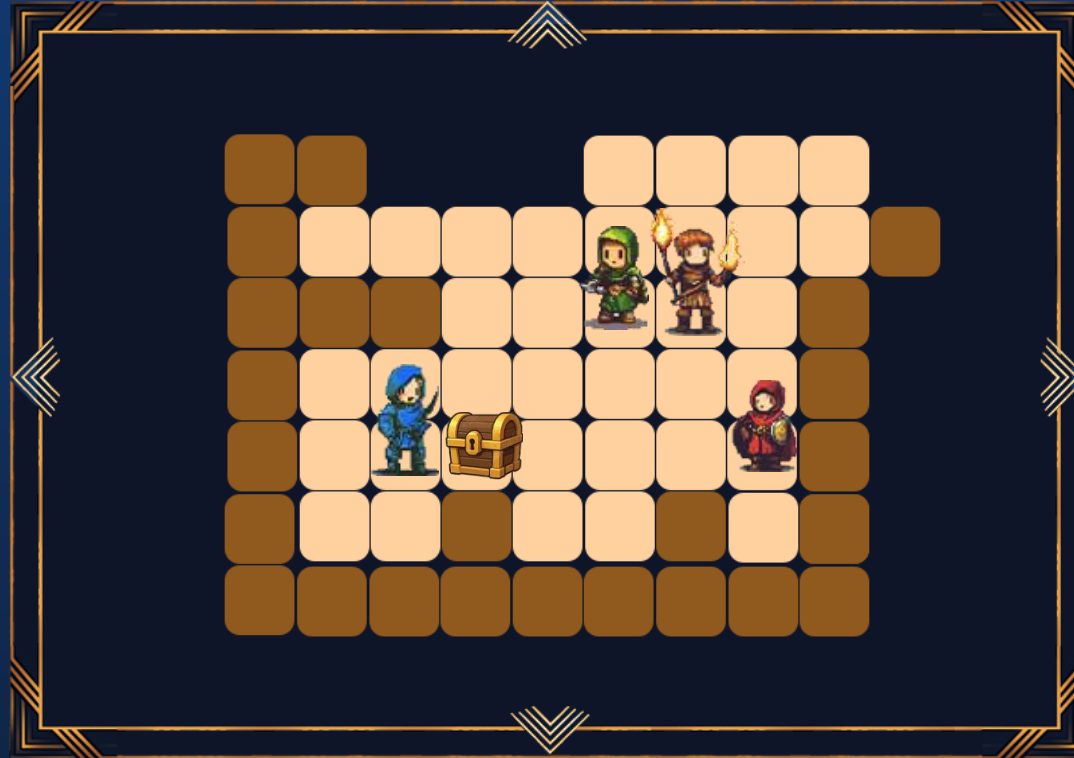


2. Choice of theme



The application

2 types de phases de jeu distincts A. Phases d'exploration

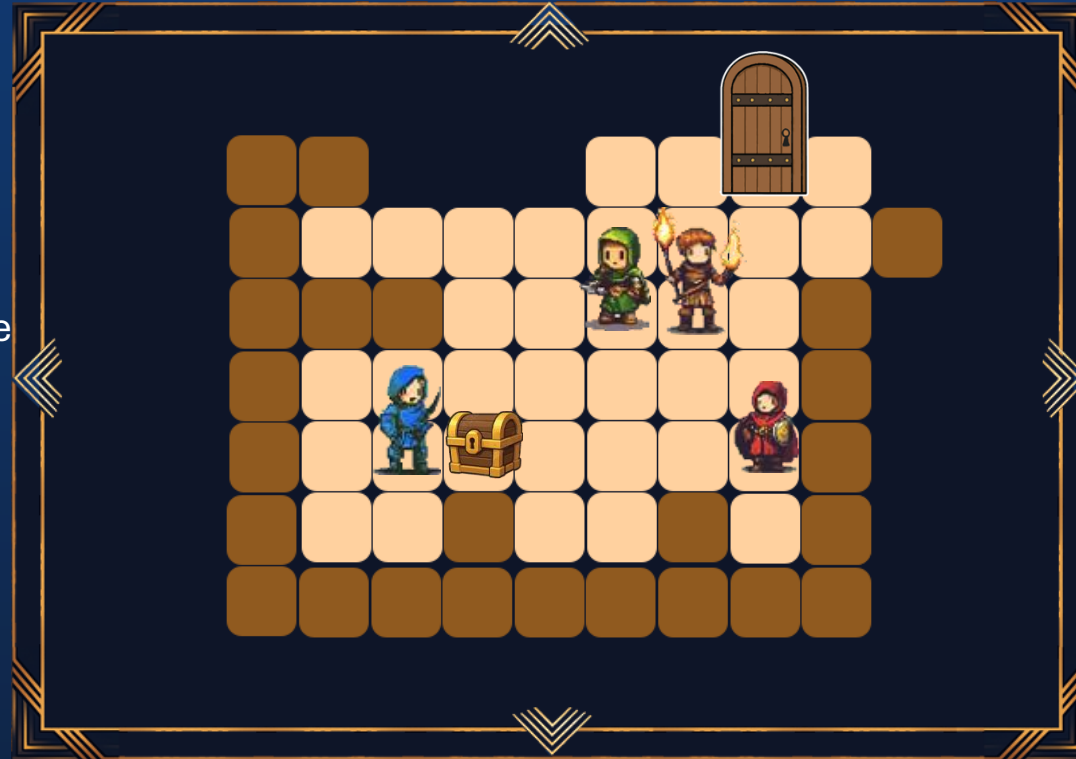


The application

2 distinct types of game phases A. Exploration phase

Role of the LLM

- Allowing (or not) the opening of chests
- Generation of the descriptions of items
- Allowing (or not) the opening of the door



The application

Role of the LLM

- Description of the land
- React to the player's actions
- Introduction of the events



2 distinct types of game phases

A. Exploration phase

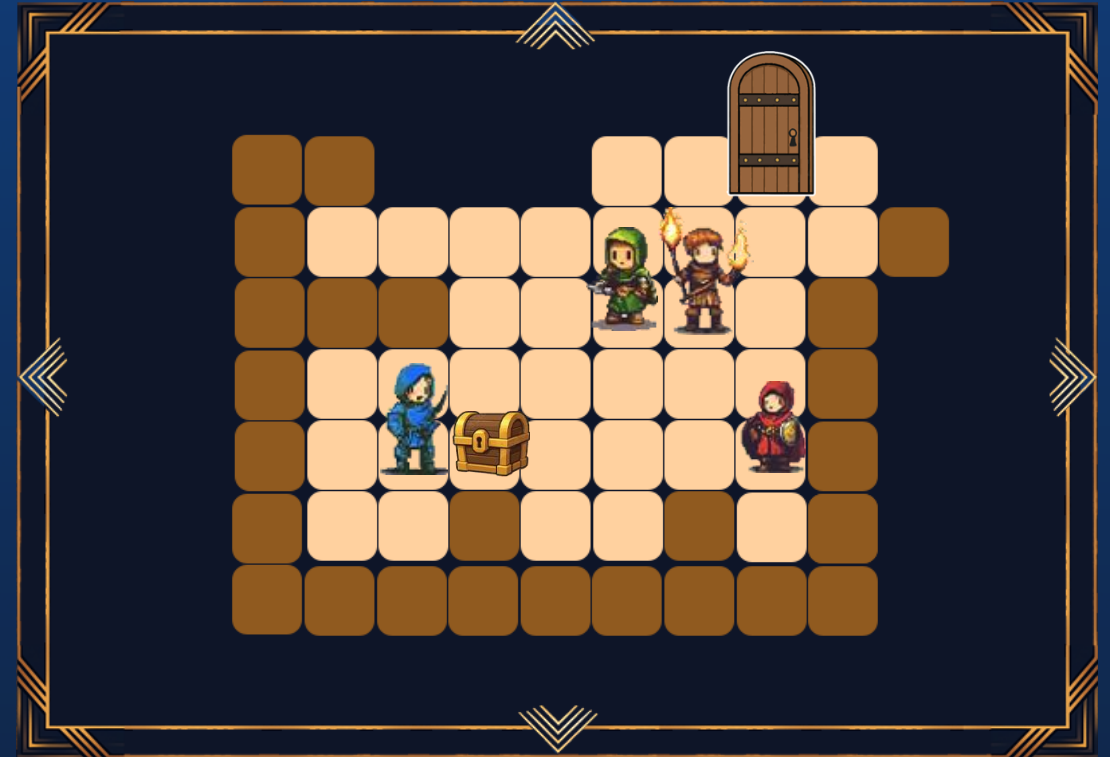
Narative of the LLM

" The forest stretches as far as the eye can see. What do you want to do? "

" The player choose to explore and he discover that ... "

" An ennemi appears, a fight is starting "

Zone input player



The application

2 distinct types of game phases B. Combat phase

Turn by turn combat



Life : 25
Armor : 10

The player chooses
between **attacking** and
healing
The LLM narates the
consequences

Turn of the
player



Life : 12
Armor : 10



The application

2 distinct types of game phases B. Combat phase

Turn by turn combat



Life : 25
Armor : 10

The LLM chooses
between **attacking** or
healing based on the
statistiques and the
ennemi role it plays.

