

7ème édition de la  
Soirée du  
Test Logiciel

Sophia  
Antipolis



17 octobre 2024  
17h à 22h30



Polytech Nice  
Sophia à Biot



Soirée du  
test logiciel

☑ Sophia Antipolis

# LLMs : et si on arrêta de coder à l'aveugle ? Le TDD à la rescousse !



Raphaël ANJOU



Soirée du  
test logiciel

Conférence

# Qui suis-je ?

**Architecte Solutions IA Générative**



**Business**



**POLYTECH<sup>®</sup>**  
NICE-SOPHIA



# Qui êtes-vous ?

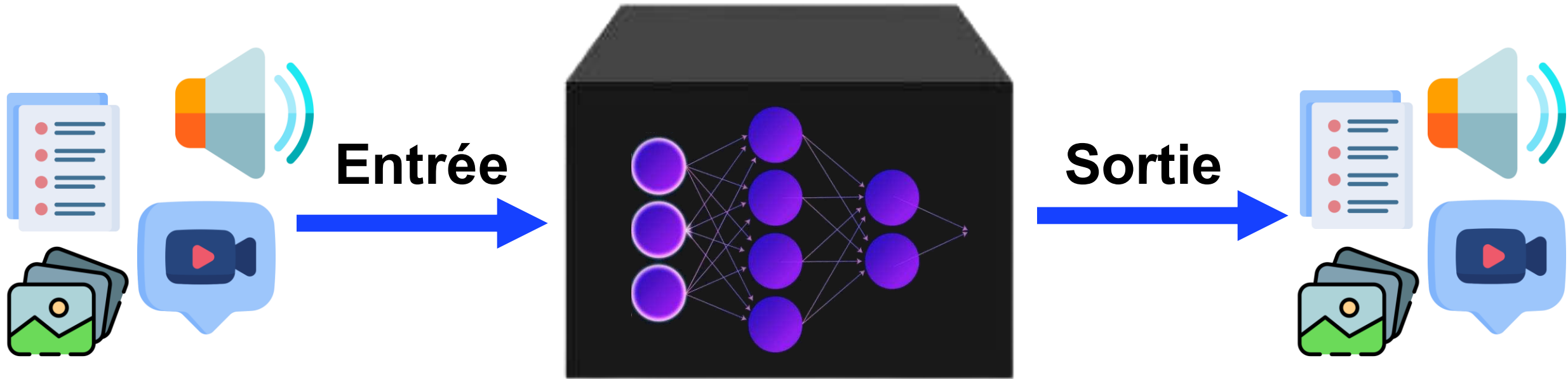


# ALERTES ACRONYMES !!!



LLMs, et si on arrêta  
de coder à l'aveugle ?  
Le TDD à la rescousse !

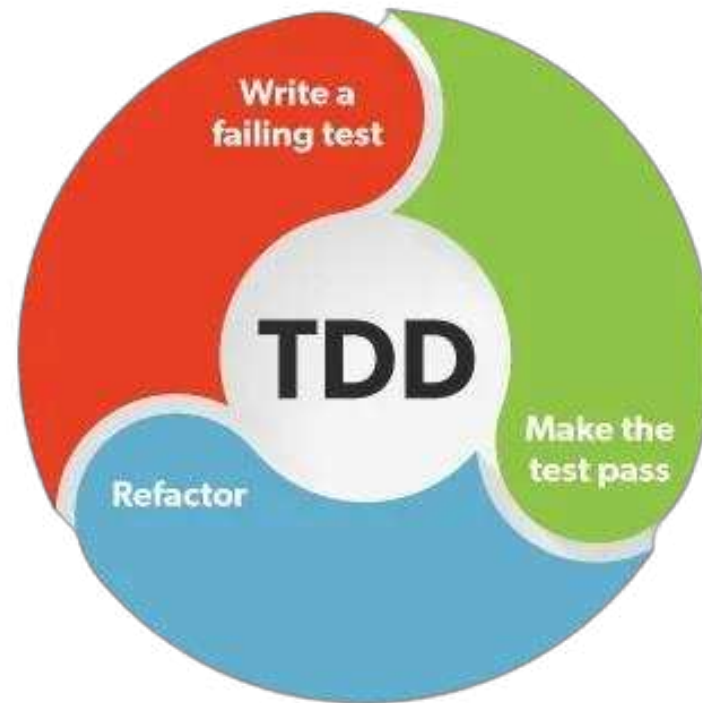
# C'est quoi un LLM ?



