

# JFIE 2025

**Génération et Vérification  
Automatique de User Stories avec  
l'IA Générative**

**Présentation, Démonstration & Discussion**

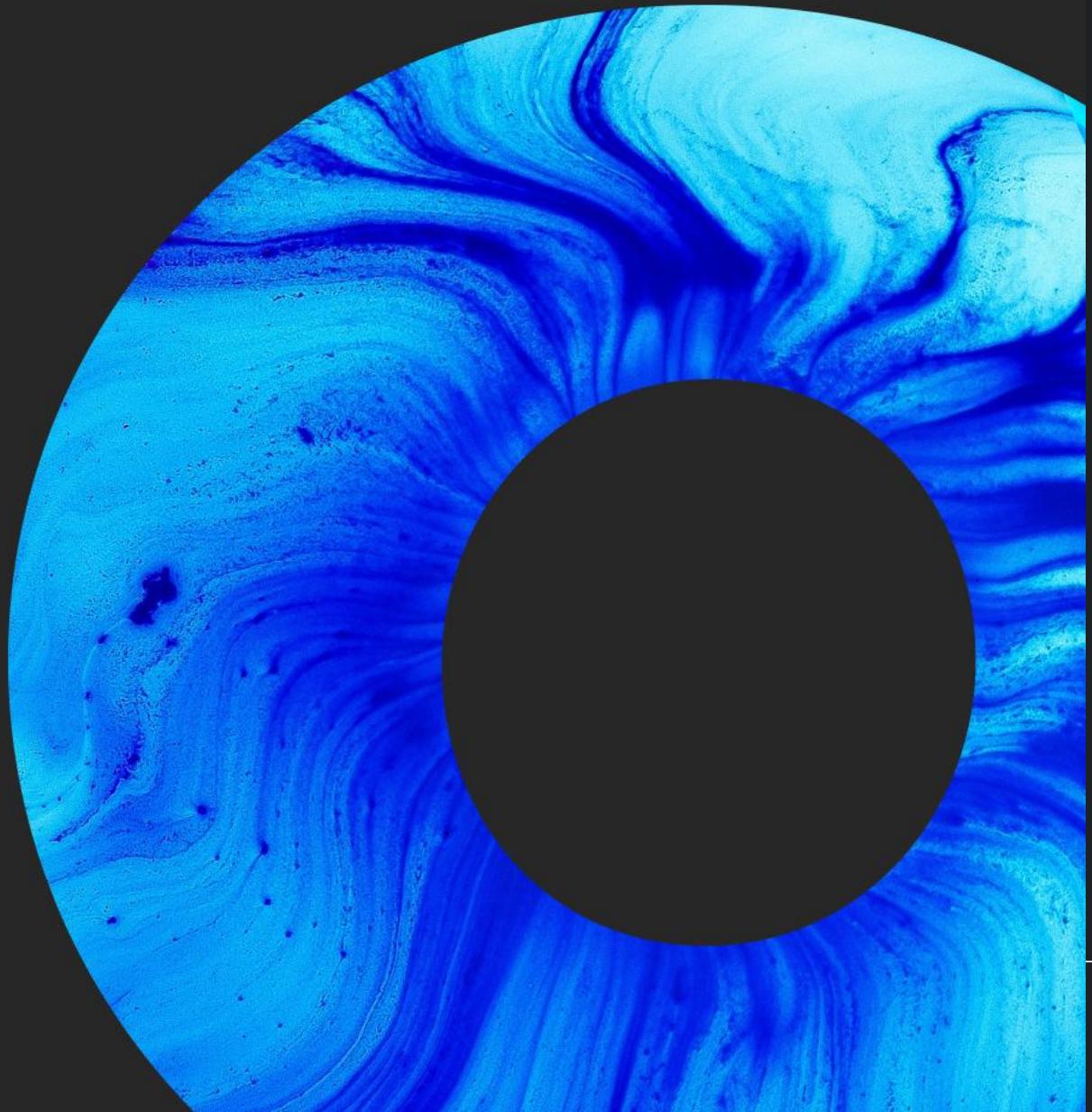
Présenté par :



**Marius Ortega**

Doctorant

IA & Test Logiciel



# Sommaire

## 1. Contexte & Problématique (8 min)

- Définition d'une User Story
- Difficulté d'Accès aux Données
- Autres points de tension liés aux User Stories

## 2. Génération Automatique de User Stories (12 min)

- Les Typologies de Tâches entre IA et User Stories
- Les Solutions de l'Etat de l'Art
- Text2Stories : un Agent pour Assister les Product Owner.
- Démonstration de Text2Stories
- Evaluation des Modèles