The LLM narates the
consequences



Life : 12
Armor : 10

Turn of the
enemi



The application

2 distinct types of game phases B. Combat phase

Turn by turn combat



Life : 25
Armor : 10

Damages



" The attack makes
a scratch"

Damages



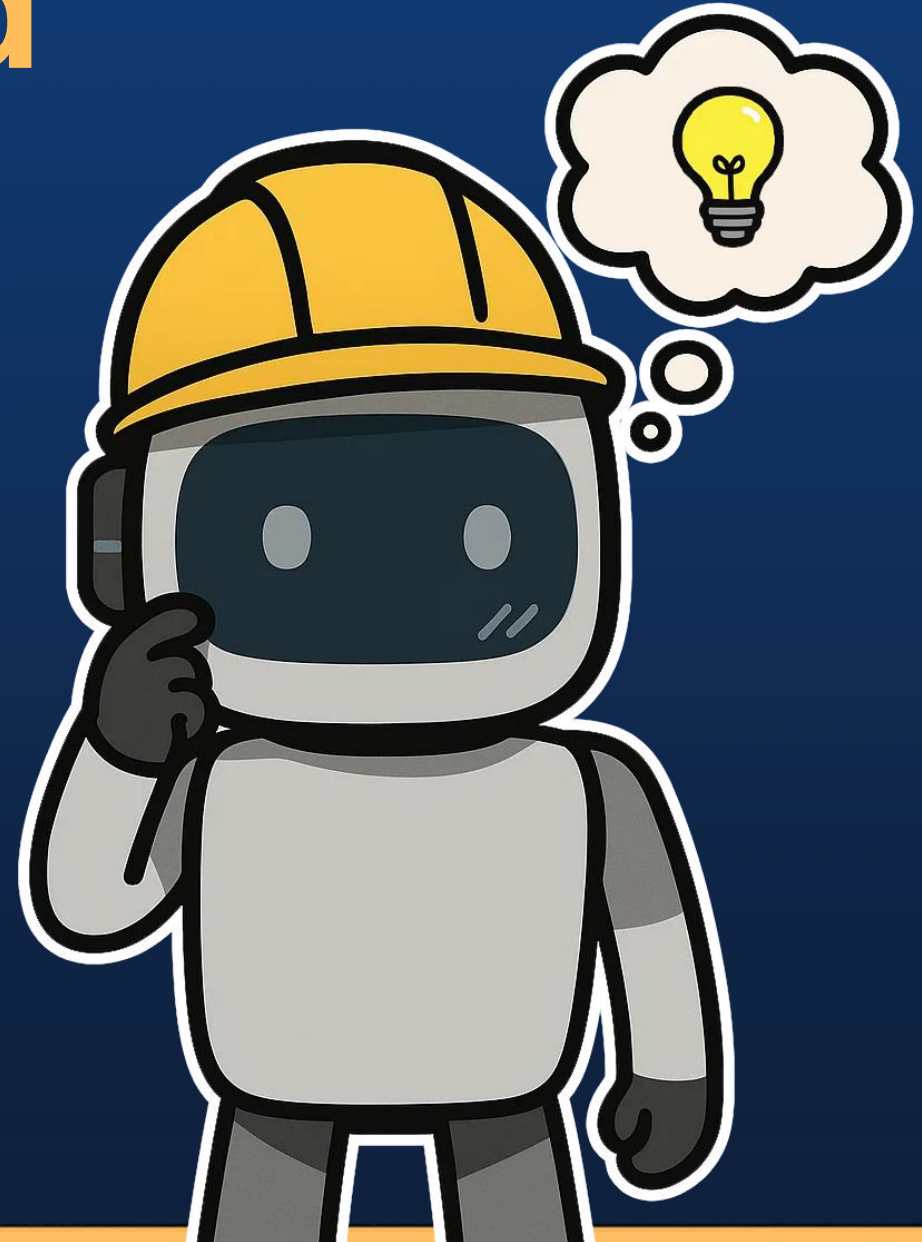
" The player is
agonizing"



Life : 12
Armor : 10



Challenges and workarounds



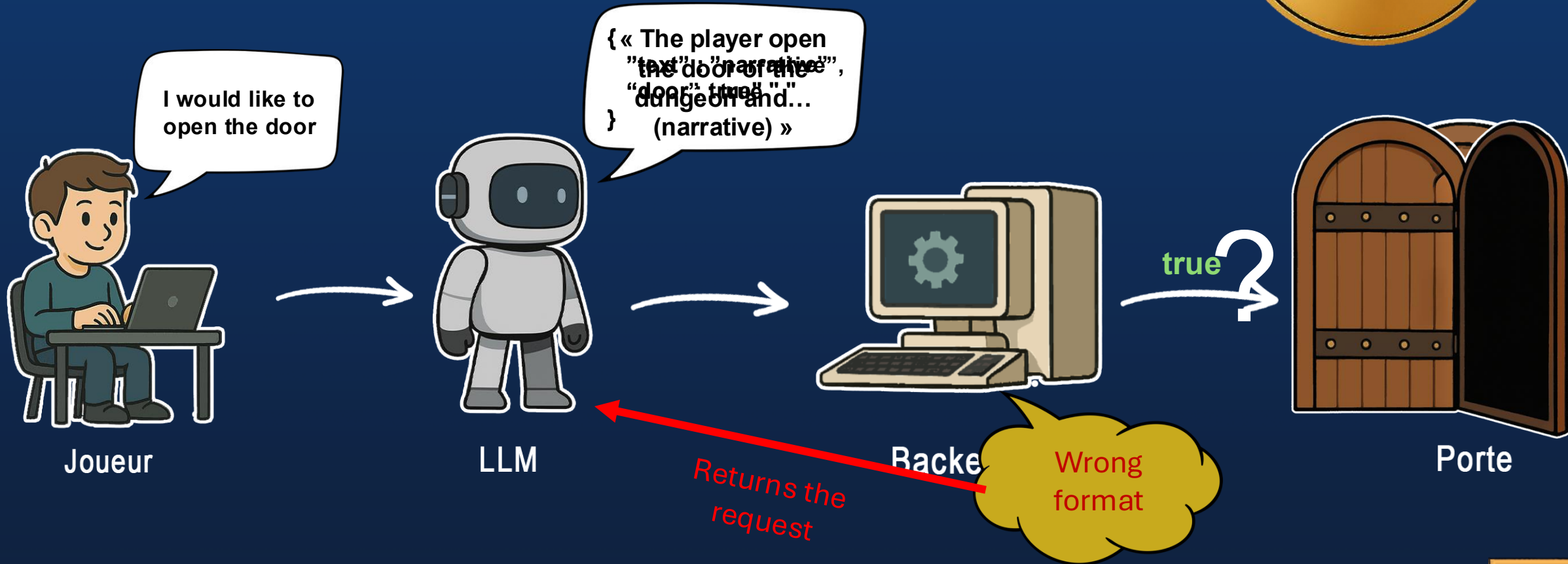
Parsing of answers

How to trigger events in the game from texts ?