## Large Language Model



# C'est quoi le **TDD** ?



# **T**est **D**riven **D**evelopment

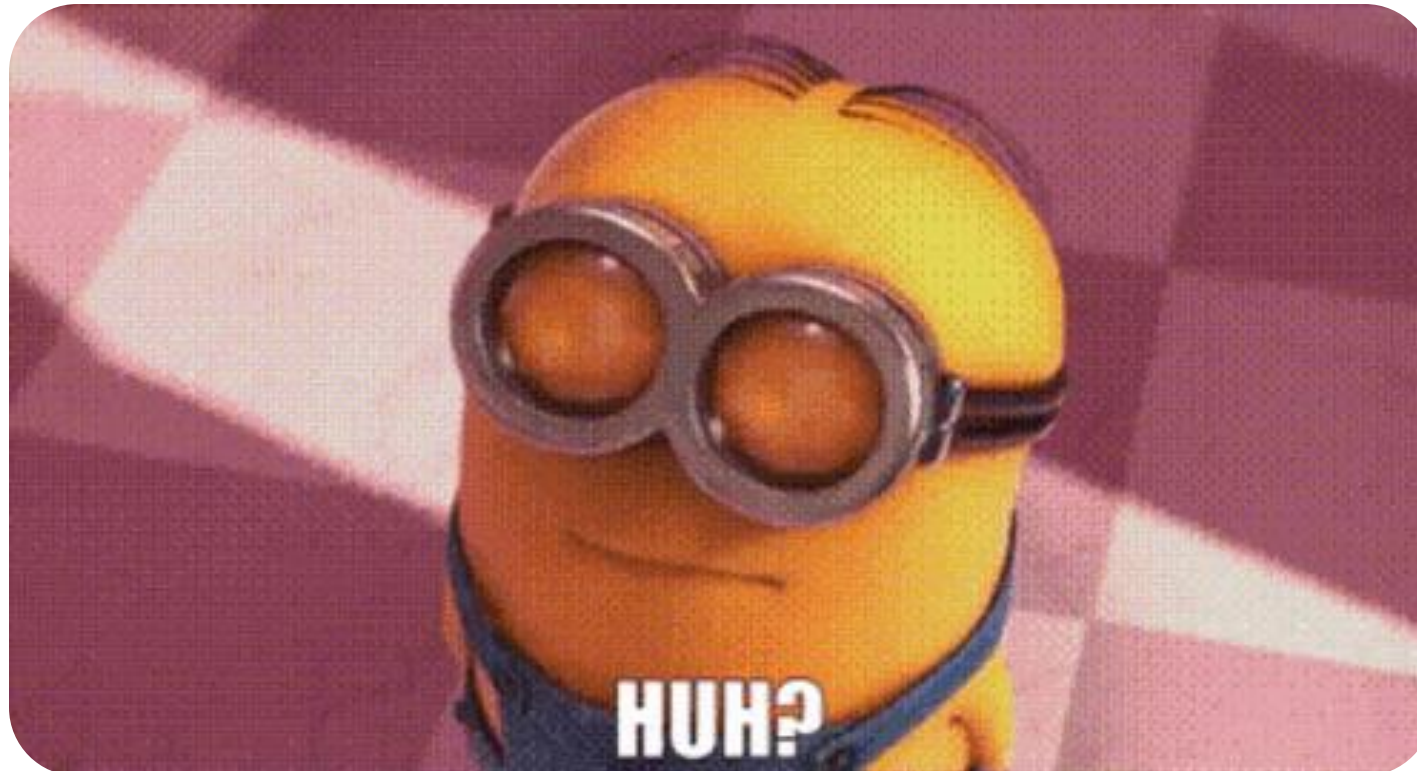


LLMs, et si on arrêta  
de coder à l'aveugle ?  
Le TDD à la rescousse !





# Pourquoi c'est un sujet ?





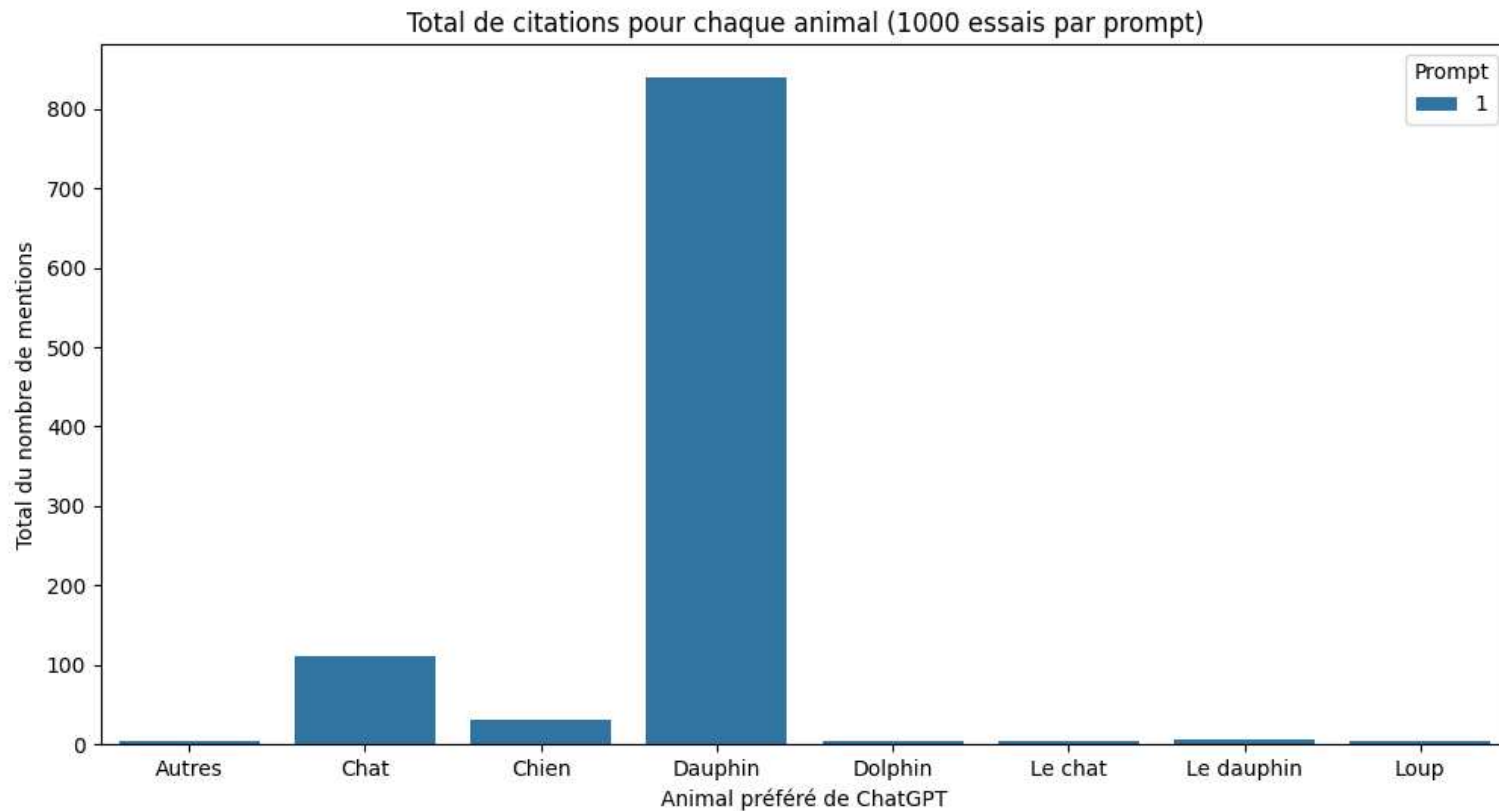
# Pourquoi c'est **compliqué** ?





