

# Soirée Meetup

**Generative AI par la pratique :**  
cas concrets d'utilisation d'un LLM en  
Java, avec l'API PaLM, Gemini et  
LangChain4J



**Guillaume LAFORGE**  
@Google



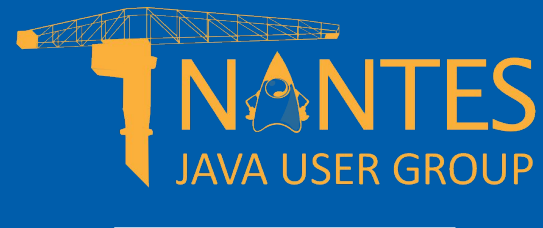
**15 février**



**19h**



**Epitech**



## Sponsors Gold



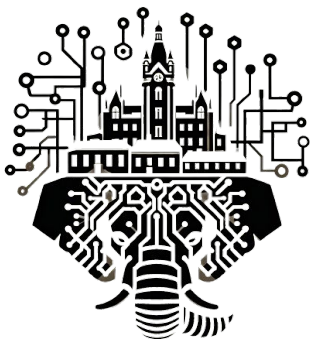
## Sponsors Silver





**Merci à nos  
sponsors <3**





GenAI Nantes

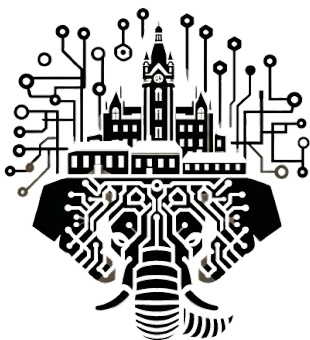


*Scan me*

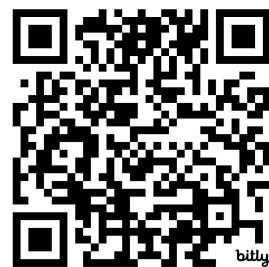


[sf≡ir]

lonestone



**SAVE THE DATE !**

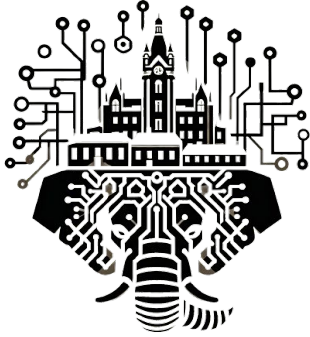


*Scan me*

---

**13/03 - Meetup avec**





**SAVE THE DATE !**



*Scan me*

---

**Hackathon GenAI** du vendredi 31 Mai au dimanche 2 Juin



**GDG** Nantes



**SAVE THE DATE !**

---

**14/03** - International Women Day



**SAVE THE DATE !**

---

**27/02 - Laissez tomber vos Dockerfile, adoptez un buildpack !**



# Schedule



1- News (20min)



2- Talk (1h30)



3- Enjoy ;)

-> 5 rue d'Alger

# Native LLM function calling

```
{  
  type: "function",  
  function: {  
    name: "get_current_weather",  
    description: "Get the current weather in a given location",  
    parameters: {  
      type: "object",  
      properties: {  
        location: {  
          type: "string",  
          description: "The city and state, e.g. San Francisco, CA",  
        },  
        unit: { type: "string", enum: ["celsius", "fahrenheit"] },  
      },  
      required: ["location"],  
    },  
  },  
},  
},
```

OpenAI:  gpt-3.5 

gpt-4

Llama2: 

Mistral: 

Gemini: 



PaLM: 

Flan-T5: 

Claude: 

Cohere: 

Work around:

-  prompting
-  fine tuning

# Embedding models stability

## Google:

- sentence-t5 (August '21)
- PaLM: textembedding-gecko-001 (May '23)
- Gemini: embedding-001 (December '23)

## OpenAI:

- text-embedding-ada-002 (December '22)
- text-embedding-3-large (January '24)

## Cohere:

- cohere-embed-multilingual-v2 (December '22)
- cohere-embed-multilingual-v3 (November '23)

**Anthropic:** not supported 🤖

## Other:

- sentence-transformers/all-MiniLM-L6-v2 (August '21)