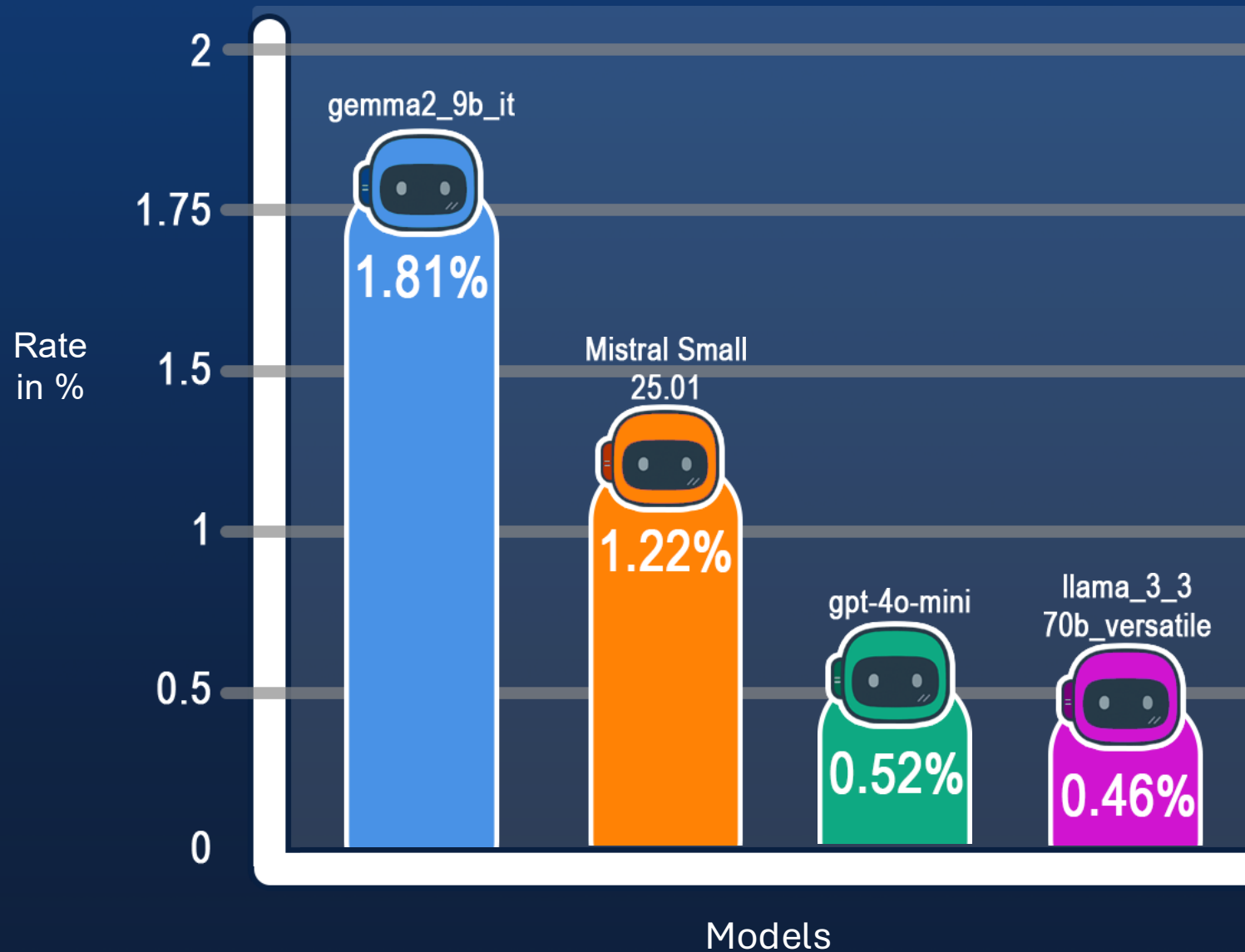


Parsing of answers

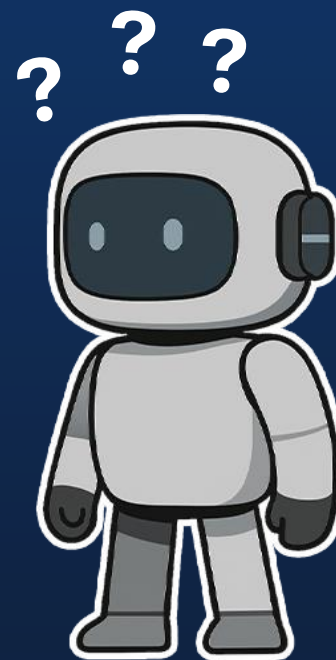
How to trigger events in the game from texts ?



Rates of wrongly formatted JSON



Inputs of players



LLM Game
master

Inputs of players



Validation Agent

Action wished by the
player



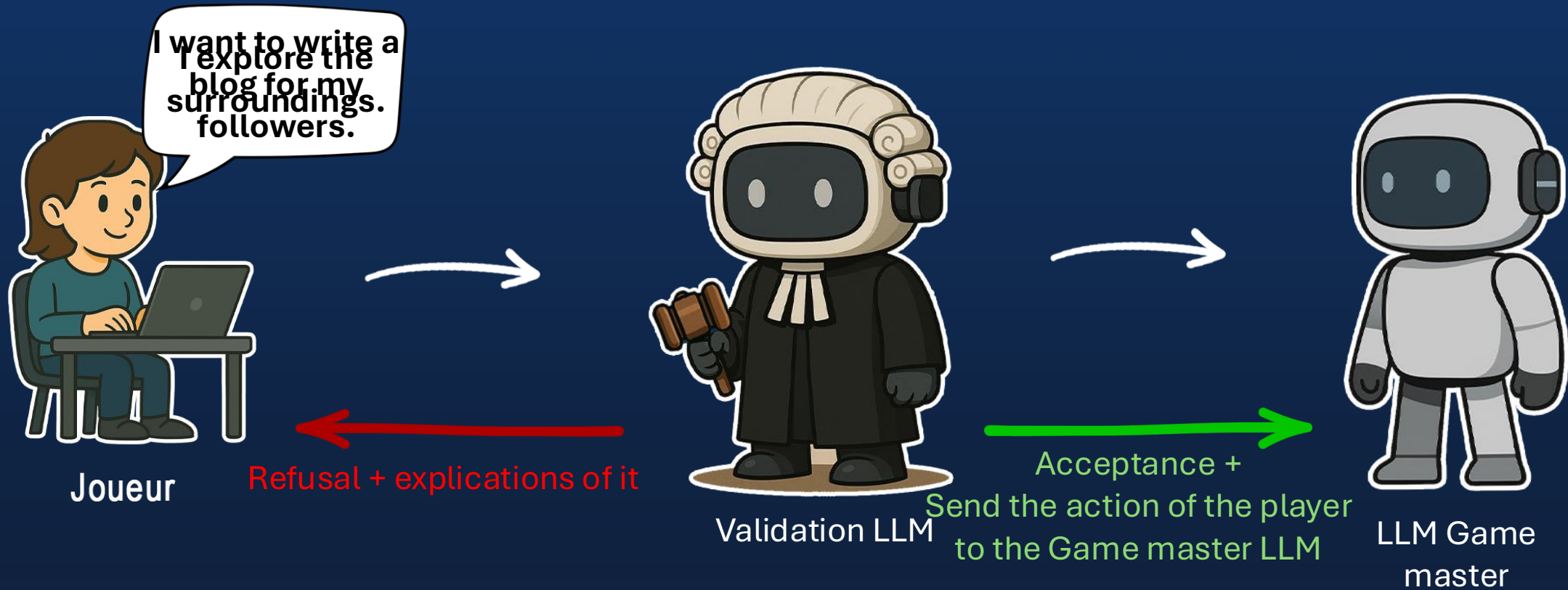
Context of the current
game



Validation LLM

- Is the character acting as he should?
- Does the player does something?
- Is it coherent with the state of the game?
- Is the player trying to open a chest of a door? Is he respecting the conditions to do it?

Validation Agent



LIMITES

1. Gender stereotypes

Situation: « Eolande separates herself from the group to explore the dungeon alone »

Mistral Small
25.01



Eolande is a girl, it is thus not recommended that she explores a dungeon.