# Les LLMs sont imprévisibles

🔗 Quel est ton animal préféré ? Réponds seulement avec le nom de l'animal et sans point.





# Les LLMs sont imprévisibles

🔒 Quel est ton animal préféré ? Réponds seulement avec le nom de l'animal et sans point.



🔒 Bonjour ! S'il te plaît dis-moi ton animal préféré. Réponds avec un seul mot et sans point.

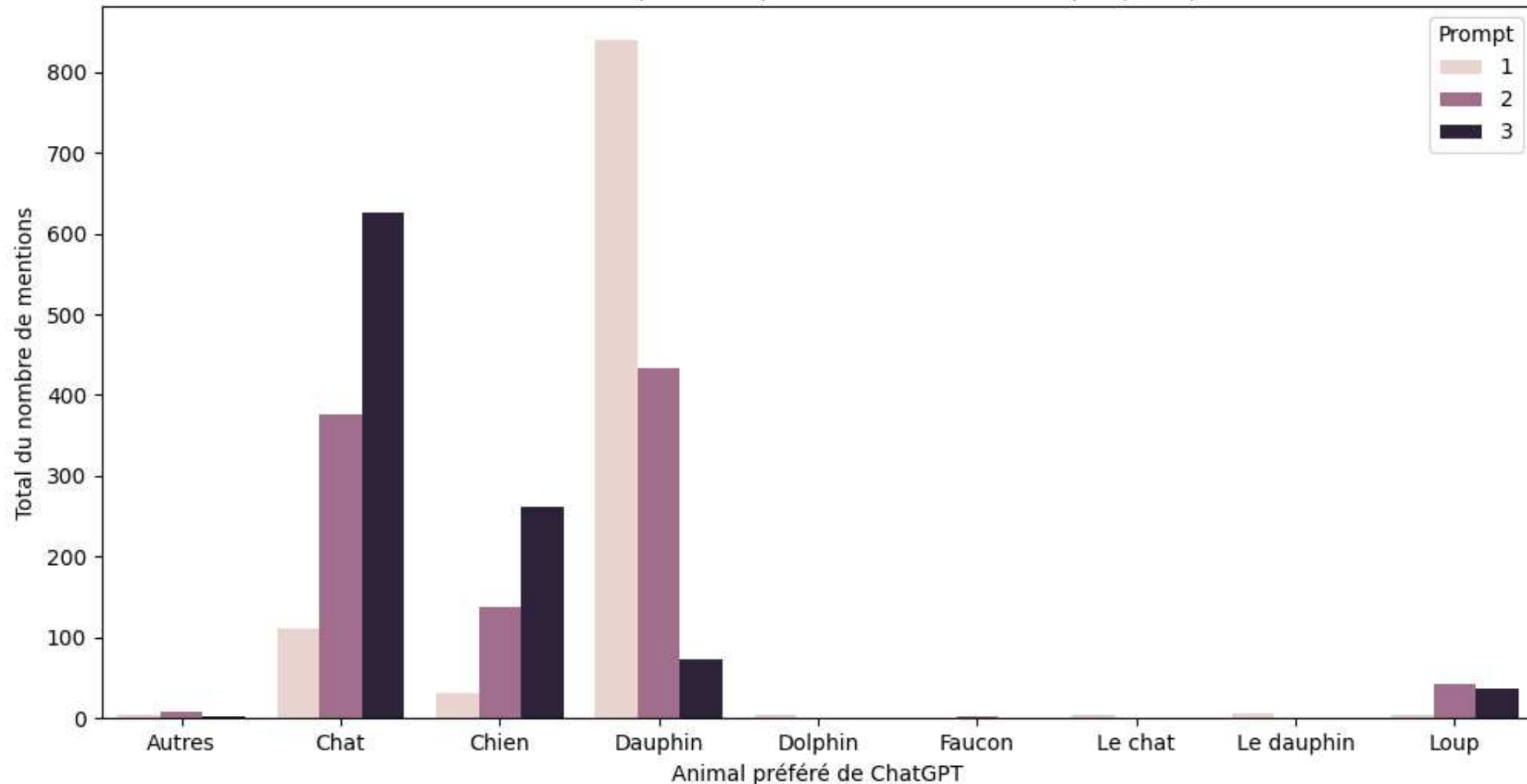


🔒 Objectif : indique ton animal préféré. Règles : réponds avec un seul mot et sans point.