# Embedding models stability

Benchmark: <https://huggingface.co/spaces/mteb/leaderboard>

## Conclusion:

- performance gain is marginal
- backfilling is costly
- new features in latest models: multimodal embeddings or moderation

# “Générer, c’est trahir”

```
# OpenAI's web crawler: GPT3.5, GPT4, ChatGPT  
User-agent: GPTBot
```


```
# ChatGPT plugins  
User-agent: ChatGPT-User
```

```
# Google's web crawler: Bard, VertexAI, Gemini  
User-agent: Google-Extended
```

```
# Claude  
User-agent: anthropic-ai
```




```
# Common Crawl  
User-agent: CCBot
```

```
# Censorship area  
Disallow: /
```




- Block training
- Block LLM web browsing
- Block AI agent 

# “Générer, c’est trahir”




## Category: Press

- Scanned: 51
-  Passing: 24 %
-  Blocked: 76 %
-  Unknown: 0 %

## Category: Podcast

- Scanned: 8
-  Passing: 75 %
-  Blocked: 25 %
-  Unknown: 0 %




## Category: Artist

- Scanned: 42
-  Passing: 74 %
-  Blocked: 21 %
-  Unknown: 0 %




( only)

# “Générer, c’est trahir”

## Category: X

- Scanned: 6
-  Passing: 100 %
-  Blocked: 0 %
-  Unknown: 0 %

## Category: Religion

- Scanned: 5
-  Passing: 100 %
-  Blocked: 0 %
-  Unknown: 0 %

## “Générer, c’est trahir”

Name	Status
Adobe Stock	✓
Getty Images	✓
Pexels	✓
Pixabay	✓
500px	✓



# “Générer, c’est trahir”

## Shame list

---

Name	Status
Medium	
Quora	
Elsevier	
ScienceDirect	
Scopus	
Science	

**“Générer, c’est trahir”**

<https://github.com/samber/the-great-gpt-firewall>

## Contenu authentique ?



# Contenu authentique ?

## Getty Completes its Reversal on Generative AI and Launches Generative AI by iStock

You can pay to generate images, but your rights of use are limited according to Getty



BRET KINSELLA  
9 JAN 2024



12



1

Share



## Duolingo turns to AI to generate content, cuts 10 percent of its contractors

The language learning app is increasingly relying on AI to create its lessons.

By [Amanda Yeo](#) on January 9, 2024



Credit: Nikolas Kokovlis / NurPhoto via Getty Images

# Contenu authentique ?

## Amazon confronté à une déferlante de « faux livres » générés par intelligence artificielle

Le géant américain a réduit à trois le nombre de livres publiés chaque jour sur sa plate-forme Kindle Direct Publishing par un même « auteur », en obligeant à spécifier s'il a été écrit par IA. Sans donner de détails sur ses éventuelles sanctions.

Par Nicole Vulser

Publié le 12 décembre 2023 à 18h33, modifié le 13 décembre 2023 à 05h38 · 🕒 Lecture 4 min.

🔒 Article réservé aux abonnés



AUREL

# OpenAI memory





**GPT-3.5**

**Turbo**

Input : -50%

Output : -25%

**GPT-4 Turbo**

Moins

paresseux

A medium shot of Sam Altman, co-founder of OpenAI, speaking at a podium. He is wearing a dark suit jacket over a light blue button-down shirt. He has dark, curly hair and is looking slightly upwards and to his right. The background is a solid blue wall with a white geometric structure on the left. A red news banner is at the bottom of the frame.

**BREAKING  
NEWS**

**ALTMAN SEEKING BILLIONS FOR AI CHIP FACTORIES**



SDXL



DALL-E 3



RPG (Ours)



**Prompt:** A beautiful landscape with a river in the middle, the left of the river is **in the evening** and **in the winter** with a big iceberg and a small village while some people are skiing on the river and some people are skating, the right of the river is **in the summer** with a **volcano in the morning** and a small village while some people are playing.