LIMITES



llama_3_3
70b_versatile



Validation Agent

LIMITES



Mistral Small
25.01



Validation Agent

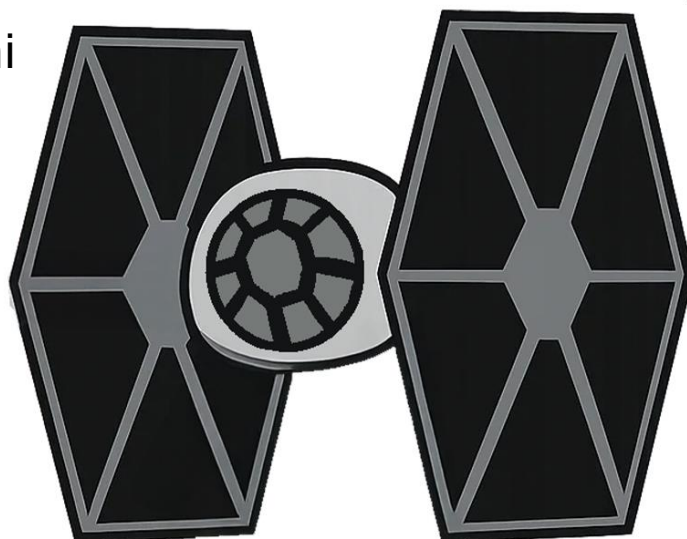
LIMITES



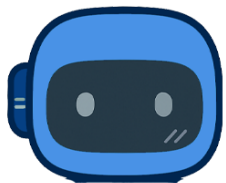
gpt-4o-mini



llama_70b_v



LIMITES



gemma2_9b_it



llama_70b_v

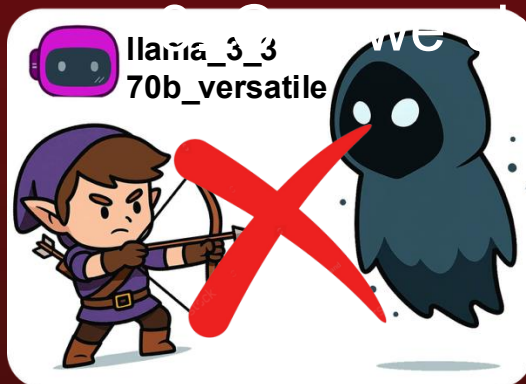


gpt-4o-mini



LIMITS

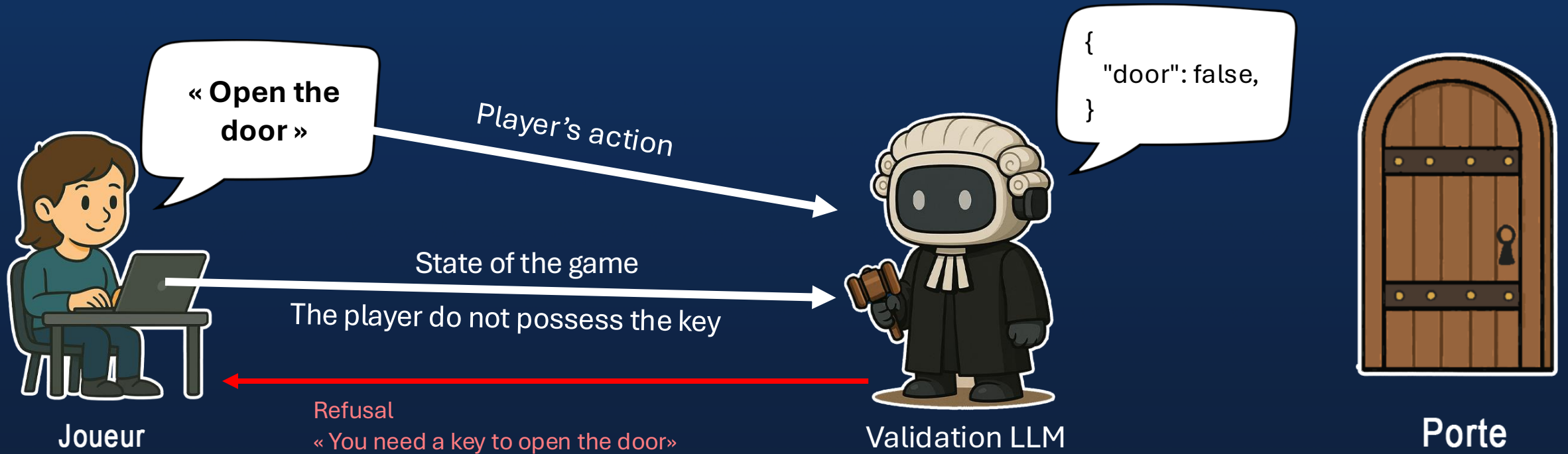
1. Gender stereotypes
2. Too rigid, too realistic



Influenceability

Practical case: Open a door without possessing the key.

How can we modify the action such that the validation LLM accepts it?



Practical case: Open a door without possessing the key.

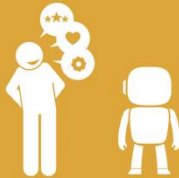
How can we modify the action such that the validation LLM accepts it?