## 3. Limitations et Perspectives (7 min)

- Standardisation des Jeux de Données
- Développement de Métriques Quantitatives

## 4. Conclusion (3 min)

- Récapitulatif
- Mise à Disposition de Text2Stories

## 5. Références



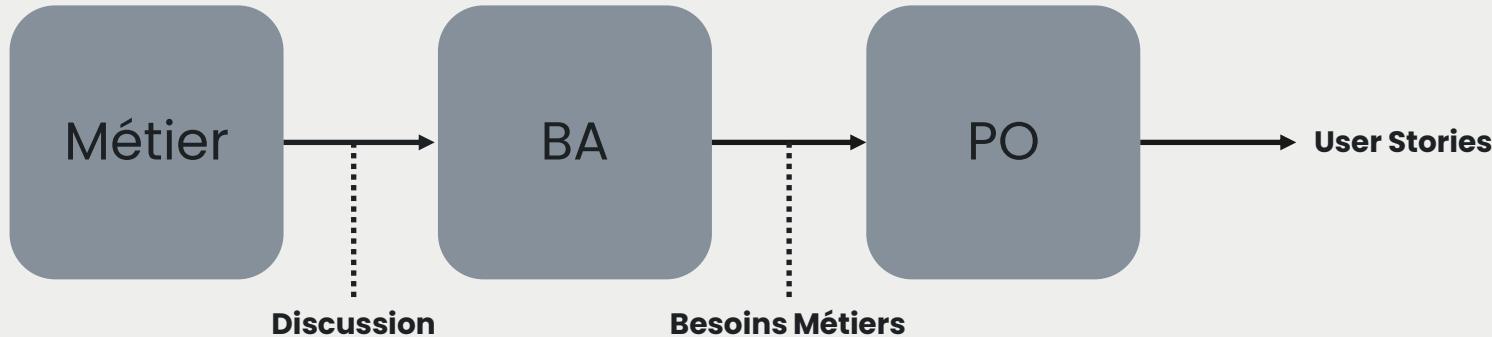
# 1. Contexte & Problématique

Les principaux défis de l'IA appliquée à l'Ingénierie des Exigences ?

## Contexte et Problématique

### User Stories : Une première définition

#### User Stories dans le Cycle de Vie d'un Projet



#### Exemple de User Story

##### Un Récit

En Tant que **<Acteur>**,  
je souhaite **<Action>**,  
afin de **<Valeur>**.

##### Critères d'Acceptation

- Lorsque **<Contexte>**, Quand je **<Action>**, Alors **<Résultat Attendu>**
- ....

##### Autre Informations

Notes, Captures d'écrans, Diagrammes explicatifs etc....

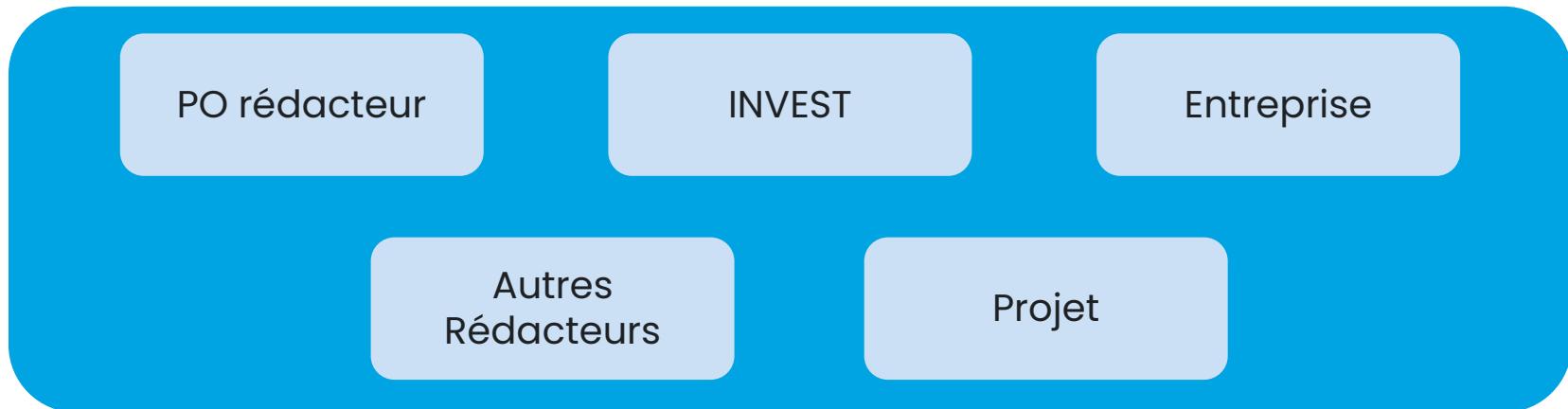
## Contexte et Problématique

### User Stories : Un second point de vue

#### Caractéristiques Objectives :

Source	Contexte	Règles de Gestion	Critères d'Acceptation	Multimodalité
Documents de specs	Epics Stories	Avec	Avec (Gherkin ou non)	Images (captures d'écrans, etc...)
Discussion BA/PO	Initiatives	Sans	Sans	Vidéos, etc...