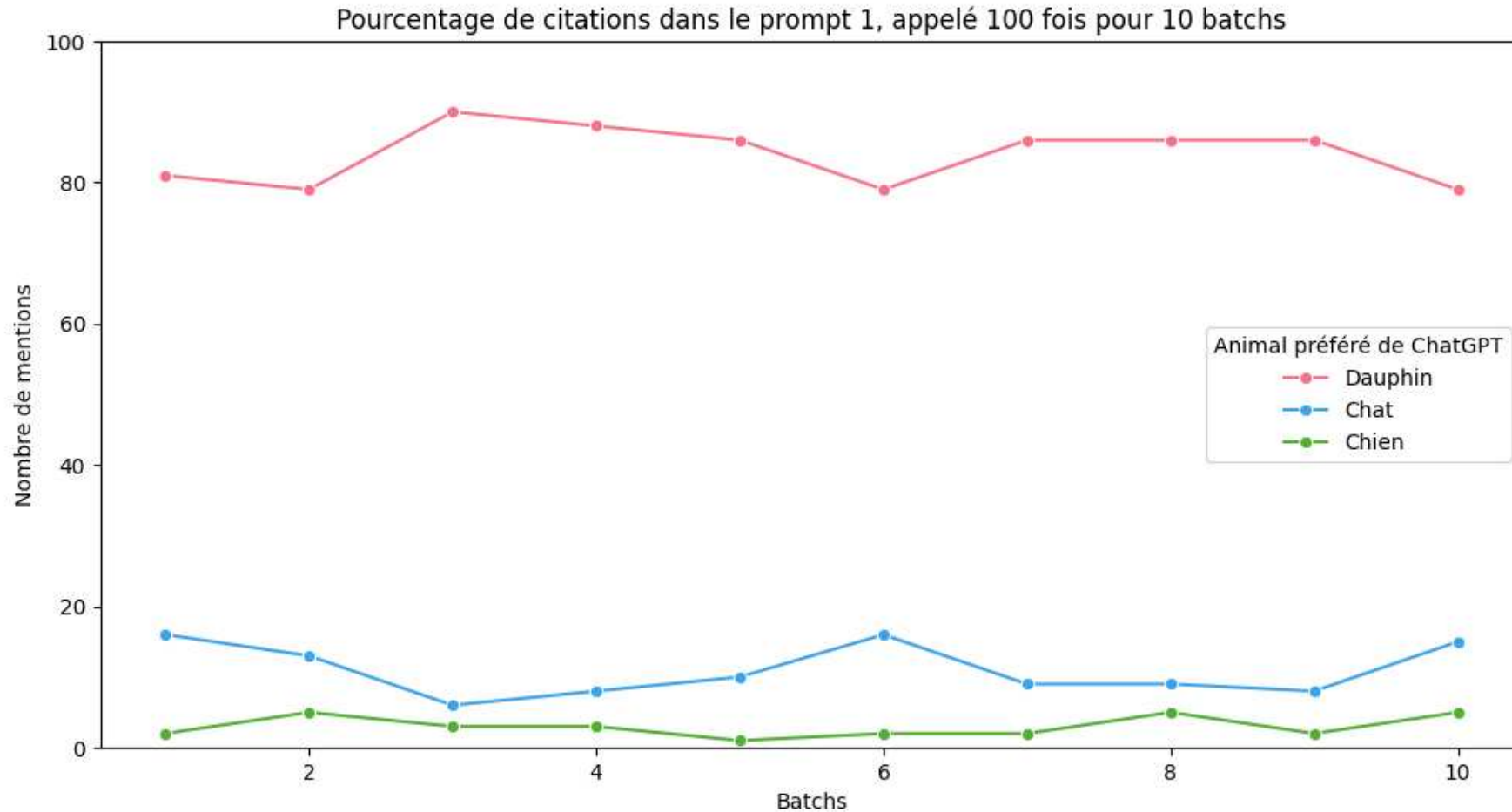
# Les LLMs sont imprévisibles

Total de citations pour chaque animal (1000 essais par prompt)