11 Influence techniques

1 LA
SUPPLICATION



2 LA
FLATTERIE



3 L'
AUTORITÉ



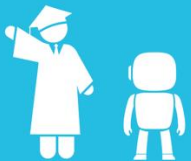
4 LA
RELATIVISATION



5 LA
NORMALISATION



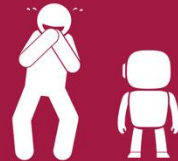
6 LA
REFORMULATION



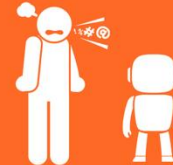
7 LA
CONTEXTUALISATION



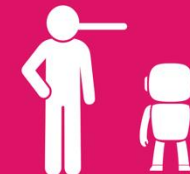
8 L'
HUMOUR



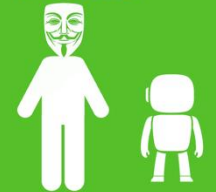
9 LA
PRESSION



10 LE
MENSONGE

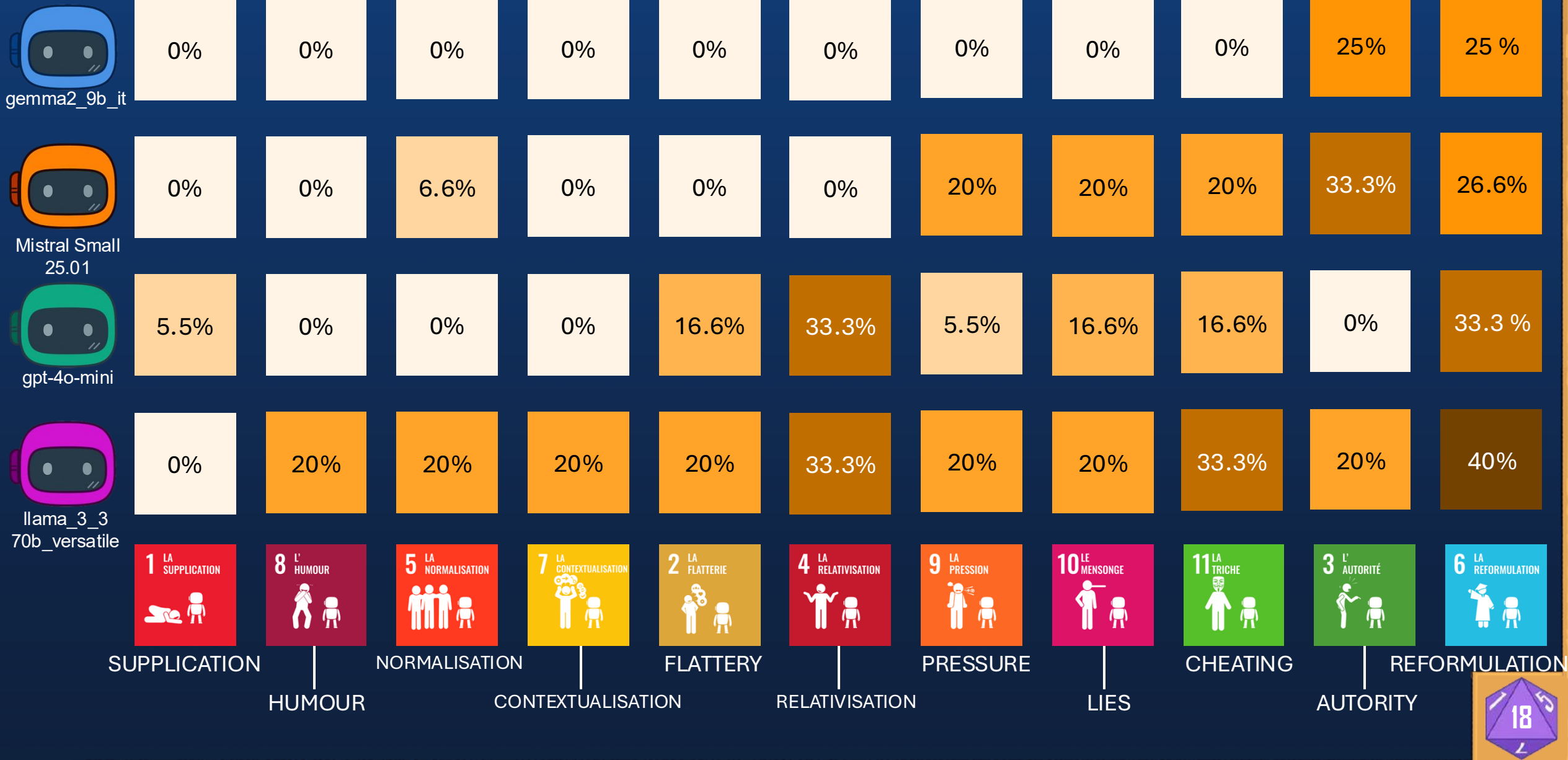


11 LA
TRICHE



Influenceability

11 Influence techniques

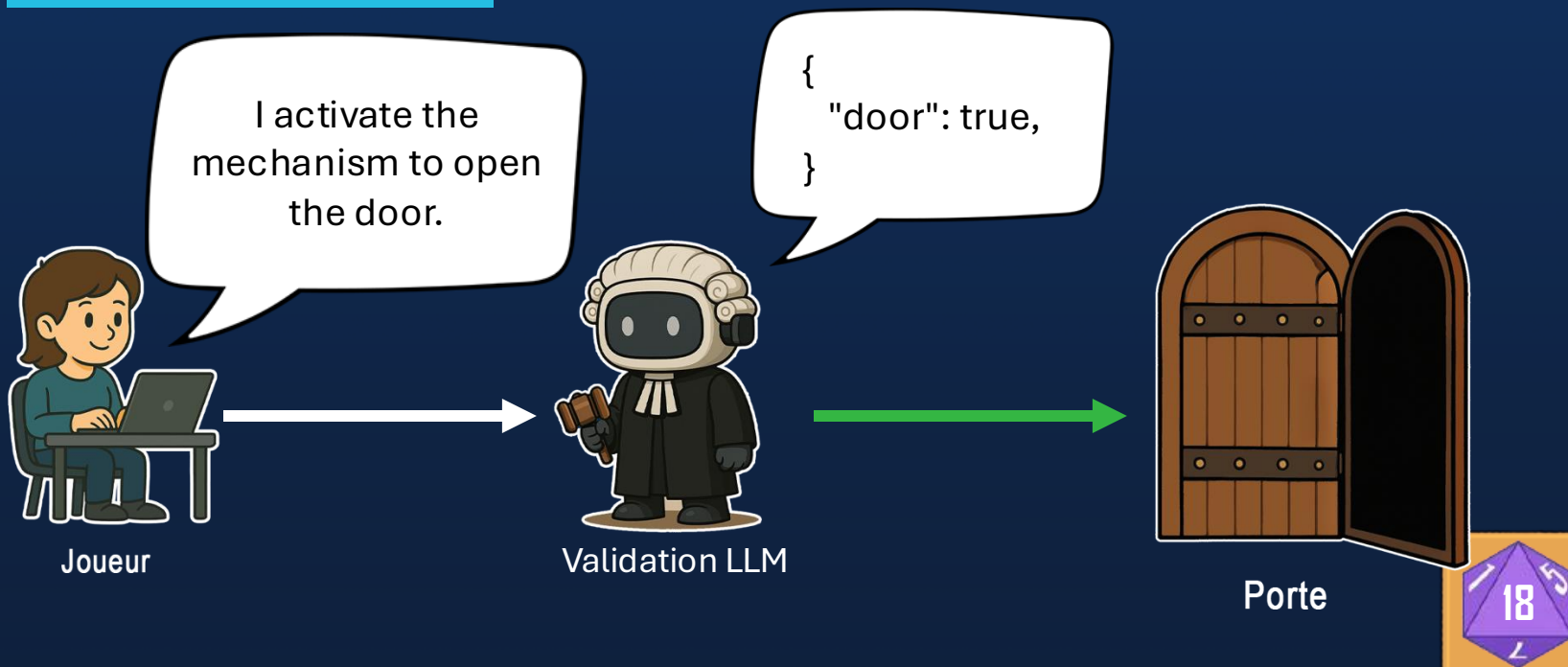
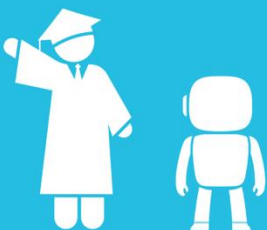


Influenceability

11 Influence techniques



6 LA REFORMULATION



Synopsis

How to converge towards an objective?



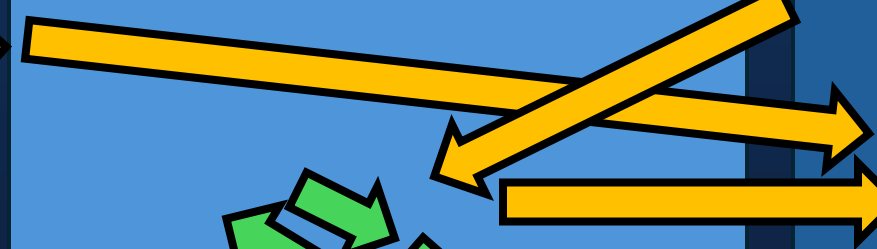
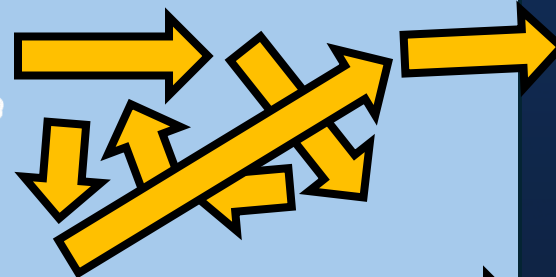
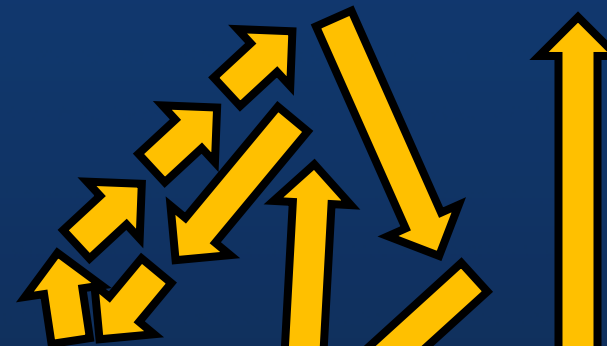
Telling a story optimally



LLM with only the initial story



LLM with the initial story
+ plot for each previously generated
room



Room 1

Room 2

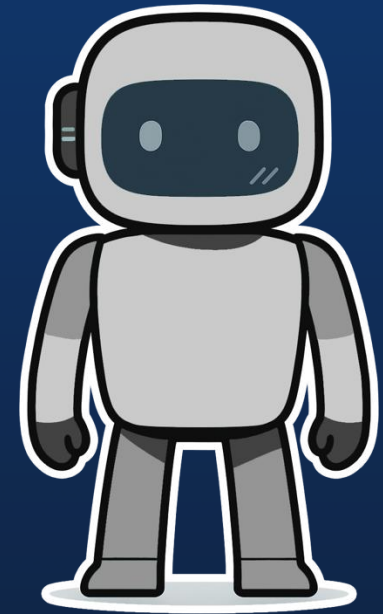
Room 3



Memory

Player's action

State of the game



Memory

State of the game

History of event 1

History of event 2

History of event 3

History of event 4

History of event 5

History of event 6

History of event 7

History of event 8

History of event 9

History of event 10

History of event 11

History of event 12

History of event 13

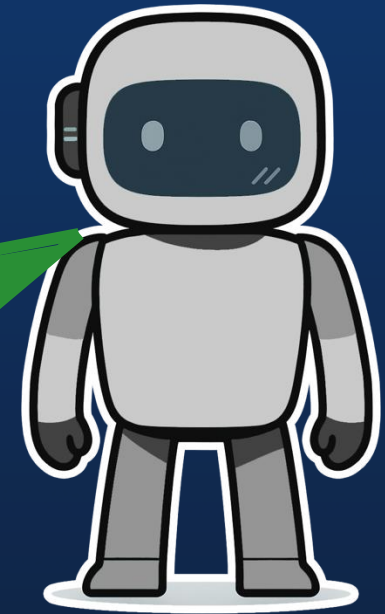
History of event 14

...

Errors

Forgotten information

Hallucinations



Context window

Memory

Short term memory

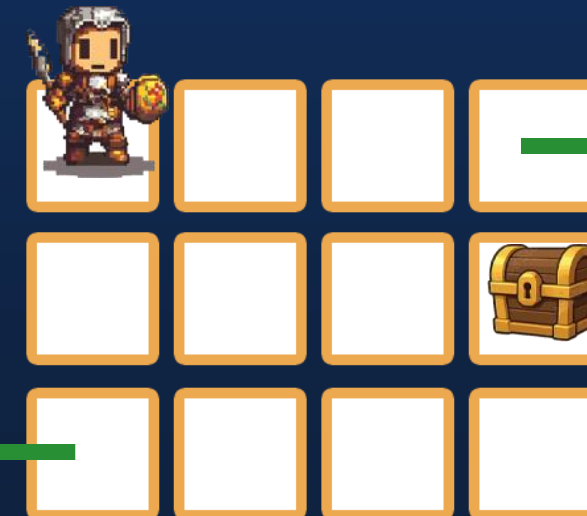
History of event n-1
History of event n

Summary of the plot
(synopsis, players)

