

# 1001 RAISONS D'ÉCHOUER



Créer un compte

## LEMAGIT

### IA générative : les entreprises face au défi de l'industrialisation

Les PoC en IA générative se sont par entreprises. Mais per an complexités. État des =

**THENEWSTACK** 

a

**OPERATIONS** 

From POC to Production: Why GenAl Projects Often Stall Successfully transitioning GenAl projects from POC to production requires a strategic approach grounded in effective

governance.

Gartner

Newsroom

Newsroom / Information Technology / Press

par Christophe Auffray, LeN

Apr 11th, 2025 10:00am by Arti Raman **Gartner Predicts 30% of Generative** Al Projects Will Be Abandoned After Proof of Concept By End of 2025

SYDNEY, Australia, July 29, 2024

L'IA en Entreprise : entre prom infinies et défis à relever pour

Consulting Manager | Digital Transformation

Les usages de la genAI s'étendent,

mais au prix d'un grand nombre

Mojuste EGBEWOLE

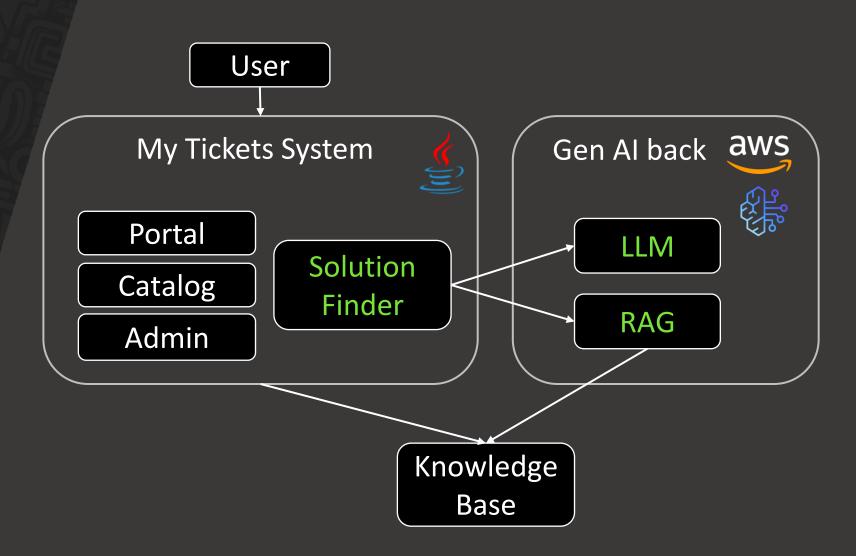
d'échecs

le 27 Mars 2025

28 janvier 2025

### SOLUTION FINDER

Super POC parti en prod trop tôt



## **DEVOXX NEWS**

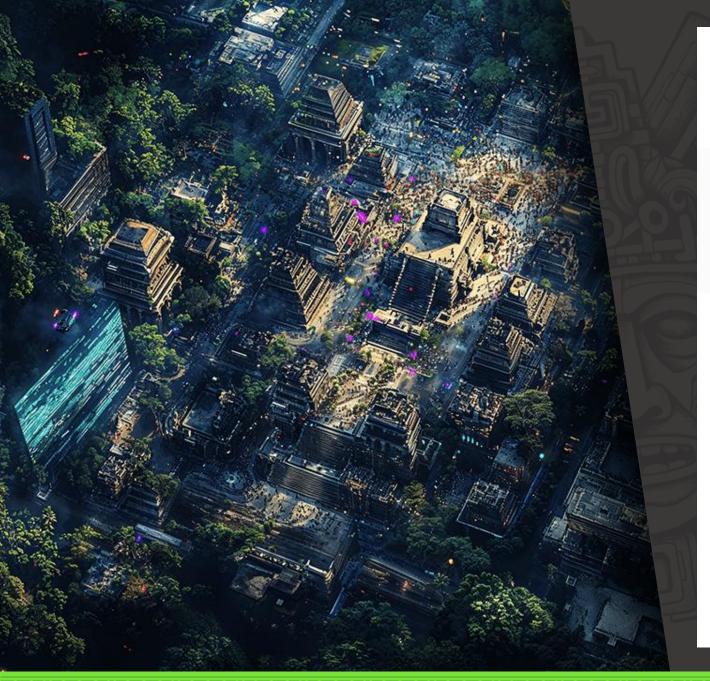
# Guerre commerciale : La France interdit l'usage des outils IA américains



Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo egestas. Praesent vitae faucibus nisi.

Interdum et malesuada fames ac ante ipsum primis in faucibus. In aliquam mauris sed risus fermentum porta. Curabitur quis dolor quis risus accumsan dictum. Nullam efficitur tellus eu est congue maximus. Donec vitae urna sodales enim venenatis vestibulum. Donec gravida varius lectus id accumsan. In tortor orci, molestie quis lorem ut, euismod molestie quam.

Sed nec mauris nec urna





# ARCHITECTURE PRINCIPES

#### Bonnes pratiques à conserver



- Well Architected Framework
- SOLID
- Clean Architecture

# ARCHITECTURE PRINCIPES

#### Bonnes pratiques à conserver



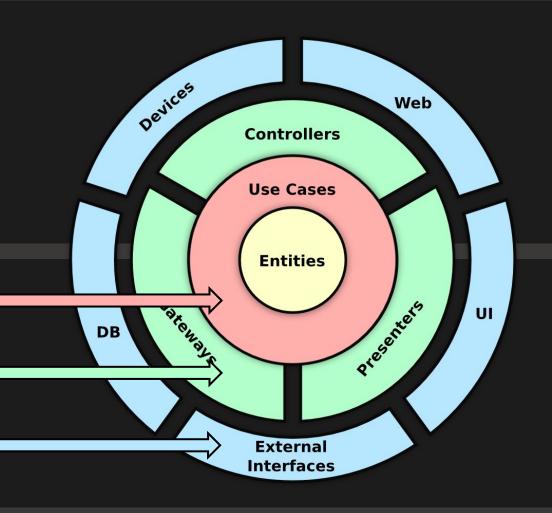
- Well Architected Framework
- SOLID
- Clean Architecture



**Solution Finder feature** 

**Gen AI controller** 

**GenAl** backend interface

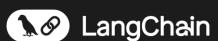


## ARCHITECTURE **FRAMEWORKS**

#### Frameworks



- SDKs des fournisseurs de modèles
- Frameworks spécialisés LLM / Agents
- Modules stacks principales Spring Al



LlamaIndex

Semantic Kernel 24k 🌟



40k

**G**Github





Langages















## ARCHITECTURE **FRAMEWORKS**

#### **Frameworks**



- SDKs des fournisseurs de modèles
- Frameworks spécialisés LLM / Agents
- Modules stacks principales Spring Al





40k

**G**Github





Langages







Semantic Kernel 24k

LlamaIndex

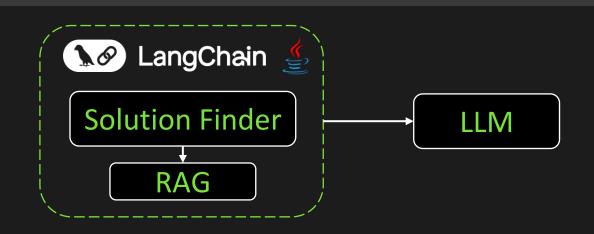












## **DEVOXX NEWS**

### Scandale après une fuite de données dans « Solution Finder »



ac scelerisque. Nam in dapibus. Sed sed diam

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam

Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo Praesent vitae faucibus

Interdum et malesuada fames ac ante ipsum primis in faucibus. In aliquam mauris sed risus fermentum Curabitur quis dolor quis risus accumsan dictum. Nullam efficitur tellus eu est conque maximus. Donec vitae urna sodales vestibulum. Donec gravida varius lectus id accumsan. In tortor orci, molestie quis lorem ut, euismod molestie quam.