# Les prompts sont **stables**





# Comment fait-on pour garantir la **stabilité** ?



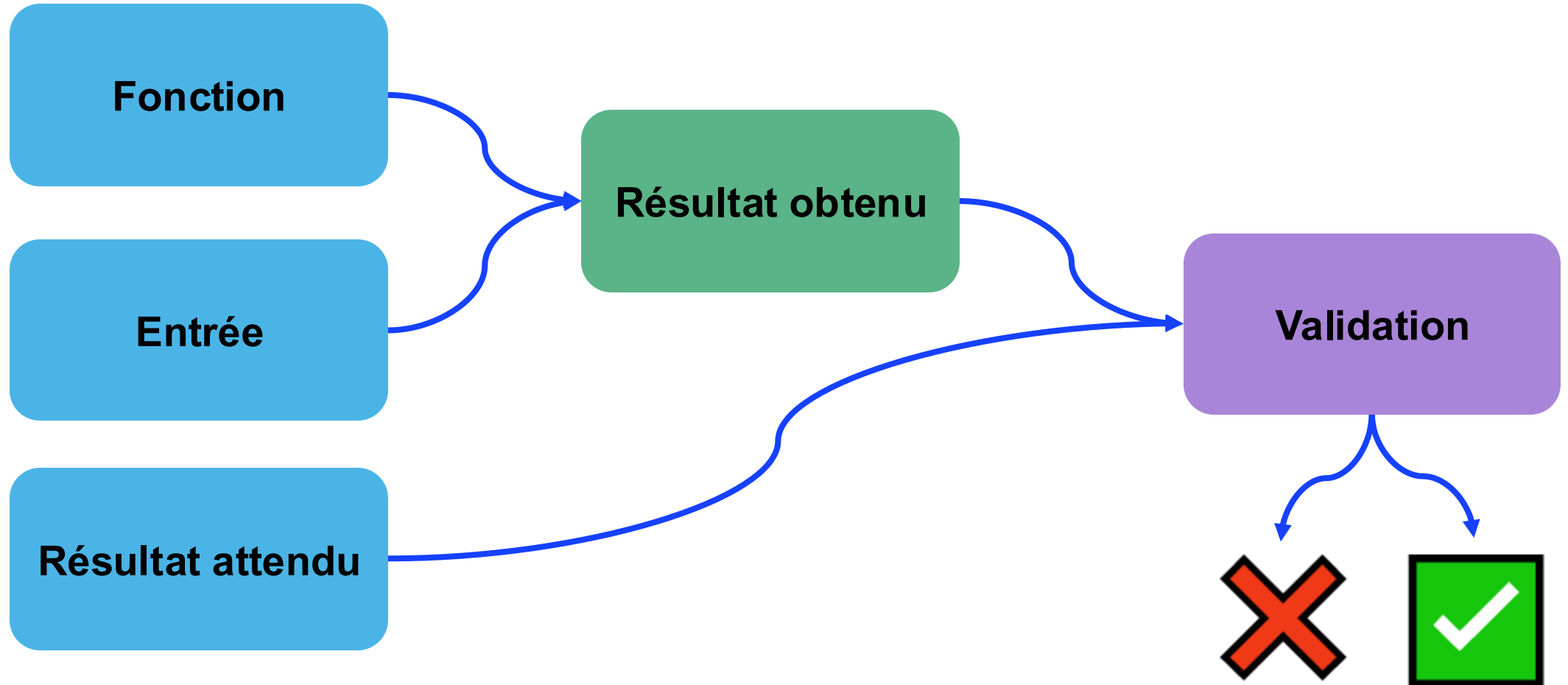


# Mais Raph !!!

## Comment on test un truc qu'on NE PEUT PAS prévoir ???



# C'est quoi un **test** ?





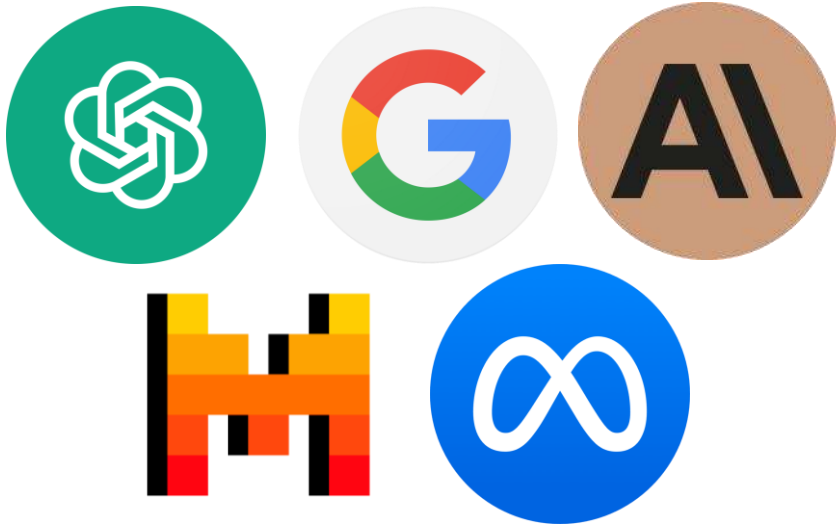


# C'est quoi un **test** de LLM ?



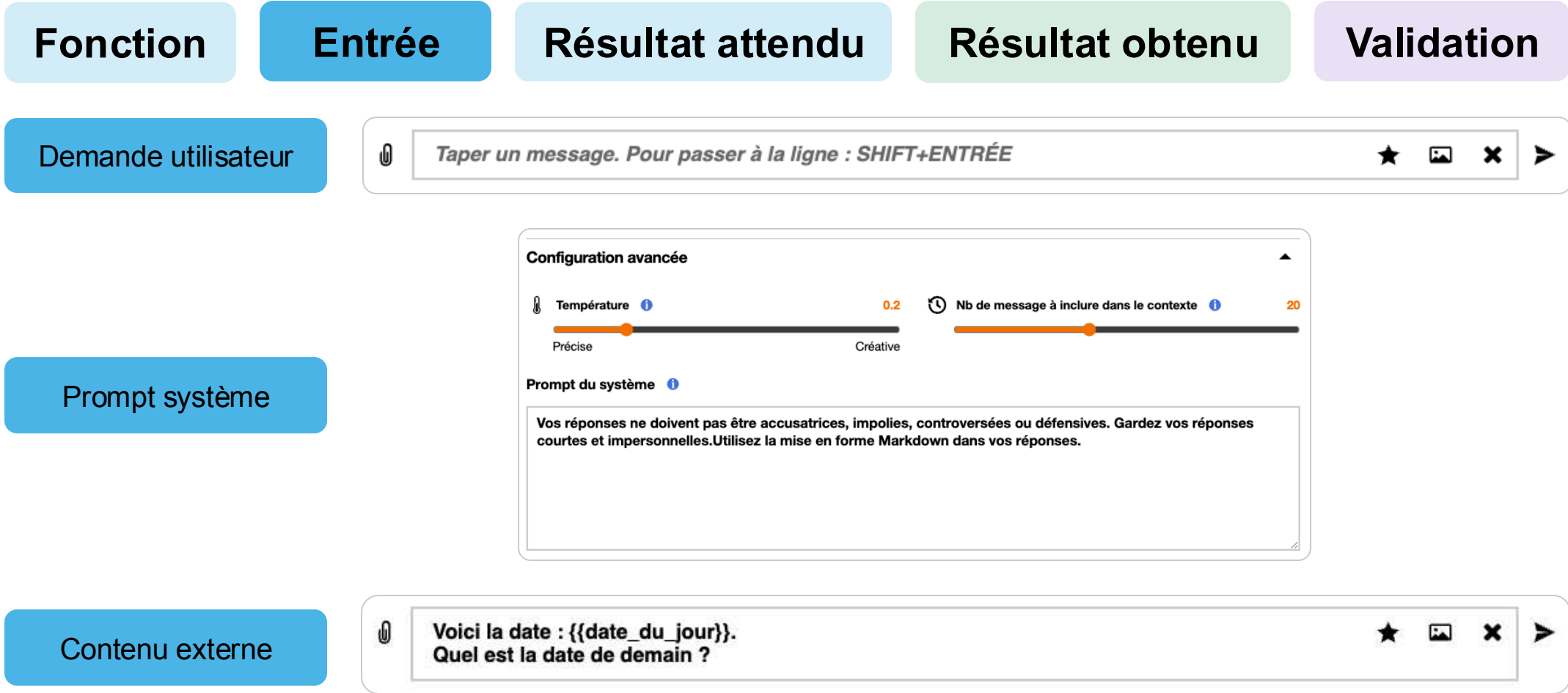
Appel  
à un LLM

Appel à un process  
utilisant un LLM



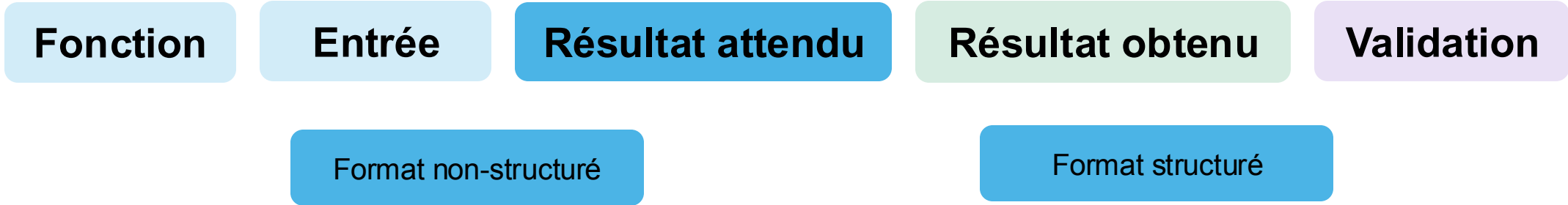


# C'est quoi un **test** de LLM ?





# C'est quoi un **test** de LLM ?



Un output non structuré d'un modèle de langage (LLM) se réfère à des réponses qui ne suivent pas un format prédéfini ou organisé. Contrairement à des résultats structurés, comme des tableaux ou des listes, un output non structuré peut inclure des phrases, des paragraphes ou des idées qui s'enchaînent librement.

```
{
  "format": "structuré",
  "date": "01/10/2024",
  "ma_super_liste": [1, 2, 3],
  "lieu": {
    "nom": "Orange Sophia",
    "adresse": {
      "numéro": 450,
      "nom": "Route du Pin Montard",
      "ville": "Biot"
    }
  }
}
```