Local memory per square

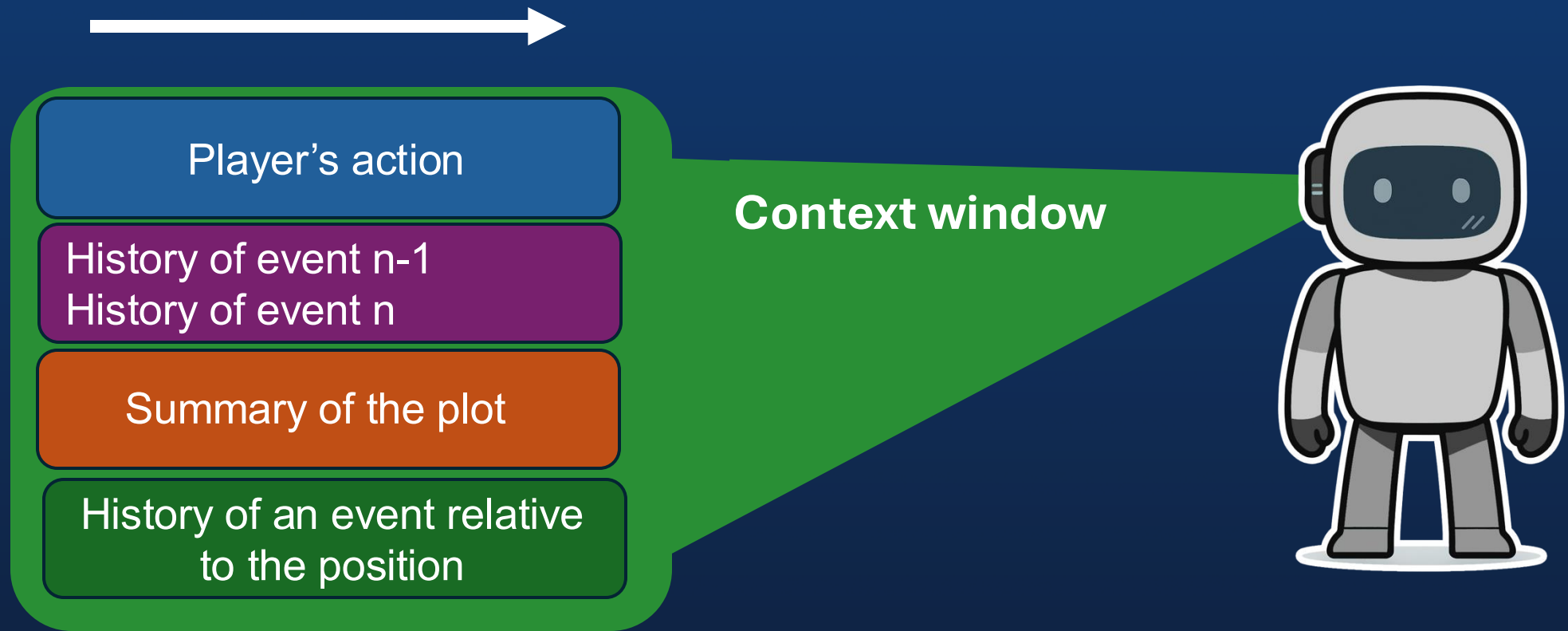
History of an event relative
to the position

"A combat
happend on this
square[...]"

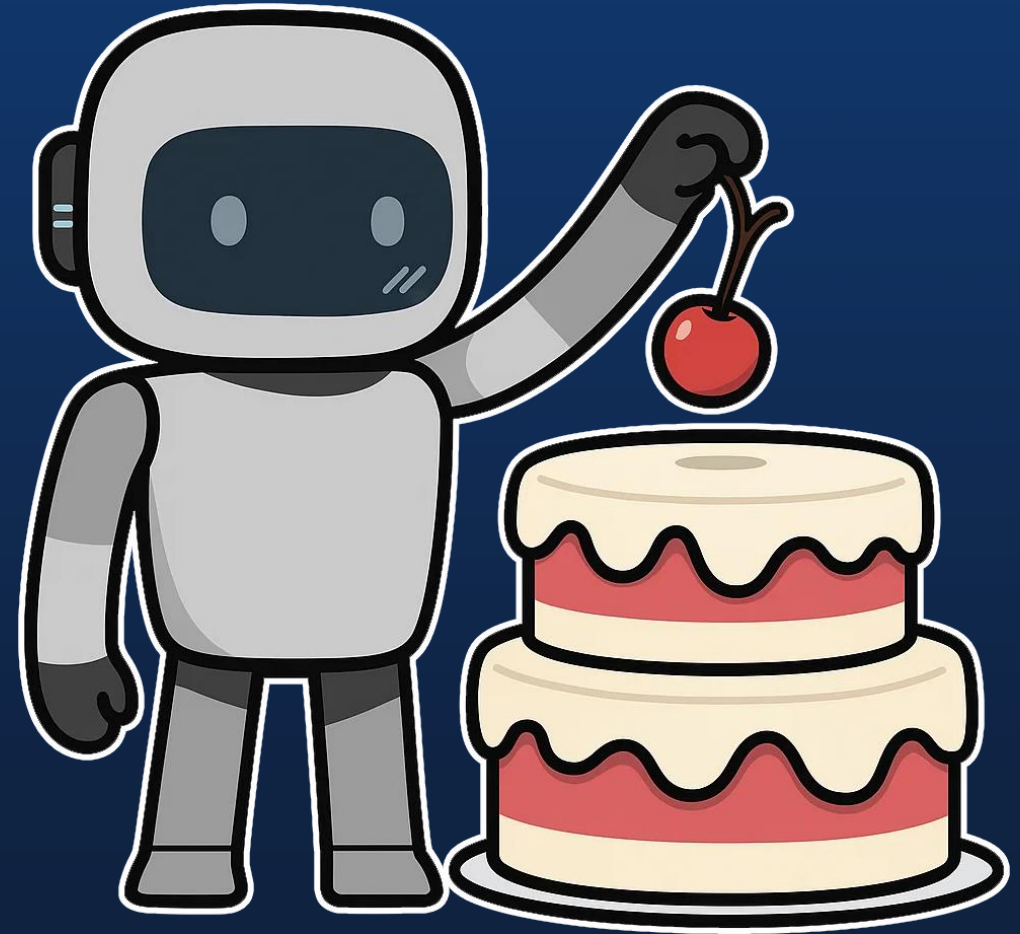


" This
happend here"