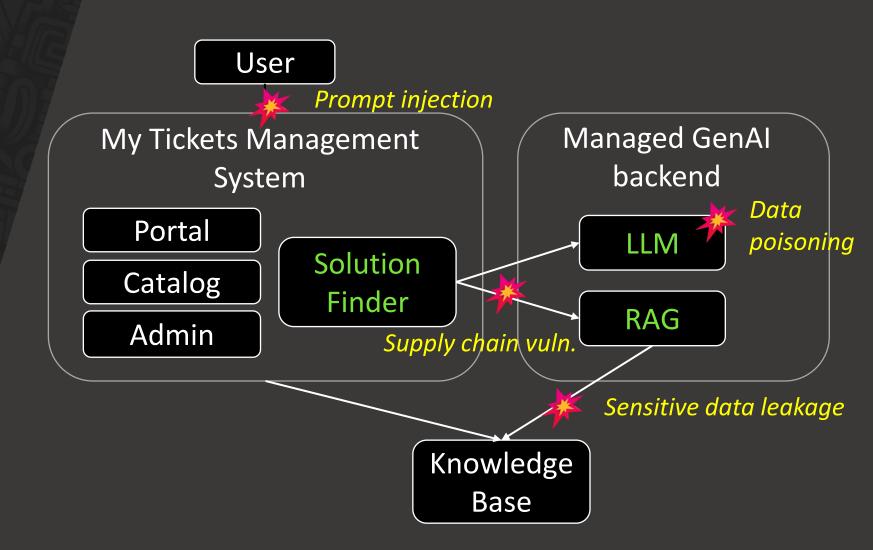
Sed nec mauris nec urna

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim urna sed orci gravida



## **SÉCURITÉ**

NOUVELLES SURFACES D'ATTAQUE



## **SÉCURITÉ**

### OWASP TOP 10 FOR LLM

**01**: Prompt injection

02 : Sensitive info. disclosure

03 : Supply chain

**04**: Data and model poisoning

**05**: Improper output handling

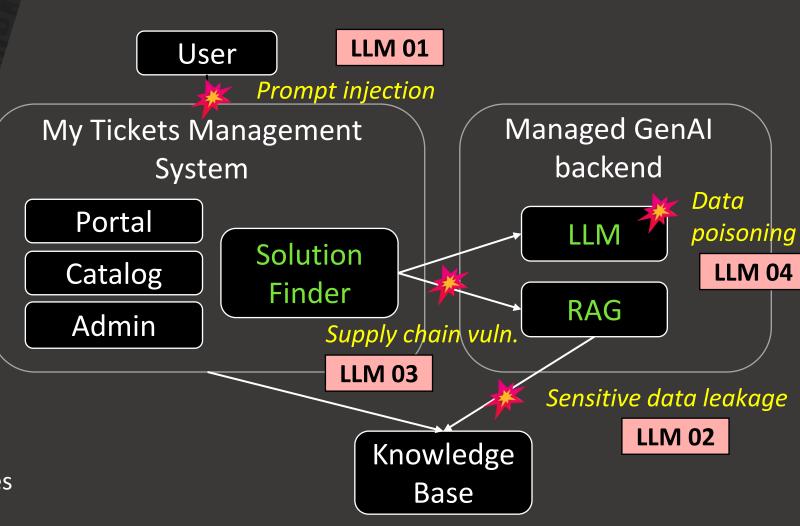
**06**: Excessive agency

**07**: System prompt leakage

**08**: Vector & embedding weaknesses

09: Misinformation

10: Unbounded consumption



## SÉCURITÉ TESTER

#### Grille d'audit



- **OWASP** Top 10 for LLM
- 35 recommandations de sécurité pour un système d'IA générative de l'ANSSI

### **Outils: approche Red Team**



PyRIT







DeepEval.