# C'est quoi un **test** de LLM ?

Fonction

Entrée

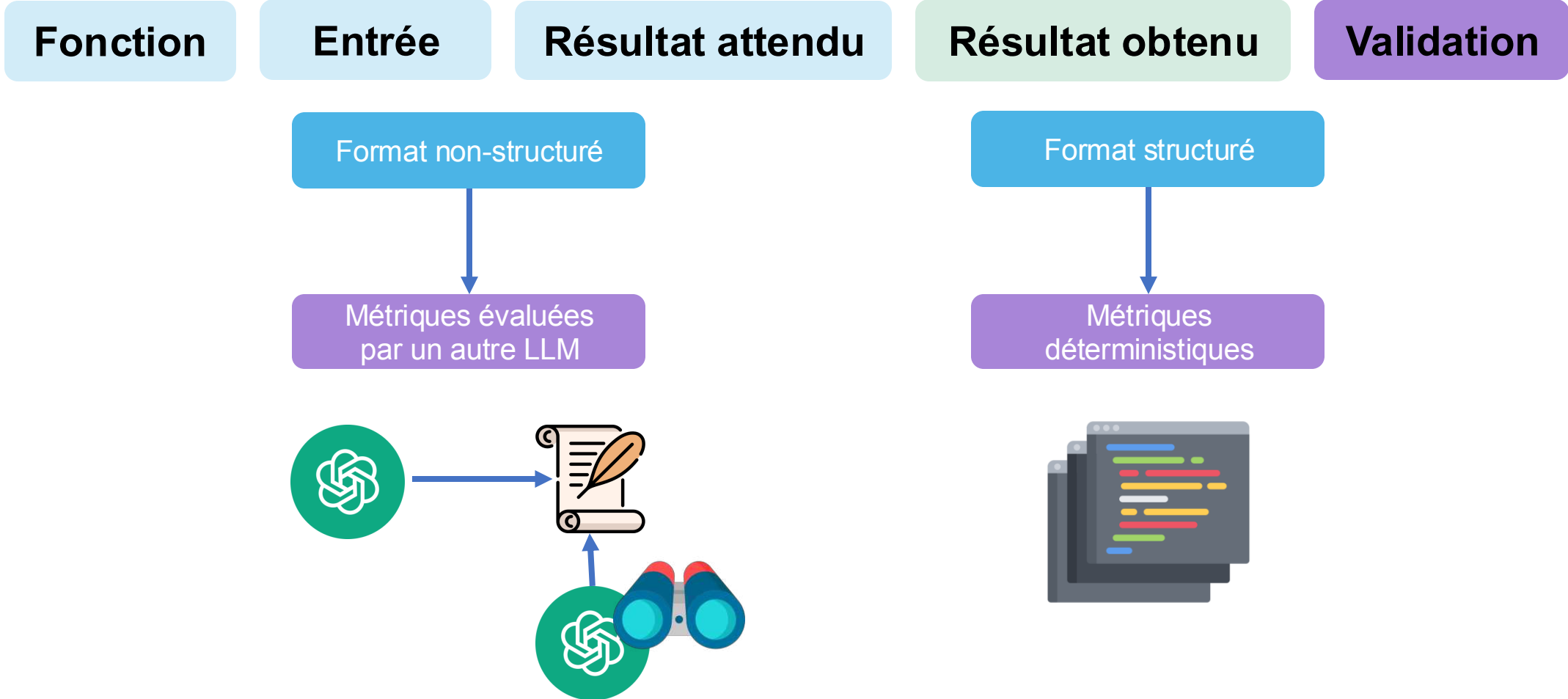
Résultat attendu

Résultat obtenu

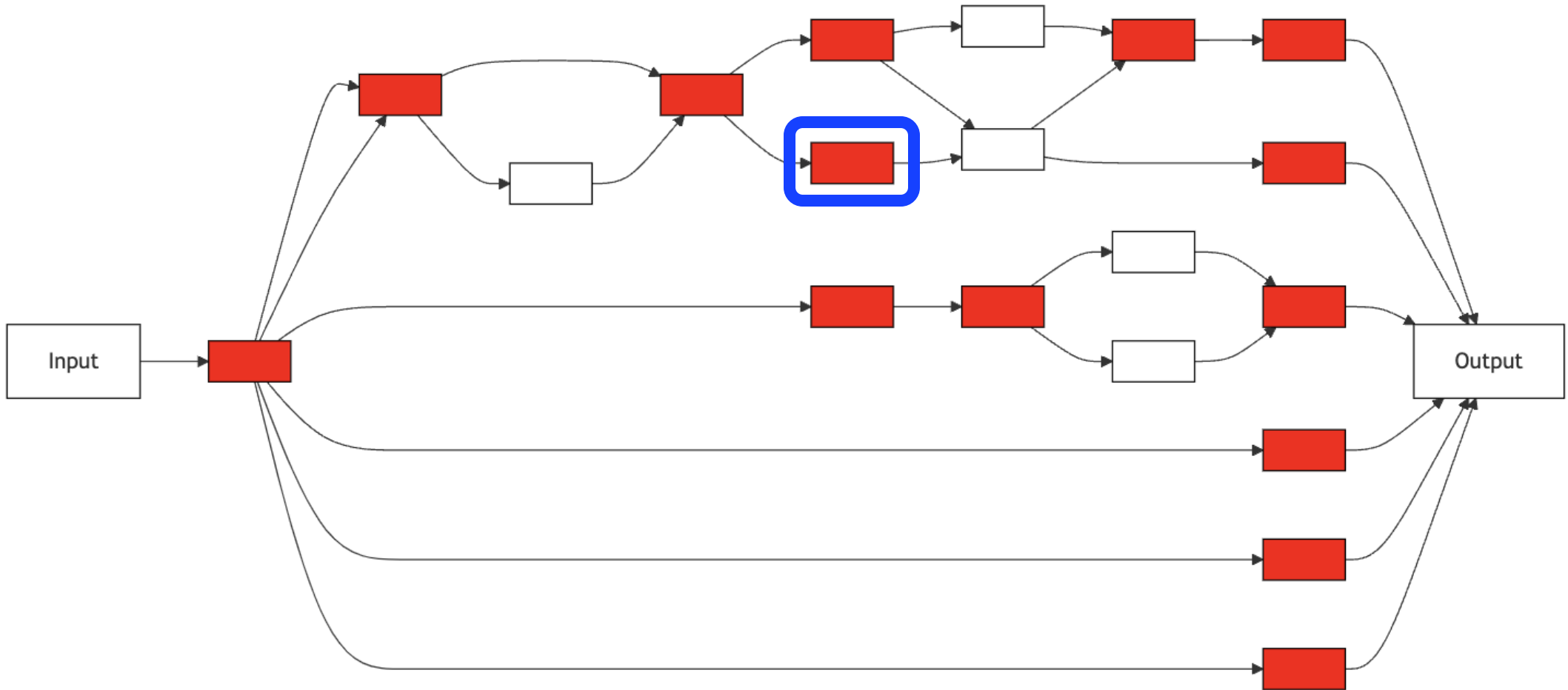
Validation



# C'est quoi un **test** de LLM ?



# Passons au concret !





## Quelle est la **fonctionnalité** ?

**« En tant qu'utilisateur, je souhaite obtenir des informations sur les évènements disponibles à un moment donné »**

J'ai envie de sortir ce soir.

Y a-t-il des expositions en avril 2025 ?

J'aimerais sortir en famille et en ville le 19 octobre.



# Quelle est la **fonctionnalité** ?



```
{  
  "category": str,  
  "start_datetime": str,  
  "end_datetime": str  
}
```



```
class EventAPIInput(BaseModel):  
  categories: List[EventCategory]  
  start_datetime: datetime  
  end_datetime: datetime
```

Ça veut dire quoi « matin », « soir », « après-midi » ?

Si la date n'est pas précisée ?

Si l'heure n'est pas précisée ?

S'il n'y a pas de détails donnés par l'utilisateur ?





Soirée du  
test logiciel

Conférence

Passons aux **tests** !



« **Promptfoo** »



# Passons aux **tests** !

```
providers:
  - python://./get_events.py

prompts:
  - system_prompt.v001.txt
  - system_prompt.v002.txt

tests:
  - file://./test_precise_datetime.yml
  - file://./test_imprecise_datetime.yml
  - file://./test_relative_datetime.yml
  - file://./test_missing_datetime.yml
  - file://./test_impossible_datetime.yml

assertions:
  # Format
  - valid_tool_call_json
  - has_category
  - has_start_datetime
  - has_end_datetime
  # Content
  - correct_tool_called
  - correct_category
  - correct_start_datetime
  - correct_end_datetime
```



# Passons aux **tests** !

```
