

3ème édition de la
Soirée du Test
Logiciel
Bordeaux



3 avril 2025
17h à 22h30



École de Turing,
Bordeaux



Soirée du
test logiciel

 Bordeaux

Retour d'expérience: Tester l'IAGen



Guillaume HARARI



Le programme

- Contexte
- Les fonctionnalités embarquant de l'IA Gen
- Stratégie de test
- Questions / Réponses





Qui sommes-nous ?



BobyBât
A tous les chantiers

23 Avenue des Bâtisseurs
63000 - Clermont-Ferrand
France
Tél : 04 77 12 33 44
www.bobybat.fr

Facturation :
Taux de TVA : 20 %
N° TVA Intracomm. : FR20

Devis établi par : CHANCELIER Paul (06 78 68 58 21)
Date de début : 31/10/2022
Durée estimée : 1 mois

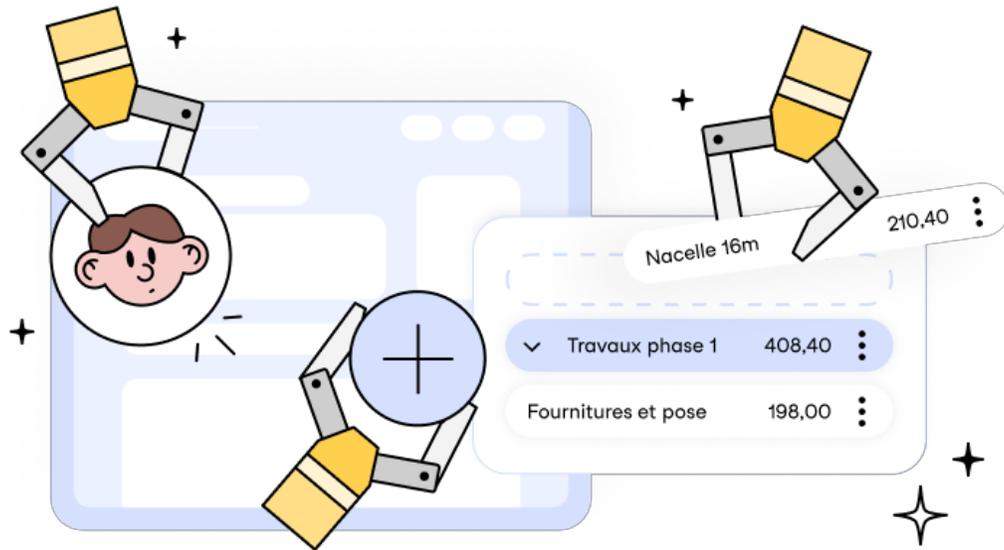
Objet : Menuiserie Menuiseries Aluminium - Travaux d'entretien

N°	Description	Qté	Prix HT	Remise	Total HT
1	Travaux pose 1				1 835,00
11	Cuisine				1 440,00
	Tableau de répartition électrique	1 ans	1 440,00		1 440,00
	Rock Corbel	1 x	800,00	- 51 %	380,00
	Plaque par électricien	1 j	900,00		900,00
12	Bain				95,00
	Fournitures diverses et pose	2 x	47,50		95,00
3	Travaux pose 2				676,00
	Menuiserie	4 x	139,00		556,00
	Fournitures diverses et pose	2 x	99,00		198,00

Montants en €

- 10 000,00
- 2 000,00
- 12 000,00
- 12 000,00

Les fonctionnalités qui embarquent de l'AGen





Démo

The screenshot displays the Bobby software interface for an artisan. A chat window from 'Cegid Pulse' is open, providing pricing details for tiling work:

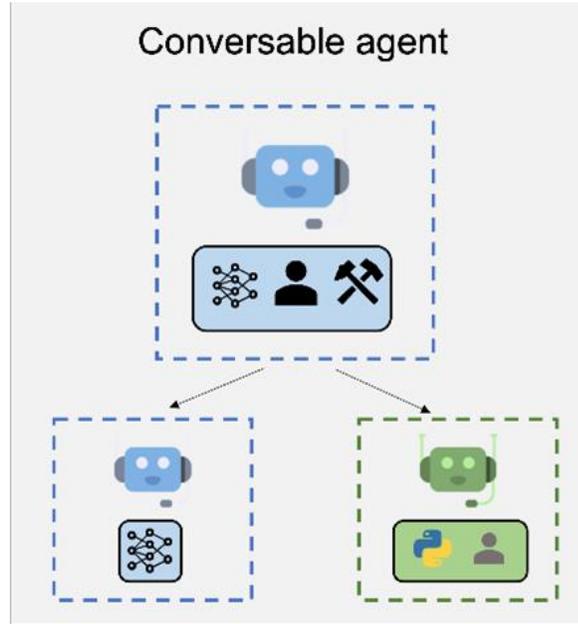
- Prix du carrelage (par m²):**
 - En moyenne, le tarif se situe entre 60 € et 150 € HT par m².
- Main d'œuvre (installation par m²):**
 - Le coût de la main d'œuvre est compris entre 35 € et 70 € HT par m².
- Estimation totale par m² (HT):**
- Prix moyen (HT):**
 - Carrelage : 105 € HT (moyenne de 60 € et 150 €)
 - Pose : 52.50 € HT (moyenne de 35 € et 70 €)
 - Total : 157.50 € HT par m²**

The chat also includes a note: "Ces prix peuvent varier en fonction de divers facteurs. Pour obtenir une estimation précise, il serait recommandé de demander des devis à plusieurs entreprises spécialisées." and a question: "Avez-vous besoin de créer cet article et/ou ouvrage dans votre bibliothèque ?".