# SÉCURITÉ **TESTER**

#### Grille d'audit



- **OWASP** Top 10 for LLM
- 35 recommandations de sécurité pour un système d'IA générative de l'ANSSI



#### **Outils: approche Red Team**



PyRIT





promptfoo



DeepEval.

#### PROTÉGER #GUARDRAILS

#### Services managés



- **Azure** Al Content Safety
- **Amazon** Bedrock Guardrail
- **Google Cloud** Vertex Al Safety Filters

#### Librairies / Frameworks / APIs

- Solutions intégrées aux frameworks
- **Nvidia** NeMo Guardrails
- **OpenAl** Moderation API

## **DEVOXX NEWS**

### J-Corp chute en bourse après une indisponibilité de 12h de « My Tickets »



Phasellus cursus et enim urna sed orci gravida dapibus. Sed sed diam

Lorem ipsum dolor sit adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam

Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo Praesent vitae faucibus nisi.

Interdum et malesuada fames ac ante ipsum primis in faucibus. In aliquam mauris sed risus fermentum Curabitur quis dolor quis risus accumsan dictum. Nullam efficitur tellus eu est conque maximus. Donec vitae urna sodales vestibulum. Donec gravida varius lectus id accumsan. In tortor orci, molestie quis lorem ut, euismod

Sed nec mauris nec urna

molestie quam.

Lorem ipsum dolor sit amet, adipiscing elit. ac scelerisque. Nam in



## OPS

### AUANT LE DÉPLOIEMENT

#### Tests des outputs LLM



- Outputs non déterministes
- Outils dédiés ou extensions
- Human-in-the-loop





DeepEval.

- ✓ **Assertions** et **métriques** : contains, schémas réponses, ...
- ✓ **LLM as a judge** : cohérence factuelle, pertinence, fidélité, ...

## OPS

### AUANT LE DÉPLOIEMENT

#### **Tests des outputs LLM**



- Outputs non déterministes
- Outils dédiés ou extensions
- Human-in-the-loop





DeepEval.

- ✓ **Assertions** et **métriques** : contains, schémas réponses, ...
- ✓ **LLM as a judge** : cohérence factuelle, pertinence, fidélité, ...

#### Tests de performances



- Risques principaux : latence et limites
- Couvrir les Guardrails et Fallbacks
- Coûts et capacités -> mocks / bouchons







#### **Observabilité**

- Nouvelles métriques : Sessions, coûts, qualité, latence, erreurs, ...
- 3 types de solutions :

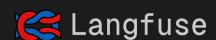


S

Elastic Stack

Grafana Stack

Datadog







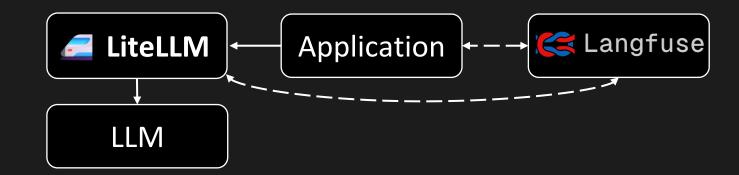
Spécialistes observabilité

Spécialistes observabilité IA

#### **Stack LLMOps**



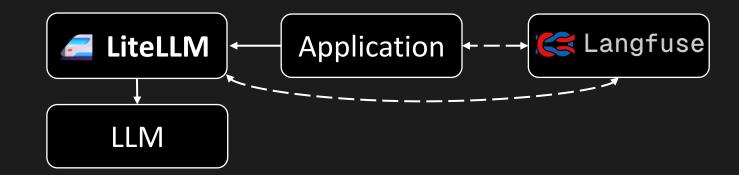
- Cloud ou **Open Source**
- LiteLLM: LLM Gateway
- Langfuse : Observabilité