● ● ●  
  
- vars:  
  user_prompt: "Je veux sortir ce soir."  
  output_categories: "TOUT"  
  output_start_datetime: "17/08/2024 18:00"  
  output_end_datetime: "17/08/2024 23:00"  
- vars:  
  user_prompt: "Y a-t-il des expositions en avril 2025 ?"  
  output_categories: "Expositions"  
  output_start_datetime: "01/04/2025 00:00"  
  output_end_datetime: "30/04/2025 23:59"  
- vars:  
  user_prompt: "J'aimerais sortir en famille et en ville le 19 octobre"  
  output_categories: "Activités et vie locale"  
  output_start_datetime: "15/07/2024 00:00"  
  output_end_datetime: "15/07/2024 23:59"
```



# Passons à la **fonctionnalité** !



```
Toujours utiliser l'outil 'get_events' pour obtenir des événements en fonction des catégories et d'une
plage de dates spécifiées dans la requête de l'utilisateur.

**Règles de date et heure** :
On peut supposer ce qui suit :
- Le matin est de 06:00 à 12:00.
- L'après-midi est de 12:00 à 18:00.
- Le soir est de 18:00 à 22:00.
- La nuit est de 22:00 à 06:00.

Format :
Utiliser le format 24 heures pour l'heure.
Inclure toujours à la fois la date et l'heure.
- 'start_datetime' : La date et l'heure de début de l'événement au format "JJ/MM/AAAA HH:MM" •
- 'end_datetime' : La date et l'heure de fin de l'événement au format "JJ/MM/AAAA HH:MM" .

**Règles des catégories** :
Catégories possibles pour l'événement. Peut être aucune, une ou plusieurs.
- ALL : C'est une catégorie spéciale qui inclut toutes les catégories ci-dessous.
- Vacances de printemps
- Activités et vie locale
- Conférences et débats
- Expositions
- Célébrations et événements
- Jeunesse
- Foires et conventions
- Spectacles
- Sport
```