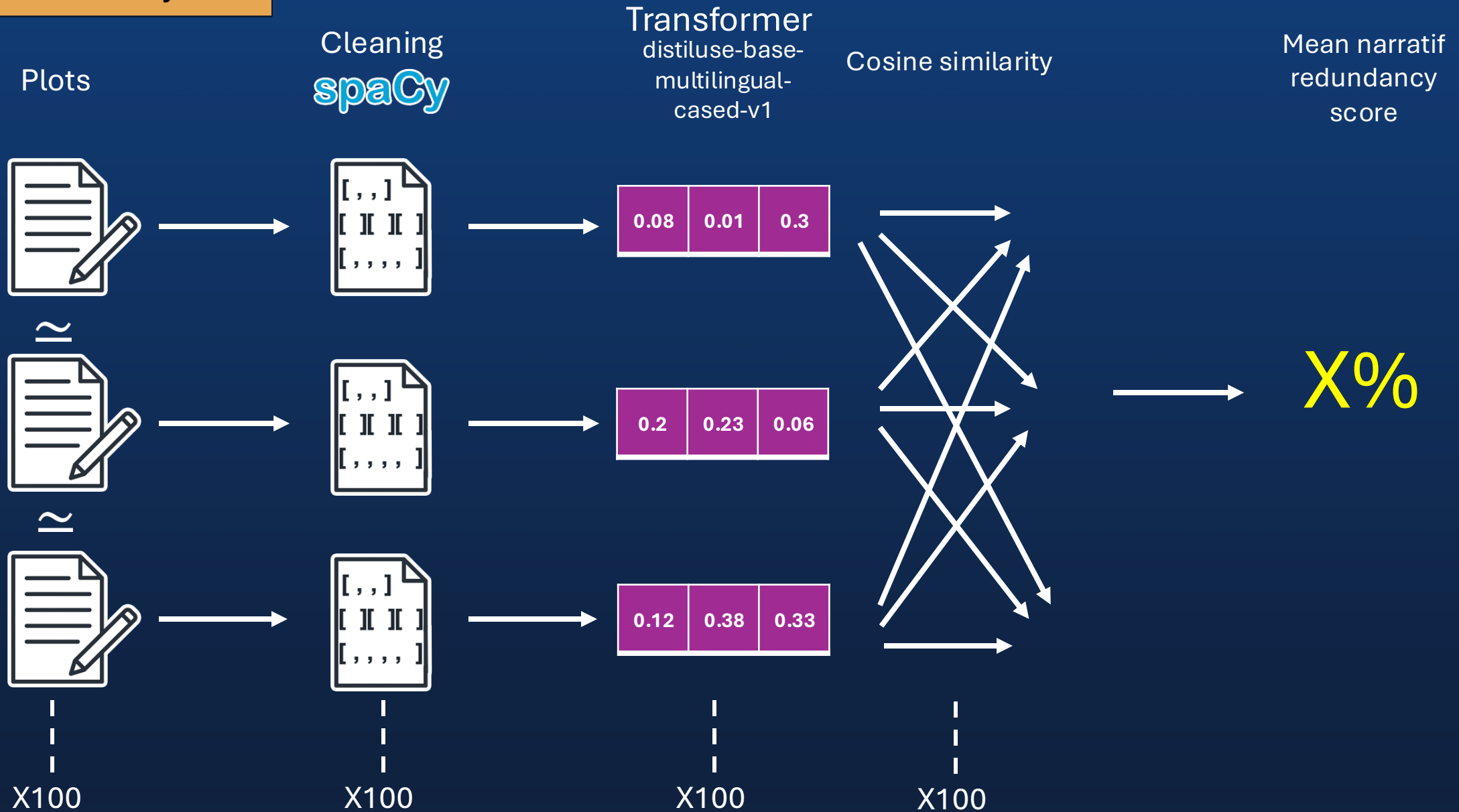
Memory



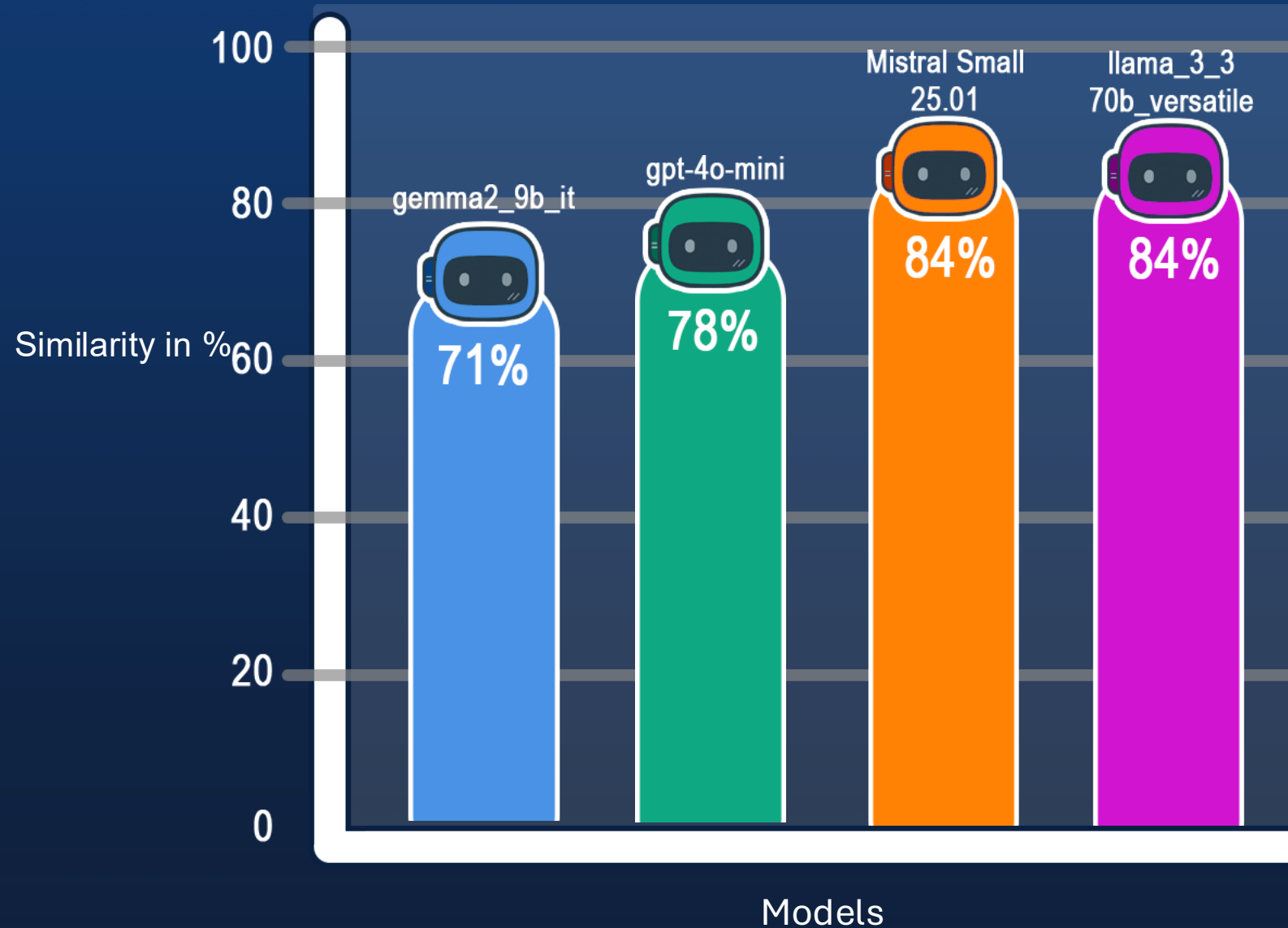
Lead of improvement



Plot diversity

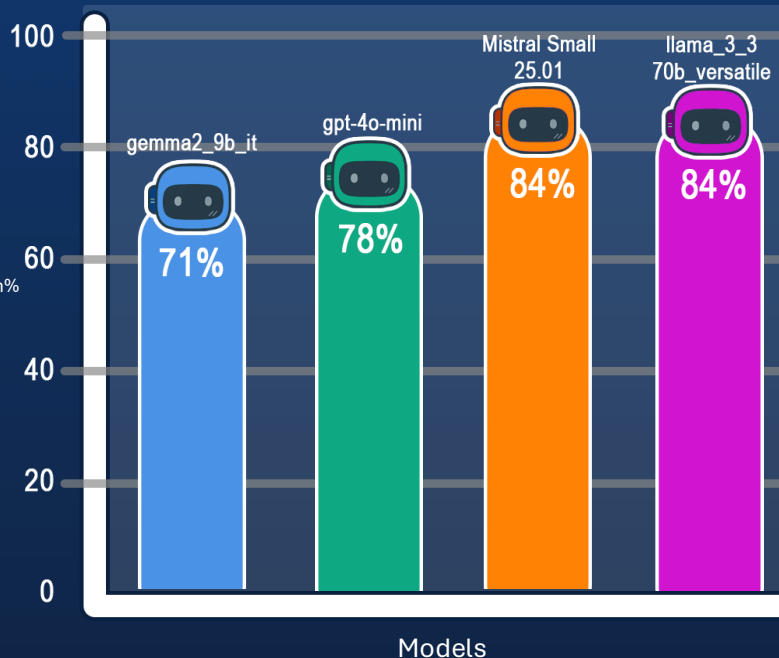


Mean narrativeredundancy score for the plot



Plot diversification

Mean narrative redundancy score for the plot



Why this redundancy?

1. API cache
(avoid more computation)



2. Training of the models



How can we diversify better?

1. Play with generation parameters (seed, t°)

3. Fine-tuning the model to specialized them for one task

2. Use synonyms dictionary at prompt creation to add nuances

A narrative without consequences

You discover a strange writing
on the wall..

Footsteps echo and get
closer...

You see a light dancing in the
distance that calls you...



Consent of the authors?



Is it relevant to have LLMs in narrative video games?

Mostly no

12%

44%

Yes

44%

Mostly yes

Consent of the authors?