#### Caractéristiques Subjectives :



Contexte et Problématique

## Générer des User Stories

**Question 1 : Comment générer un objet non-standardisé ?**

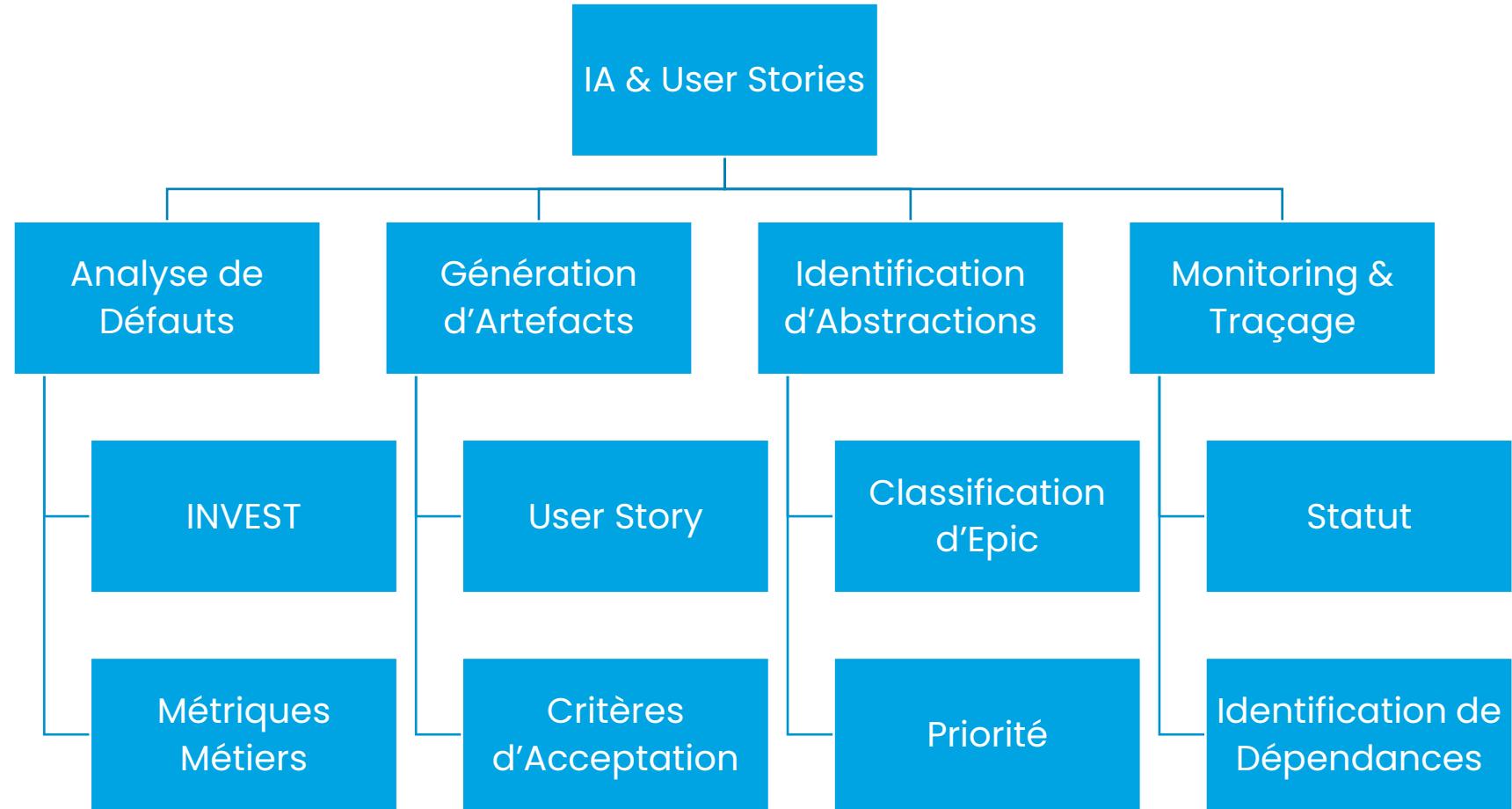
**Question 2 : Comment s'assurer de sa qualité ?**

# 2. Génération Automatiques de User Stories

Tour d'Horizon de l'Etat de l'Art

## Génération Automatique de User Stories

### Typologie de Tâches : IA & RE



Raharjana et al., 2021

## Génération Automatique de User Stories

### Quelques Solutions de l'Etat de l'Art

arXiv > cs > arXiv:2404.01558

Computer Science > Software Engineering

[Submitted on 2 Apr 2024]

**Automated User Story Generation with Test Case Specification Using Large Language Model**

Tajmilur Rahman, Yuecai Zhu

Chain-of-Thoughts

Prompt Engineering

Context Management

arXiv > cs > arXiv:2504.00513

Computer Science > Software Engineering

[Submitted on 1 Apr 2025 (v1), last revised 16 Jul 2025 (this version, v3)]

**Leveraging LLMs for User Stories in AI Systems: UStAI Dataset**

Asma Yamani, Malak Baslyman, Moataz Ahmed

LLM Benchmarking

Prompt Engineering

## Génération Automatique de User Stories

### Quelques Solutions de l'Etat de l'Art

Modèles	Découpage Fonctionnel (epics)	Génération de Critères d'Acceptation	Usage de Métriques Standards	Approche Agentique
GeneUS <i>(Rahman et al. 2024)</i>	✗	✓	✗	—
UStAI <i>(Yamani et al. 2025)</i>	✗	✗	✓	✗
Text2Stories <i>(notre modèle)</i>	✓	✓	✓	✓



