

Un RAG spécialisé pour l'IA générative au service de la Vérification & Validation des systèmes complexes.

Bruno Besace, Arnaud Bouzy, Frédéric Tamagnan
THALES & SMARTESTING



| SOMMAIRE

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

1

Le projet VVaMIA

2

Architecture technique

3

Démonstration

4

Améliorations futures

5

Conclusion

| Le projet VVaMIA

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER



| Le projet VVaMIA

Points de peine et bénéfices attendus

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

Points de peine du métier	Bénéfices de l'assistant IA
Effort de compréhension et d'analyse des spécifications	Réduire les reprises dues à la mécompréhension du besoin
Qualité des entrants	
Validité du périmètre à tester	
Mesure de la complétude	Connaître et améliorer la complétude des tests
Gestion du changement	Faciliter les analyses d'impact Réduction du temps de traitement d'un fait technique
Shift left	Intercepter plus tôt les défauts et accroître leur taux de détection

Architecture technique

5

| Construction autour de Yest

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

- Outil de Model-Based Testing
- 3 services principaux
 - Conception de test graphique et collaborative
 - “IDE du testeur”: accélérateur d’implémentation des tests
 - Support à l’automatisation par mots-clés
- Lancé en 2018 - 300 utilisateurs



THALES



Dior

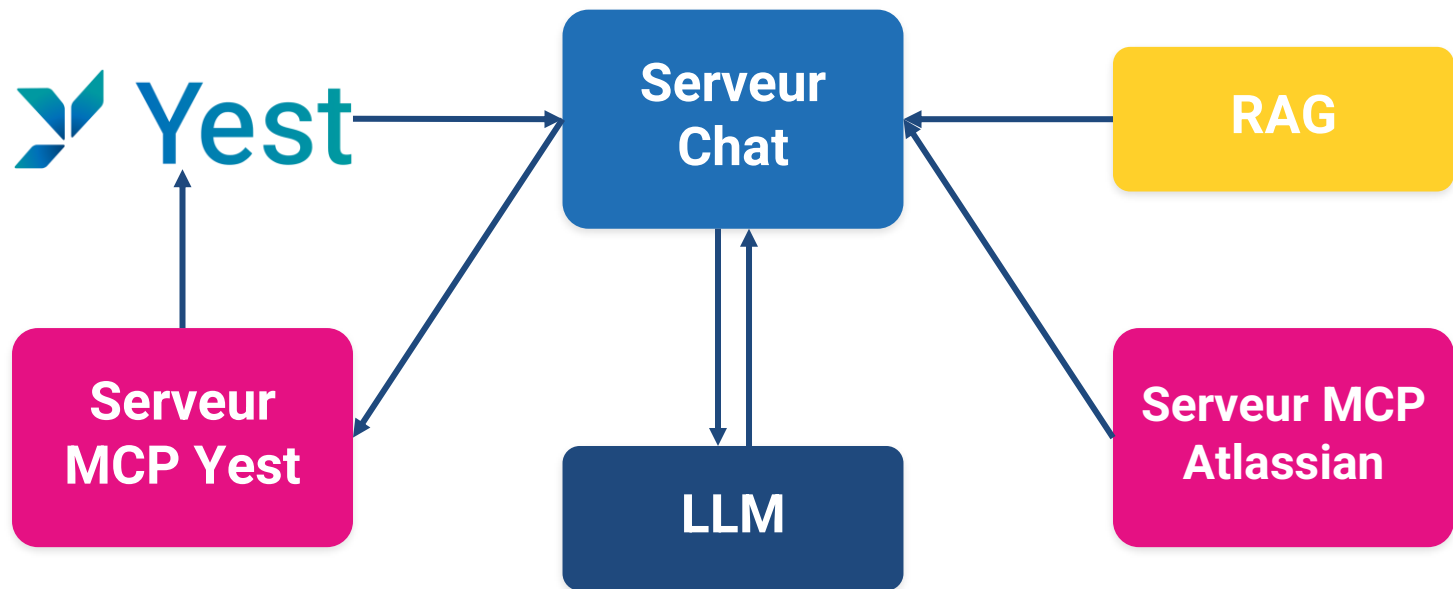


Lufthansa



| Vue globale de l'architecture

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

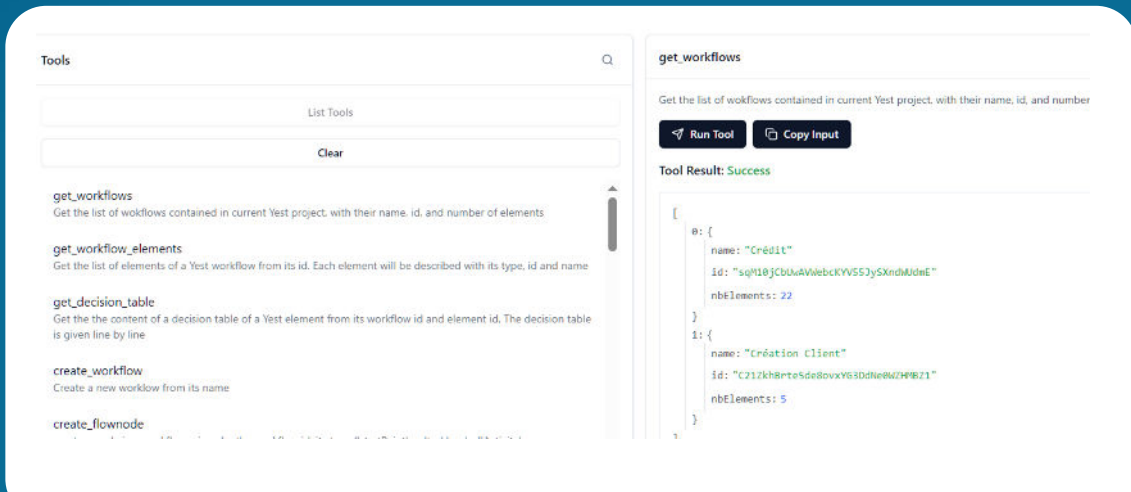
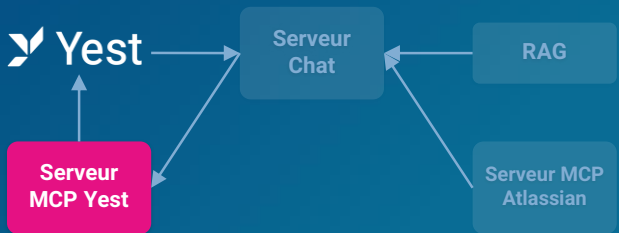


| Serveur MCP Yest

MCP : Model Context Protocol

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

- Un serveur MCP expose des tools offrant des services sur un domaine particulier.
- Les tools seront utilisés par un LLM. MCP est le protocole de communication de ce serveur.
- Le serveur MCP Yest propose des services pour consulter et modifier le projet Yest ouvert dans la session.



Démonstration

9

| Construction autour de Yest

- **Augmentation plutôt que délégation**
 - Maîtrise du choix des tests
 - Contrôle sur les hallucinations
 - Contrôle de la couverture des tests
 - Décomposition de la complexité

- **Communication/Collaboration**
 - Représentation visuelle
 - Lisibilité
 - Maintenabilité des tests déjà produits

Améliorations futures

11

| Améliorations futures

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

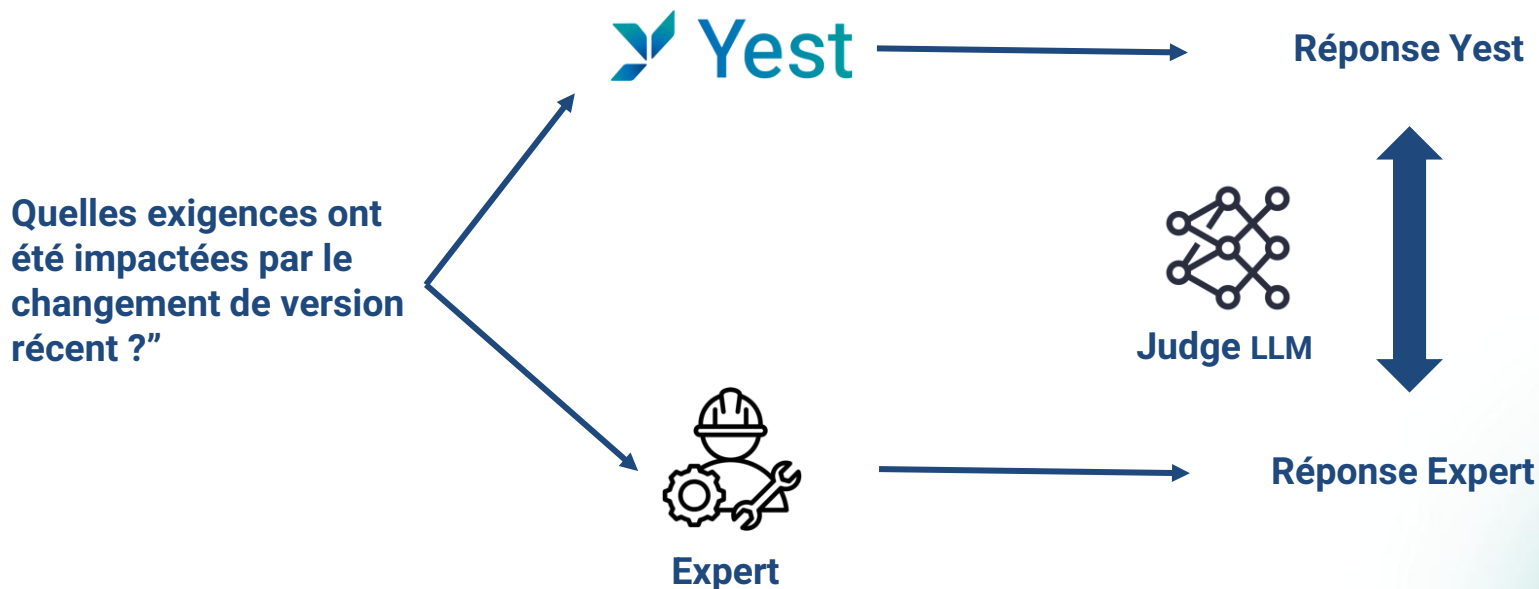
Intégration au système d'un RAG :