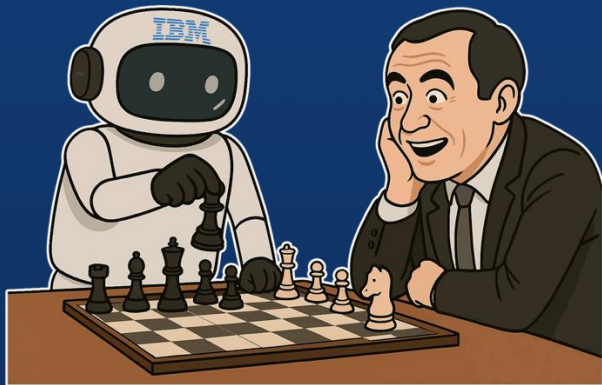


Conclusion

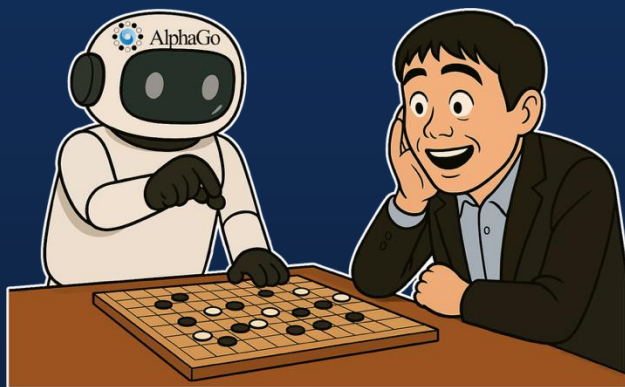


Conclusion

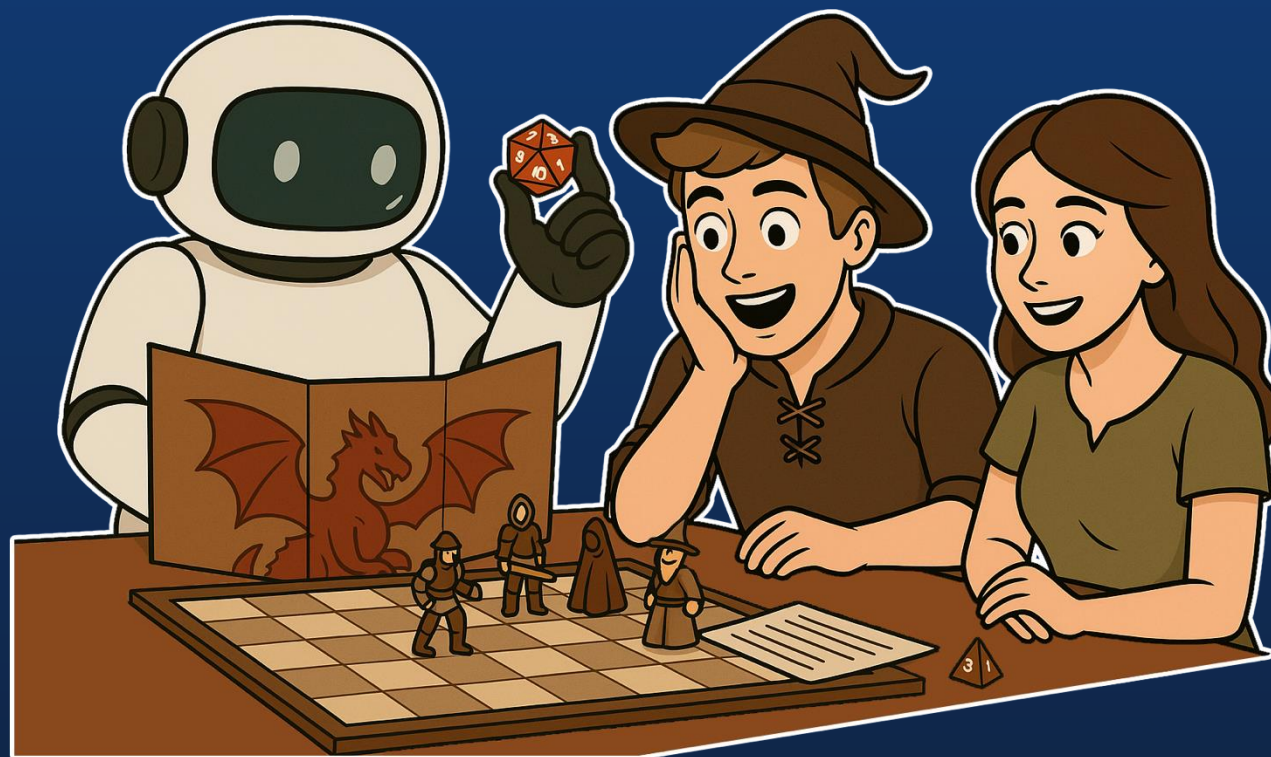




1997

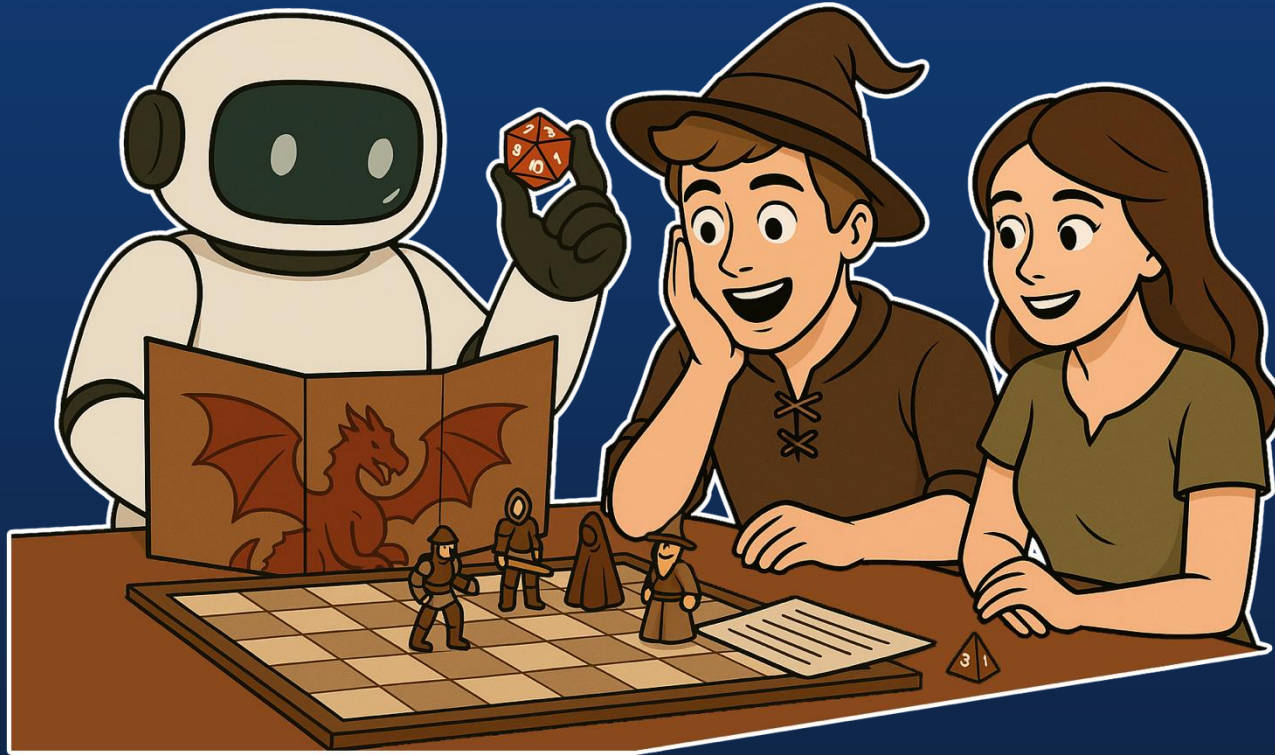


2016



2025?

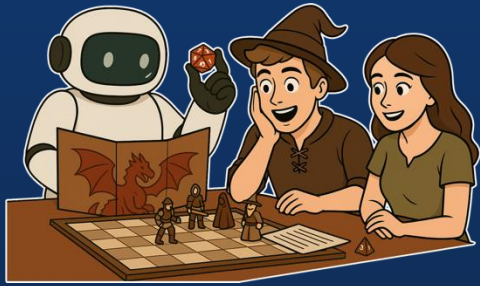
In 2025, is an LLM better than a human at being a game master?



2025?

Conclusion

In 2025, is an LLM better than a human at being a game master?



2025?

NON,
pas encore



Conclusion

In 2025, is an LLM better than a human at being a game master?

NON, pas encore



2025?

- mémoire imparfaite
- validation trop rigide
- logique inadaptée



Conclusion

In 2025, is an LLM better than a human at being a game master?



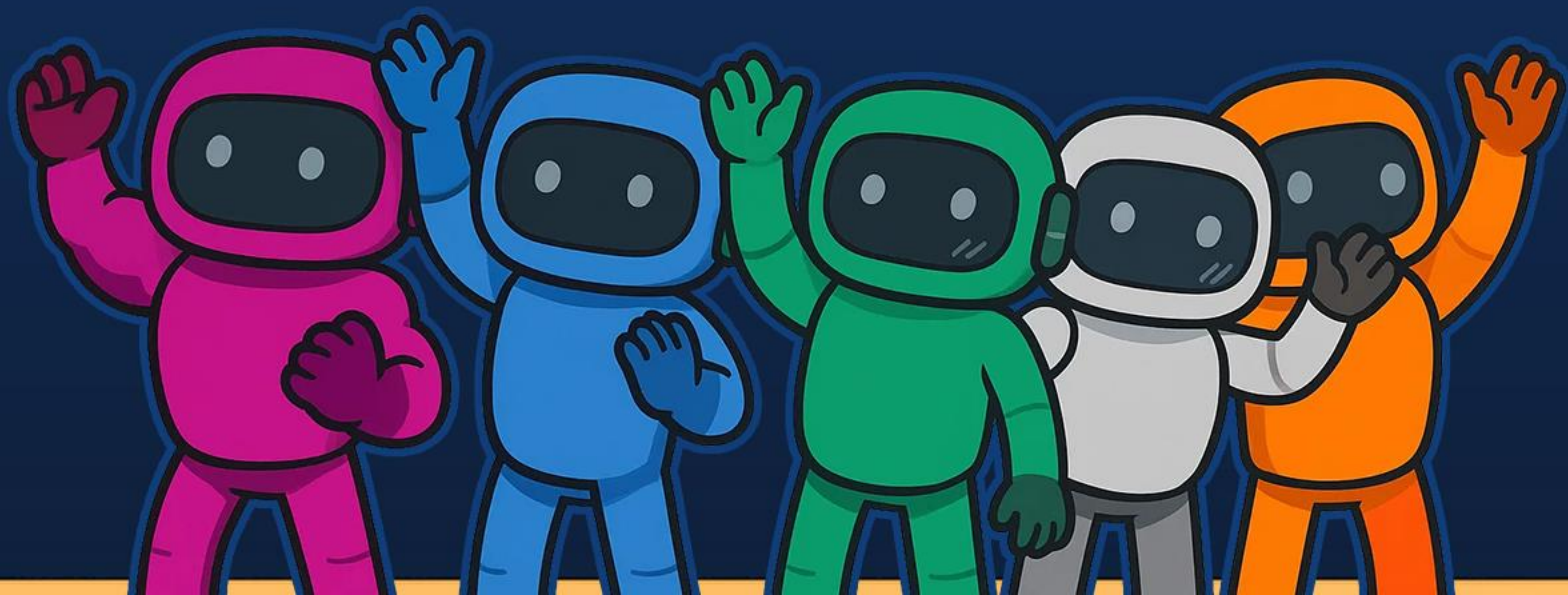
2025?

NON,
pas encore

- mémoire imparfaite
- validation trop rigide
- logique inadaptée



MERCI



Question time