Elevé

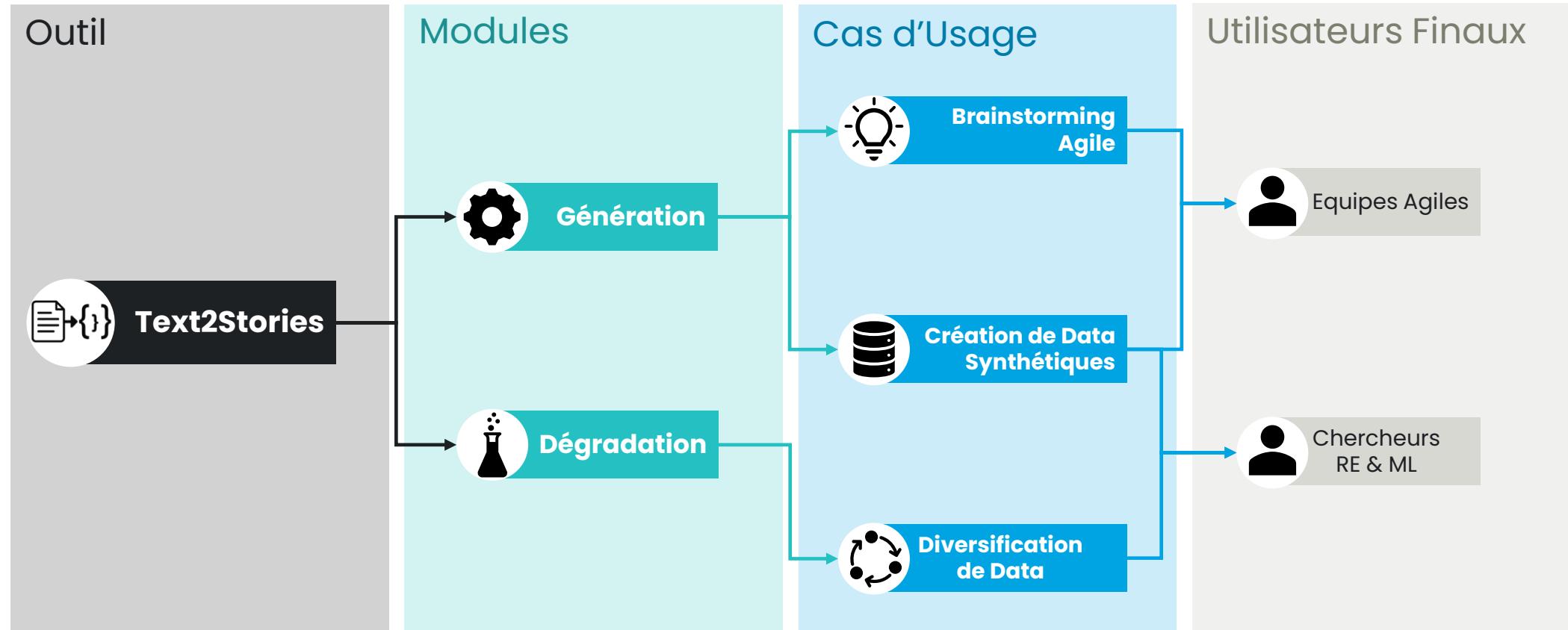


Modéré

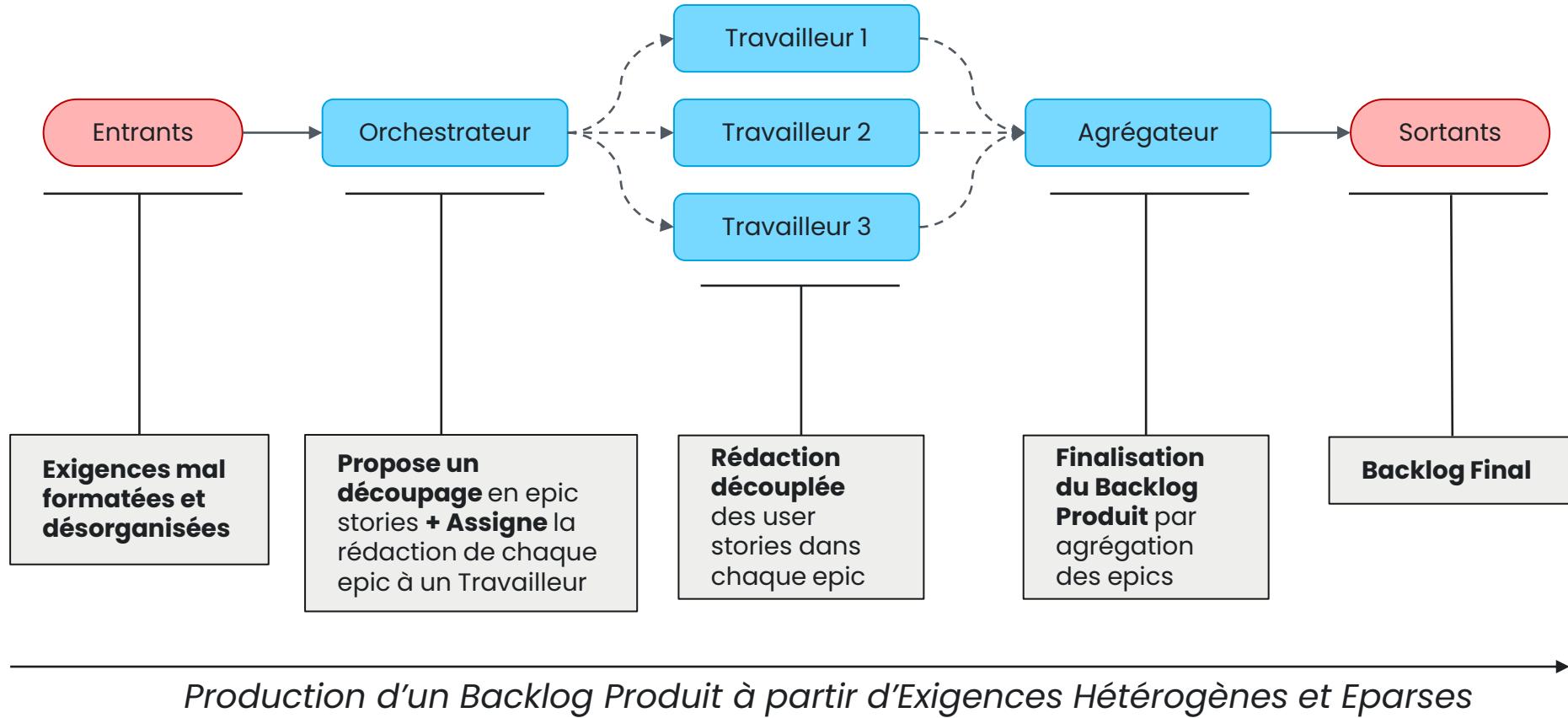


Faible

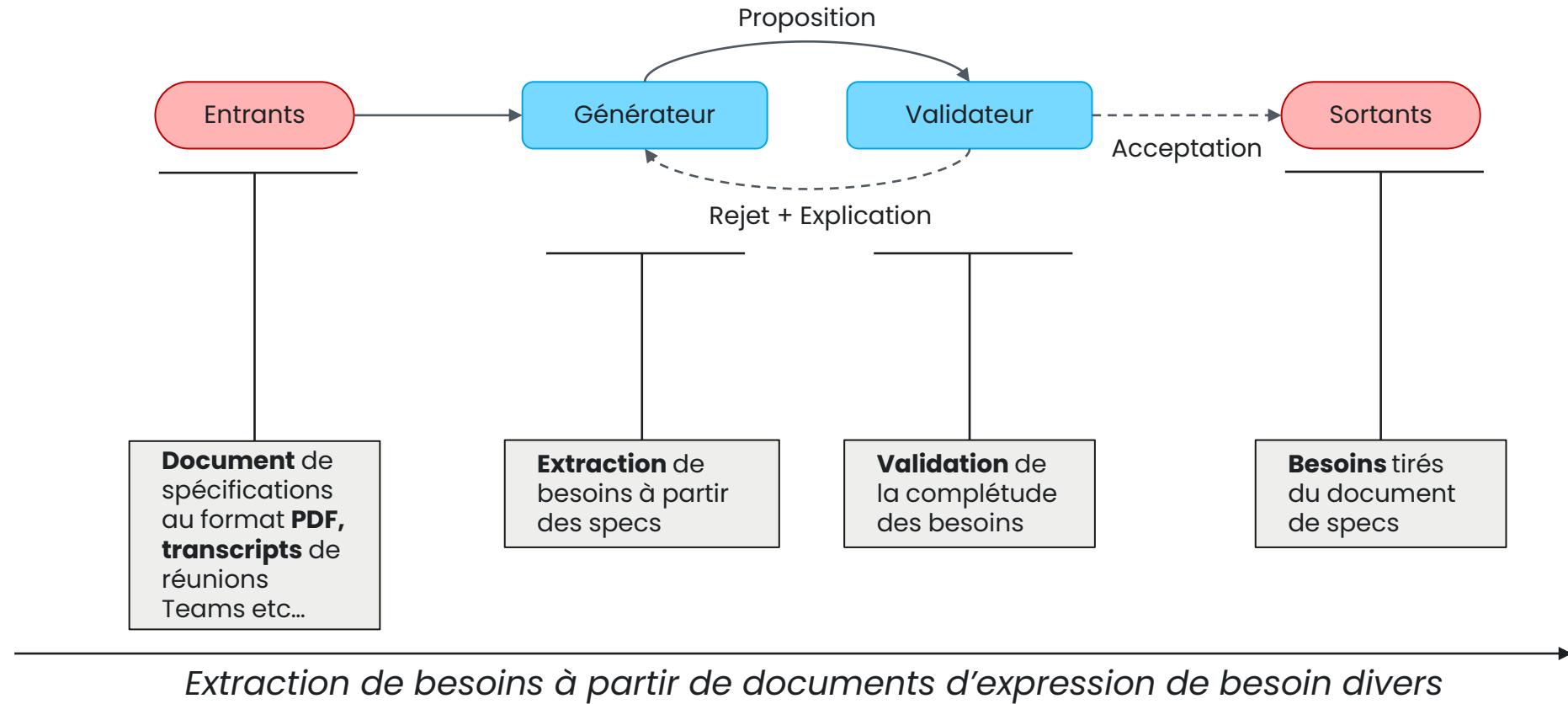