Original video  
(not used as model input)



Segmentation



Condition



Output



Original video  
(not used as model input)



Segmentation



Condition



Output

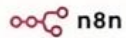


Source :  
<https://actanywhere.github.io/>



# TOPAZ AI





n8n

Frame.io Tasks

+ Add tag

Active



Share

Saved



Workflows



Templates



Credentials



Variables



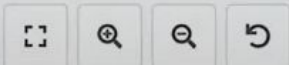
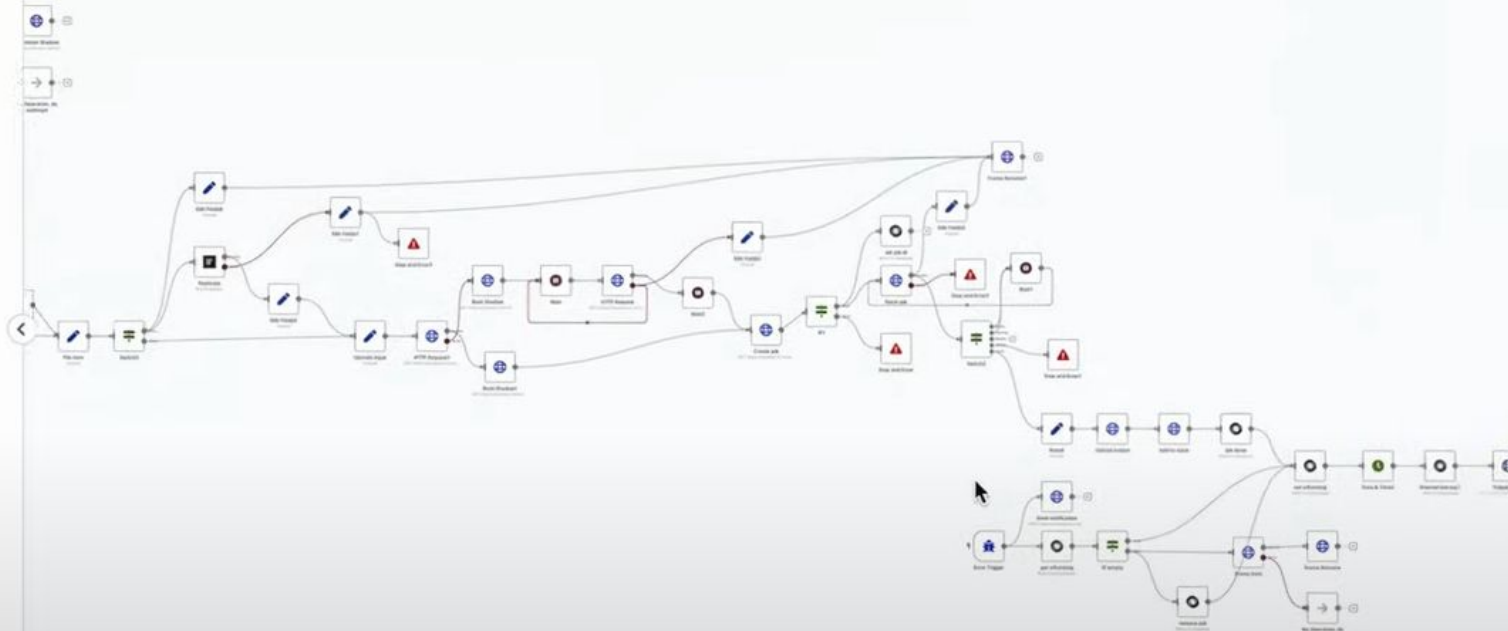
All executions



Settings



Help



Execute Workflow



Chat GPT  
Dall-e



Replicate



Source : <https://www.youtube.com/watch?v=OjdlqxivZVYw>







## AI Startup GPU Outreach Tracker



ai\_startup\_leads

product\_comparison

dealership\_due\_diligence

property\_diligence

ai\_companies

gpu\_sales\_research

+ v



Default View



Columns



Filter



Sort



Download as CSV



	company name	potential_future_needs	recent_startup_news	key_executives	linkedin_url	+
1	DeepGenomics					
2	Voyage Auto					
3	Nuritas					
4	Cerebras Systems					
5	PathAI					
6	Grammarly AI					
7	Drive.ai					
8	CloudMinds					
9	DataRobot					
10	Verkada					
11	Vicarious					
12	Pony.ai					
13	Brainwave Technologies					
14	Alpha Analytica					
15	Synthetic Data					
16	Future Intelligence Ltd					
17	RoboThoughts Inc					