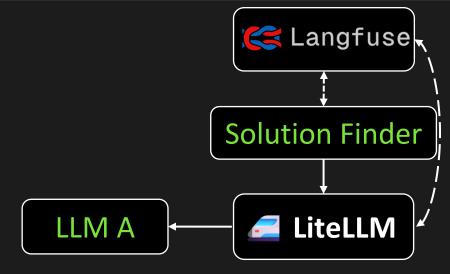
#### **Stack LLMOps**



- Cloud ou Open Source
- LiteLLM: LLM Gateway
- Langfuse : Observabilité



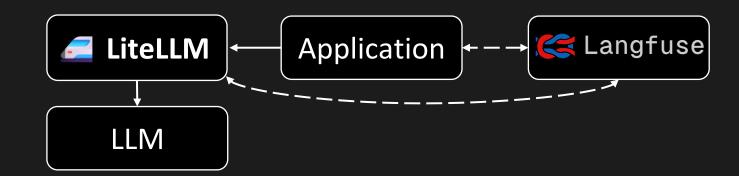




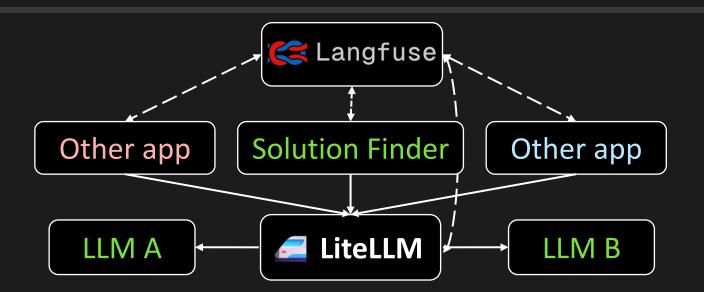
#### **Stack LLMOps**



- Cloud ou Open Source
- **LiteLLM**: LLM Gateway
- Langfuse : Observabilité







## 1001 AUTRES RAISONS D'ÉCHOUER

GOUVERNANCE

**FORMATION** 

THE RESERVE OF THE PARTY OF THE

CONFIANCE

ÉTHIQUE

CONFIDENTIALITÉ

USE CASES MÉTIER

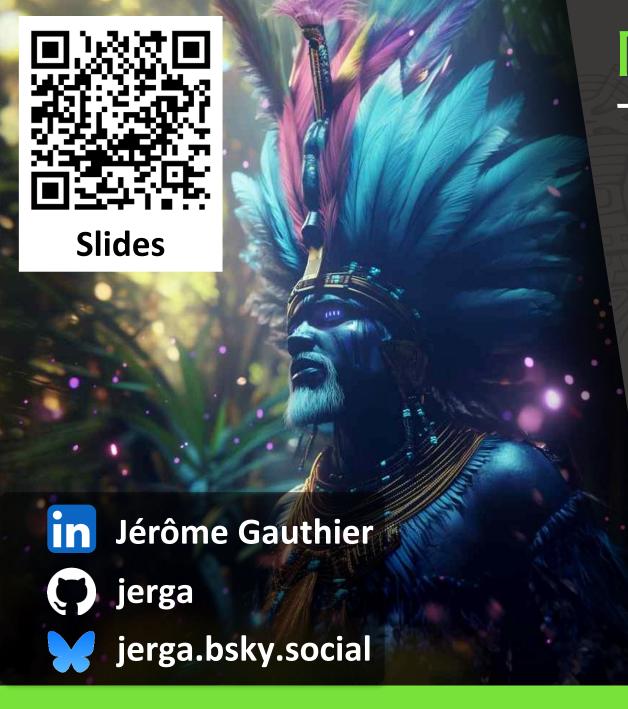
QUALITÉ DES DONNÉES

**ACCOMPAGNEMENT** 

RÈGLEMENTATIONS

SOUVERAINETÉ

DURABILITÉ



# ROAD TO PROD TAKE AWAY