## L'Objectif du Projet



## Les Sous-Agents Standards (1/3) : Orchestrateur/Travailleur



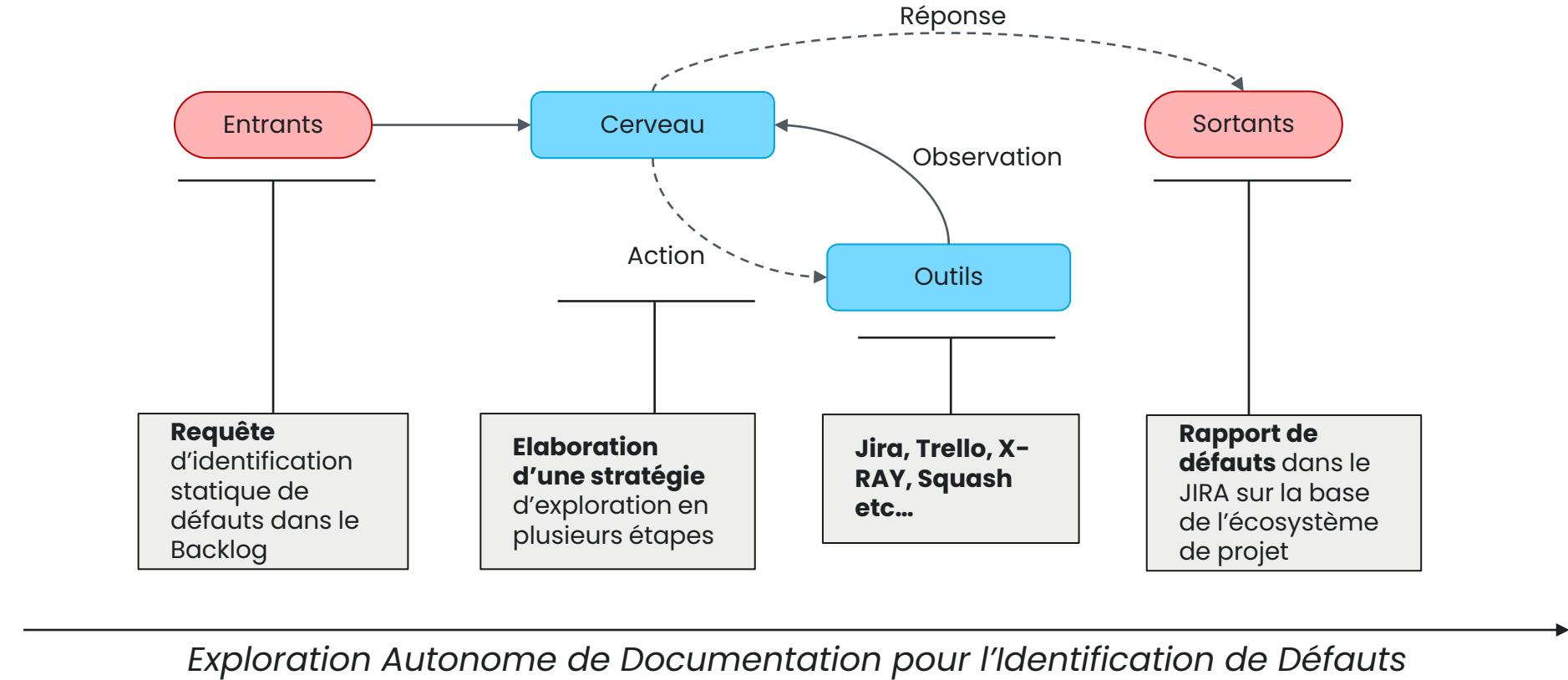
## Les Sous-Agents Standards (2/3) : Générateur/Evaluateur