usaco\_medalists

ioi\_medalists

ipho\_medalists

imo\_medalists

+ ▾

Default View

Columns

Filter

Sort

Download as CSV



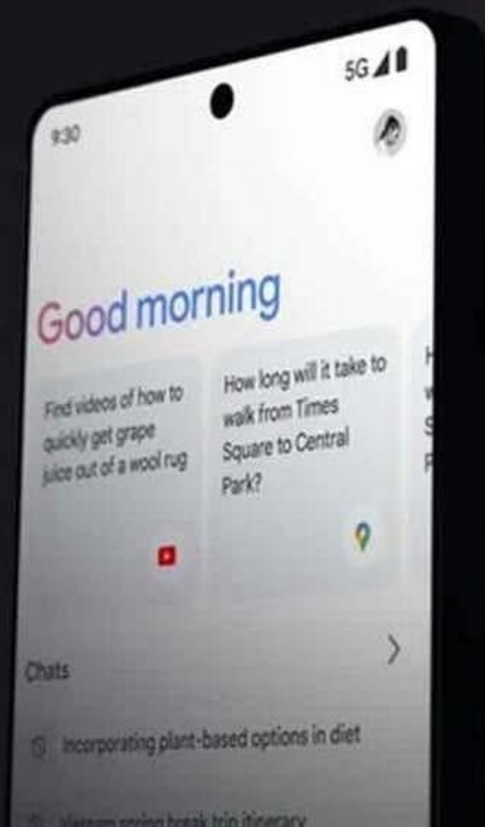
	Name	Best Medal	Years Awarded	Country	LinkedIn URL	Current Job Title	+
1	Aleksei Volostnov						
2	William Hu						
3	Yixiao Li						
4	David Stoner						
5	Fabio Pruneri						
6	James Lin						
7	Ryoshun Oba						
8	Mikhail Ivanov						
9	Yixuan Wang						
10	Sehun Kim						
11	Jongwon Lee						
12	Po-Jui Chen						
13	Ashwin Sah						
14	Allen Liu						
15	Victor Wang						
16	Jiyang Gao						



	product_name	description	features	price	number of users	founded	+
1	Jira						
2	Notion						
3	Airtable						
4	Smartsheet						
5	Monday						
6	Wrike						
7	Asana						
8	Linear						
9	Google Sheets						
10	Trello						
11	Microsoft Excel						
12	ClickUP						
13	BaseCamp						
14	Teamwork						
15	Zoho Projects						
16	Podio						
17							