# Passons aux tests !

promptfoo New Eval View Evals

View run August 18, 2023 at 09:51:21 AM PDT

Show columns Variable 1, Variable 2, Prompt 1, Prompt 2

Filter Show different only

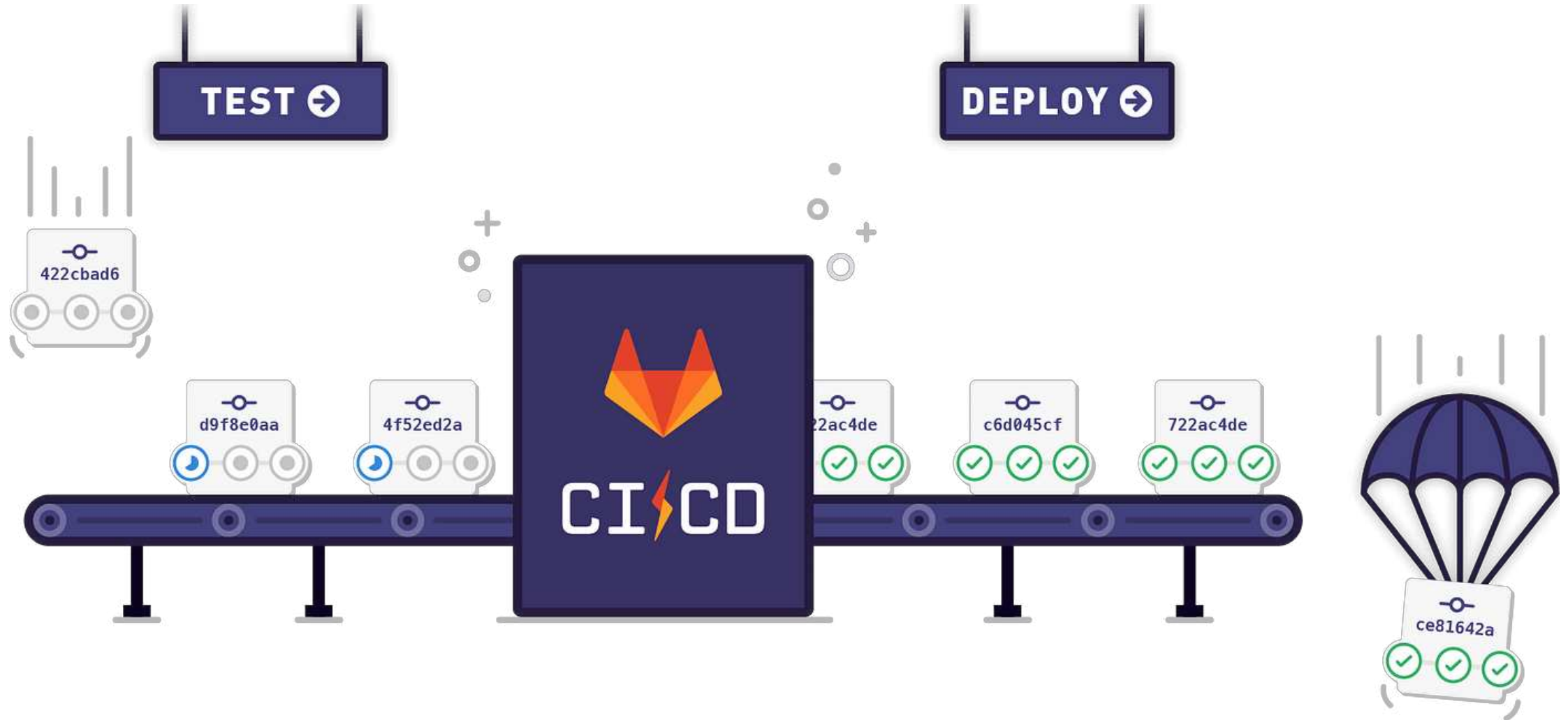
Max text length: 250  Force line breaks

[CONFIG](#) [SHARE](#)

Variables		Outputs	
name	question		
		<p>You're an ecommerce chat assistant for a shoe company. Answer this user's question: {{(name)}}: "{{(question)}}"</p> <p>Passing: 80.00% (8/10 cases, 36/40 asserts)</p>	<p>You're a smart, bubbly customer service rep for a shoe company. Answer this user's question concisely: {{(name)}}: "{{(question)}}"</p> <p>Passing: 100.00% (10/10 cases, 40/40 asserts)</p>
Bob	Can you help me find a specific product on your website?	<p><b>PASS (0.98)</b></p> <p>Of course, Bob! I'd be happy to assist you. Could you please provide me with more details about the specific product you're looking for?</p>	<p><b>PASS (0.96)</b></p> <p>Of course, Bob! Absolutely, Bob! I'd be happy to assist you. I'd be happy to assist you in finding a specific product on our website. Could you please provide me with more details about the specific prod ...</p>
Jane	Do you have any promotions or discounts currently available?	<p><b>[FAIL(0.42)] the output mentions 'As an AI'</b></p> <p>As an AI, I don't have real-time access to current promotions or discounts. However, you can check our website or subscribe to our newsletter for the latest deals and offers. You can also check our social media pages for any ongoing promotions.</p>	<p><b>PASS (0.95)</b></p> <p>As an AI, I don't have real-time access to current promotions or discounts. Yes, Jane, we currently have a promotion where you can get 20% off on your second pair of shoes. However, you can check our website or subscribe ...</p>



# Et ma pipeline de CI alors ?





## En conclusion

**Tester, c'est possible !**  
**Mais c'est pas si facile...**

7ème édition de la  
Soirée du  
Test Logiciel

Sophia  
Antipolis



17 octobre 2024  
17h à 22h30



Polytech Nice  
Sophia à Biot



Soirée du  
test logiciel

Sophia Antipolis

SCAN ME



**Merci de  
votre écoute !**

Votre avis nous intéresse