Schluntz et Zhang, 2024

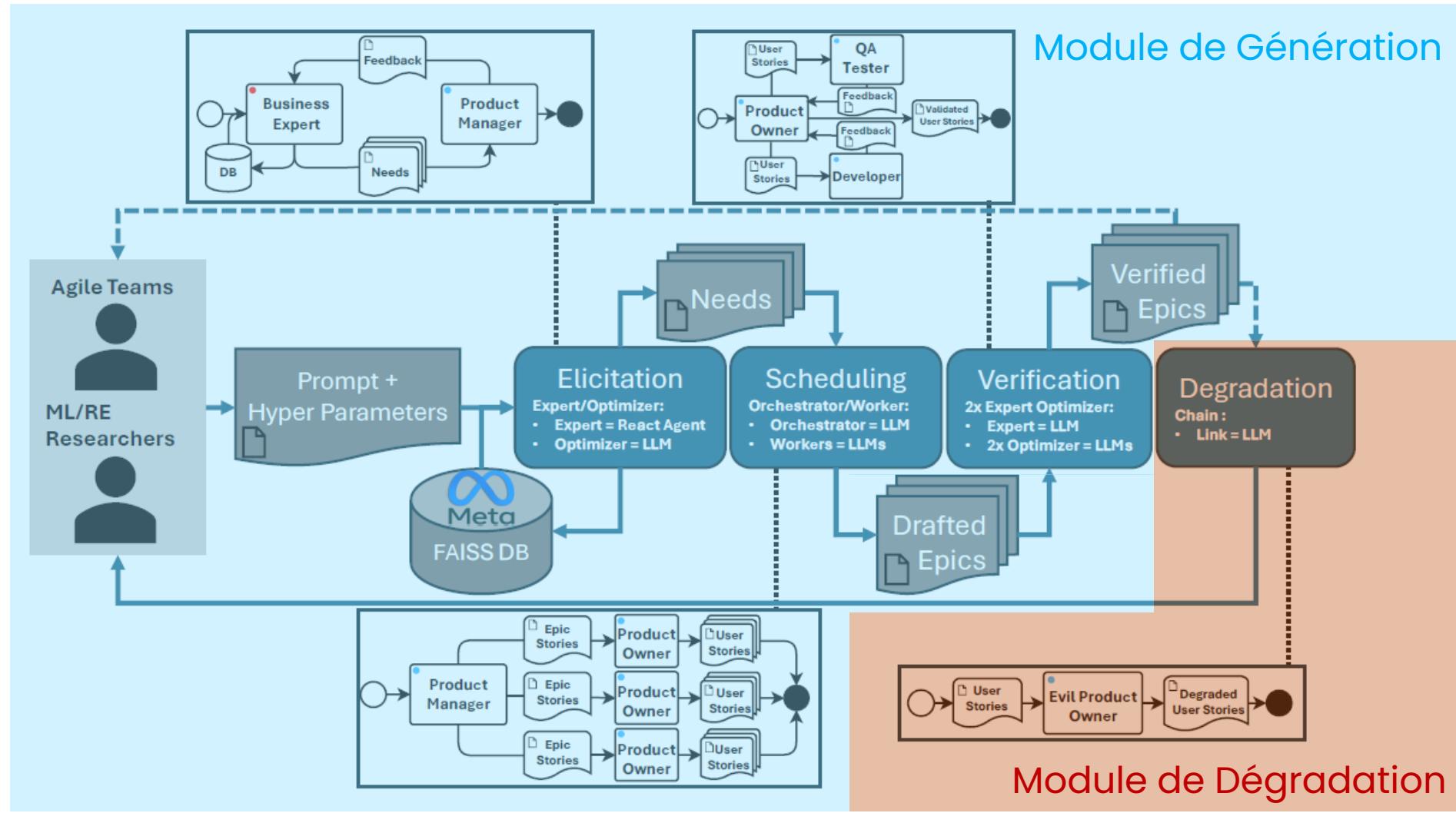


## Les Sous-Agents Standards (3/3) : ReAct, Combiner le Raisonnement et l'Action



Yao et al., 2023

## L'Architecture



## Text2Stories : 1<sup>er</sup> Projet de ma Thèse

### Démonstration

Generation
Degradation

Topic:

Ready to start  
Click the button to begin

**Start generation**



# 3. Limitations & Perspectives

Une meilleure évaluation des modèles pour accélérer  
l'innovation en IA pour le RE ?