| Améliorations futures

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

Évaluation : jeu de vérité



| Améliorations futures

Évaluation : Jeu de vérité

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER



Question 3/13

Good: 1.0 pts

5.92s

Question:

Quel est le nombre total de capacités sur l'ensemble du projet ?

EXPECTED ANSWER

le nombre total de capacités dans le projet est 86

AGENT ANSWER

86

JUDGE EVALUATION (2.24s)

Good

1.0 / 1.0

The agent's answer '86' exactly matches the ground truth that the total number of capacities in the project is 86. It correctly and completely addresses the question without any errors or omissions.

► View Execution Traces

| Améliorations futures

LLM-as-a-Judge :

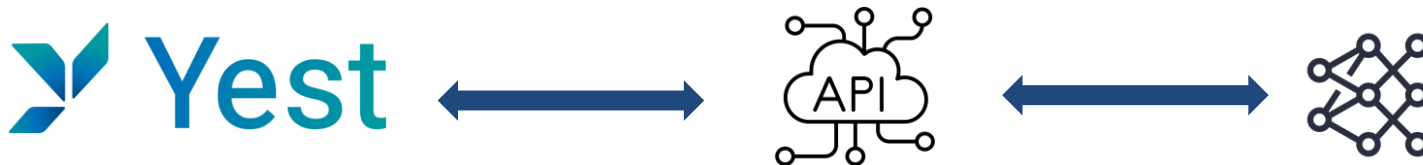
JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

Modèles	Score LLM as a Judge	Score ajusté	Open-source ?	# de paramètres
haiku-4-5	38%	88%	✗	Inconnu
devstral-small	38%	75%	✓	24B
ministral	27%	70%	✓	14B
mistral-small	42%	88%	✓	24B
gpt-oss	23%	33%	✓	20B
qwen3	35%	70%	✓	14B

| Améliorations Futures

JOURNÉE
THÉMATIQUE
IA GÉNÉRATIVE
POUR TESTER

Utilisation des tools :



- > +50 appels de tools à chaque requête
- > **Nécessite de l'optimisation**
 - Fine Tuning
 - Système agentique robuste

Conclusion

17

| Conclusion

Cas d'usage représentatif et cas d'étude

Étape 1



Cas d'usage représentatif

- **Source ouverte** « Pythadrone »
- **Représentatif de l'ontologie des artefacts d'ingénierie** (i.e. exigences, modèle Capella, modèle Simulink, corpus de test préexistant, description de l'incrément fonctionnel ciblé, variabilité de la politique produit etc.)
- **10%** du volume des artefacts d'ingénierie

KPI

Score vis-à-vis d'un **catalogue de requêtes** de référence (formant un « jeu de vérité »)

Étape 2



Cas d'étude d'un système aéroporté

- **Propriété intellectuelle Thales**
- **Produit déjà qualifié** (pour disposer d'une « vérité terrain » fiable)
- 600 exigences au niveau système + 5 000 exigences au niveau sous-système (SW et HW)
- 4 incréments fonctionnels

KPI

Score vis-à-vis du **même catalogue**

| Conclusion

Avancement du projet

- **À 40% du plan d'avancement :**
 - Spécifications et architecture déterminées
 - Début de la phase de développement et d'intégration
- **Nous recherchons des partenaires** pour exercer la solution sur de multiples environnements et types d'applications.

Des volontaires ?

Merci !



THALES

UNIVERSITÉ
MARIE & LOUIS
PASTEUR