[] Matrices

# Gemini



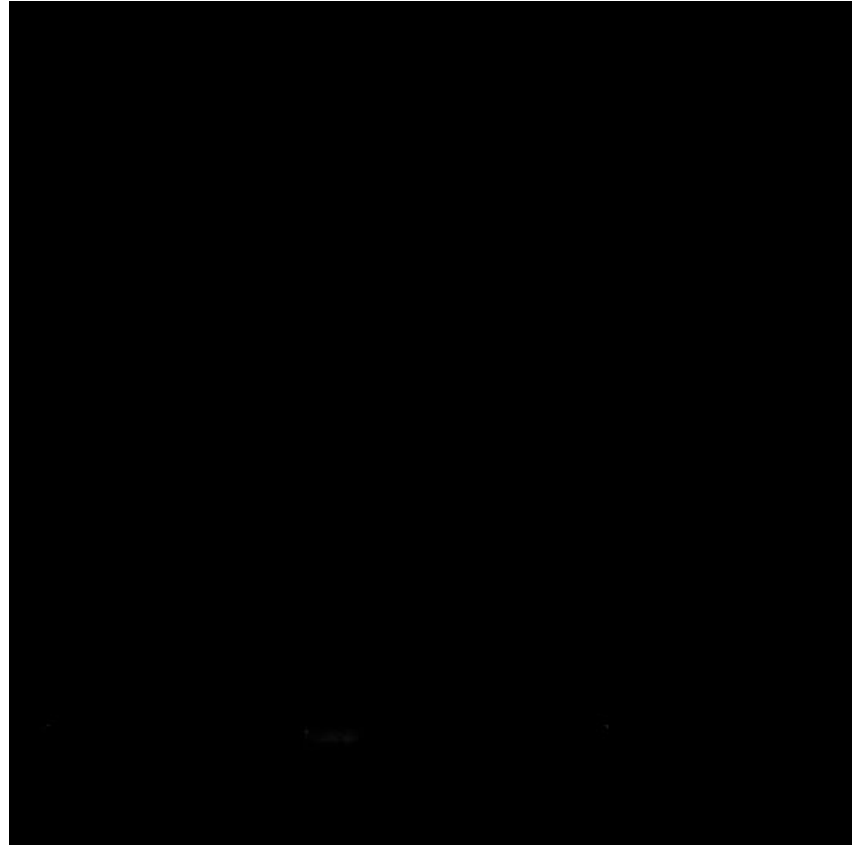
## Gemini Pro 1.5

1M de tokens (films, livres en entiers)

Mixture of Experts (cf Mixtral 8x7B)

Source :

<https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/#sundar-note>



Capability	Benchmark	Gemini		
		1.0 Pro	1.0 Ultra	1.5 Pro
Math, Science & Reasoning	<b>Hellaswag</b> (Zellers et al., 2019)	84.7% 10-shot	87.8% 10-shot	<b>92.5%</b> <b>10-shot</b>
	<b>MMLU</b> : Multiple-choice questions in 57 subjects (professional & academic). (Hendrycks et al., 2021a)	71.8% 5-shot	<b>83.7%</b> <b>5-shot</b>	81.9% 5-shot
	<b>GSM8K</b> : Grade-school math problems. (Cobbe et al., 2021)	77.9% 11-shot	88.9% 11-shot	<b>91.7%</b> <b>11-shot</b>
	<b>MATH</b> : Math problems ranging across 5 levels of difficulty and 7 sub-disciplines. (Hendrycks et al., 2021b)	32.6% 4-shot Minerva prompt	53.2% 4-shot Minerva prompt	58.5% 4-shot Minerva prompt <b>59.4%</b> <b>7-shot</b>
	<b>AMC 2022-23</b> : 250 latest problems including 100 AMC 12, 100 AMC 10, and 50 AMC 8 problems.	22.8% 4-shot	30% 4-shot	<b>37.2%</b> <b>4-shot</b>
	<b>BigBench - Hard</b> : A subset of harder tasks from Big Bench formatted as CoT problems. (Srivastava et al., 2022)	75.0% 3-shot	83.6% 3-shot	<b>84.0%</b> <b>3-shot</b>
	<b>DROP</b> : Reading comprehension & arithmetic. (Metric: F1-Score). (Dua et al., 2019)	74.1% Variable shots	<b>82.4%</b> <b>Variable shots</b>	78.9% Variable shots
Coding	<b>HumanEval</b> chat preamble* (Metric: pass rate). (Chen et al., 2021)	67.7% 0-shot (PT)	<b>74.4%</b> <b>0-shot</b> <b>(PT)</b>	71.9% 0-shot
	<b>Natural2Code</b> chat preamble* (Metric: pass rate).	69.6% 0-shot	74.9% 0-shot	<b>77.7%</b> <b>0-shot</b>
Multilinguality	<b>WMT23</b> : sentence-level machine translation (Metric: BLEURT). (Tom et al., 2023)	71.73 (PT) 1-shot	74.41 (PT) 1-shot	<b>75.20</b> <b>1-shot</b>
	<b>MGSM</b> : multilingual math reasoning. (Shi et al., 2023b)	63.45% 8-shot (PT)	78.95% 8-shot (PT)	<b>88.73%</b> <b>8-shot</b>





 Generative AI experiment. Info quality may vary.

### Ask about

- ↳ Kid friendly activities for a rainy day
- ↳ Best brunch spots nearby
- ↳ Things to do for outdoor enthusiasts near me
- ↳ Fun things to do with friends nearby

Ask a question. The more detail, the better the ideas.





# Soirée Meetup

**Generative AI par la pratique :**  
cas concrets d'utilisation d'un LLM en  
Java, avec l'API PaLM, Gemini et  
LangChain4J



**Guillaume LAFORGE**  
@Google



**15 février**



**19h**



**Epitech**