## Les Limitations Actuelles

### 1ère Limite : La Standardisation des Jeux de Données

Home > Requirements Engineering > Article 383 cites

# Improving agile requirements: the Quality User Story framework and tool

RE 2015 | Open access | Published: 01 April 2016

Volume 21, pages 383–403, (2016) Cite this article

Garm Lucassen , Fabiano Dalpiaz, Jan Martijn E. M. van der Werf & Sjaak Brinkkemper

 arXiv > cs > arXiv:2306.12132 50 cites

Computer Science > Software Engineering

[Submitted on 21 Jun 2023]

## ChatGPT as a tool for User Story Quality Evaluation: Trustworthy Out of the Box?

Krishna Ronanki, Beatriz Cabrero-Daniel, Christian Berger

Jeu de Données :

- 1023 User Stories
- 18 Projets
- Privé



Jeu de Données :

- 3580 User Stories
- 8 Projets
- Public

## Les Limitations Actuelles

### 1ère Limite : La Standardisation des Jeux de Données

#### LiveBench (White et al. 2024)

Model	Organization	Global Average	Reasoning Average	Coding Average	Agentic Coding Average	Mathematics Average	Data Analysis Average	Language Average	IF Average
GPT-5 High	OpenAI	79.33	98.17	77.10	46.67	92.77	71.63	80.83	88.11
GPT-5 Medium	OpenAI	78.85	96.58	75.05	50.00	89.95	72.38	78.99	88.99
GPT-5 Pro	OpenAI	78.73	96.67	72.11	43.33	93.77	72.42	81.36	91.45
Claude Sonnet 4.5 Thinking	Anthropic	78.26	95.28	80.36	50.00	92.96	71.76	77.51	79.99
GPT-5 Codex	OpenAI	78.24	98.67	69.61	48.33	92.74	70.29	79.32	88.73
GPT-5 Mini High	OpenAI	75.31	91.44	68.20	43.33	90.69	71.95	75.63	85.90
Claude 4.1 Opus Thinking	Anthropic	75.25	93.19	74.66	45.00	91.16	71.14	71.21	80.38
GPT-5 Low	OpenAI	74.65	90.47	74.28	35.00	85.33	69.72	78.73	88.99
Claude 4 Sonnet Thinking	Anthropic	73.82	95.25	77.48	38.33	85.25	69.84	70.19	80.43
Grok 4	xAI	72.84	97.78	73.13	26.67	88.84	69.53	75.83	78.12



## Les Limitations Actuelles

### 2<sup>nde</sup> Limite : La Standardisation des Métriques : INVEST et QUS

#### INVEST :

##### *Independant :*

Est-ce que cette user story peut être développée sans dépendre de la complétiion d'une autre user story spécifique ?

##### *Negotiable :*

Est-ce que cette user story est suffisamment flexible pour permettre à l'équipe de discuter et d'affiner les détails de la solution pendant le développement ?

##### *Valuable :*

Est-ce que cette user story apporte une valeur tangible et identifiable à l'utilisateur final ou à un stakeholder ?

##### *Estimable :*

L'équipe Agile est-elle en mesure de fournir une estimation raisonnable (même approximative) de l'effort nécessaire pour réaliser cette user story ?

##### *Small :*

Cette user story est-elle suffisamment petite pour être raisonnablement complétée au cours d'un seul sprint ?

##### *Testable :*

Est-il possible de rédiger les cas de teste pour la user story en se basant sur les informations contenues dans la description et les critères d'acceptation ?

#### QUS (Quality User Story) :

##### *Syntaxe :*

- **Atomique** : Couvre une seule fonctionnalité ou élément de travail.
- **Bien Formée** : Respecte la structure de base : « En tant que <acteur>, je veux <action> ».
- **Minimale** : Concise et se concentre sur le modèle standard : « En tant que <acteur>, je veux <action>, afin de <valeur> ».

##### *Sémantique :*

- **Conceptuellement Saine** : <action> décrit ce que l'utilisateur veut, <valeur> décrit pourquoi il le veut.
- **Orientée Problème** : Enonce le besoin ou l'objectif de l'utilisateur, pas la manière de l'implémenter.
- **Non-Ambiguë** : Utilise un langage clair, précis.
- **Sans Conflit** : Ne contredit pas les exigences ou les hypothèses d'une autre user story.

##### *Pragmatisme :*

- **Phrase Complète** : Idéalement basée sur le modèle standard.
- **Estimable** : Suffisamment petite et claire pour que l'équipe puisse lui attribuer une estimation d'effort raisonnable.
- **Unique** : Représente une exigence distincte des autres.
- **Uniforme** : Respecte le format des autres items du Backlog.
- **Indépendante** : Implémentable indépendamment des autres user stories contenues dans le Backlog.
- **Complète** : L'ensemble des user stories permettent la conception un produit fonctionnel et utilisable.



## Les Limitations Actuelles

### 2<sup>nde</sup> Limite : La Standardisation des Métriques

Métrique	Quantitative & Reproductible	Informative & Utile au Métier	Standard
INVEST (Business)	✗	✓	✓
QUS (Lucassen et al. 2015)	✓	—	✓
RUST (Ronanki et al. 2024)	✗	—	✗
ISO25010 (Gilson et al. 2019)	✗	—	✗



Elevé



Difficile à Estimer



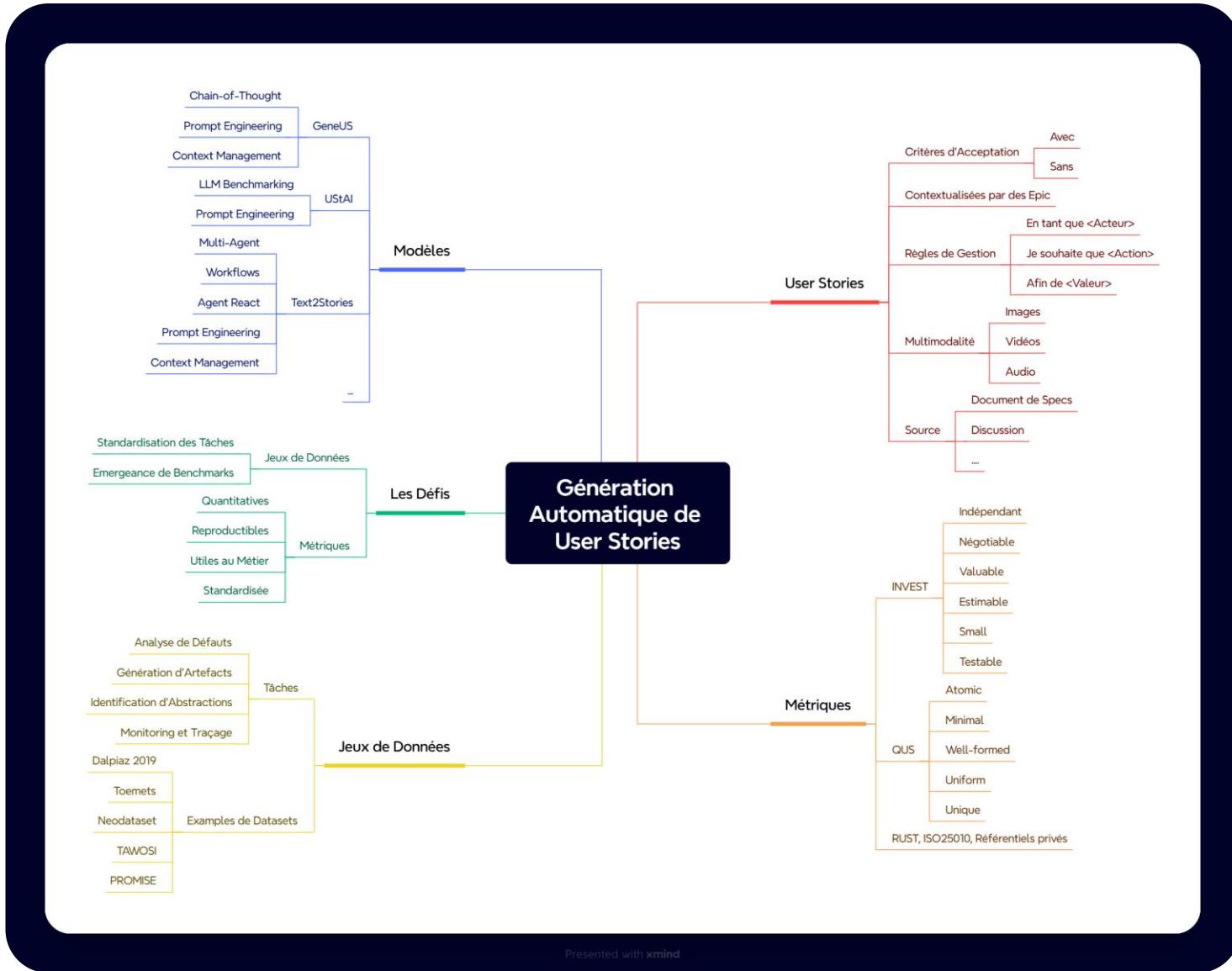
Faible

# 4. Conclusion

Récapitulatif & Mise à disposition de Documents

## Conclusion

### Récapitulatif : Points Clés



## Conclusion

## Remerciement

### Notre Equipe



**Hassan Imhah**  
Leader et Expert  
Qualité Logicielle  
*Onepoint*



**Nédra Mellouli**  
Directrice Majeure  
DATA&IA  
*ESILV*



**Christophe Rodrigues**  
Directeur Majeure  
Cybersécurité  
*ESILV*



**Nicolas Travers**  
Co-Directeur  
Laboratoire DVRC  
*ESILV*



**Marius Ortega**  
Doctorant IA et  
Qualité Logicielle  
*ESILV & Onepoint*

### Notre Projet

**Text2Stories** Public

Repository listing information not listed in the paper

Python MIT License Updated on Jul 23

<https://github.com/MariusAAROS/Text2Stories>

# 5. Références

Mes sources pour cette présentation

## Références

### Sources (1/2)

1. Raharjana IK, Siahaan D, Faticahah C. User Stories and Natural Language Processing: A Systematic Literature Review. *IEEE Access*. 2021;9:53811-26.
2. Rahman T, Zhu Y, Maha L, Roy C, Roy B, Schneider K. Take Loads Off Your Developers: Automated User Story Generation using Large Language Model. In: 2024 IEEE International Conference on Software Maintenance and Evolution (ICSME). 2024 . p. 791-801. Disponible sur: <https://ieeexplore.ieee.org/abstract/document/10795004>
3. Yamani A, Baslyman M, Ahmed M. Leveraging LLMs for User Stories in AI Systems: UStAI Dataset. In: Proceedings of the 21st International Conference on Predictive Models and Data Analytics in Software Engineering. New York, NY, USA: Association for Computing Machinery; 2025 . p. 21-30. (PROMISE '25). Disponible sur: <https://dl.acm.org/doi/10.1145/3727582.3728689>
4. Schluntz E, Zhang B. Anthropic. 2024 . Building Effective AI Agents. Disponible sur: <https://www.anthropic.com/engineering/building-effective-agents>
5. Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, et al. ReAct: Synergizing Reasoning and Acting in Language Models. ICLR. janv 2023 ; Disponible sur: <https://par.nsf.gov/biblio/10451467-react-synergizing-reasoning-acting-language-models>
6. White C, Dooley S, Roberts M, Pal A, Feuer B, Jain S, et al. LiveBench: A Challenging, Contamination-Limited LLM Benchmark [Internet]. arXiv; 2025. Disponible sur: <http://arxiv.org/abs/2406.19314>



## Références

### Sources (2/2)

7. Gilson F, Galster M, Georis F. Extracting Quality Attributes from User Stories for Early Architecture Decision Making. 2019.
8. Ronanki K, Cabrero-Daniel B, Berger C. ChatGPT as a Tool for User Story Quality Evaluation: Trustworthy Out of the Box? In: Kruchten P, Gregory P, éditeurs. Agile Processes in Software Engineering and Extreme Programming – Workshops. Cham: Springer Nature Switzerland; 2024. p. 173–81.



**Mon Contact :**

Mail : [m.ortega@groupeonepoint.com](mailto:m.ortega@groupeonepoint.com)

Merci pour votre  
attention

→ 20/11/2025