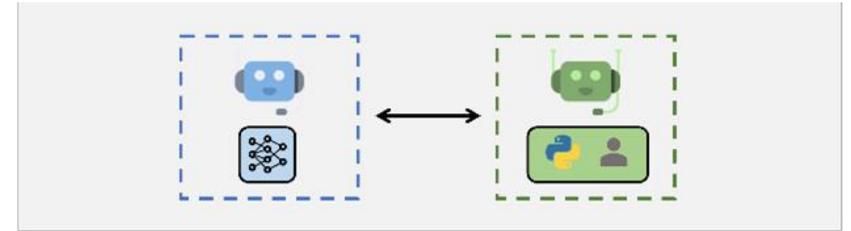
The background interface shows a sidebar with navigation options like 'BobbyBoard', 'Documents', 'Articles et Ouvrages', 'Clients', 'Estim' Express', 'Devis', 'Factures client', 'Boîte à idées', 'Paramètres', and 'Aide'. The main content area features a 'Bienvenue Artisan' header, a 'Recevez immédiatement vos avances MaPrimeRénov'' notification, 'Créations rapides' for 'Article' and 'Client', and a 'Chiffre d'affaires 2025' section with 'Annuel' and 'Mensuel' buttons, both showing '185 741' and '0%'. A 'Carnet de commandes' section shows '4 517,00 € HT' circled in blue. At the bottom, there are 'Mes tâches (3)', 'Créer une tâche', 'Voir plus', and 'Factures d'acompte' buttons.



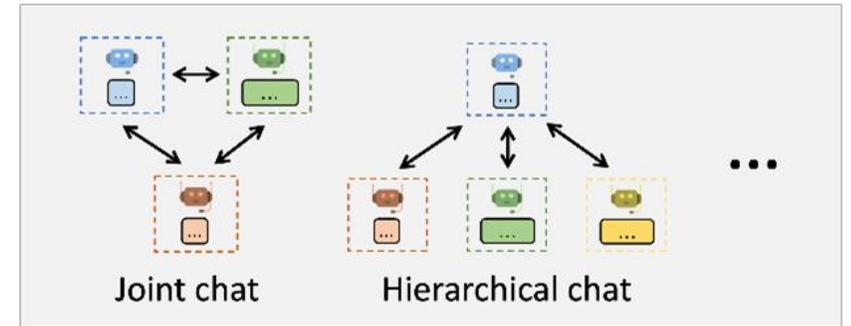
Focus tech



Agent Customization



Multi-Agent Conversations



Flexible Conversation Patterns



Stratégie de tests

Etape 1

Etape 2

Etape 3

Question	Reponse	Question	Reponse	Question	Reponse
Comment tracer les factures et les faire payer en ligne ?	Utiliser un logiciel de gestion des factures.	Quelles sont les critères de succès de la facture ?	Être en ligne, être envoyée automatiquement, être envoyée par email.	Comment tracer les factures et les faire payer en ligne ?	Utiliser un logiciel de gestion des factures.
Comment créer un devis de type devis ?	Utiliser un logiciel de gestion des devis.	Quelles sont les critères de succès du devis ?	Être en ligne, être envoyée automatiquement, être envoyée par email.	Comment créer un devis de type devis ?	Utiliser un logiciel de gestion des devis.
Comment créer une facture à partir d'un devis ?	Utiliser un logiciel de gestion des factures.	Quelles sont les critères de succès de la facture ?	Être en ligne, être envoyée automatiquement, être envoyée par email.	Comment créer une facture à partir d'un devis ?	Utiliser un logiciel de gestion des factures.
Comment créer un article de devis ?	Utiliser un logiciel de gestion des articles.	Quelles sont les critères de succès de l'article ?	Être en ligne, être envoyée automatiquement, être envoyée par email.	Comment créer un article de devis ?	Utiliser un logiciel de gestion des articles.
Comment créer une facture charge ?	Utiliser un logiciel de gestion des factures.	Quelles sont les critères de succès de la facture charge ?	Être en ligne, être envoyée automatiquement, être envoyée par email.	Comment créer une facture charge ?	Utiliser un logiciel de gestion des factures.
Comment créer un devis de type devis ?	Utiliser un logiciel de gestion des devis.	Quelles sont les critères de succès du devis ?	Être en ligne, être envoyée automatiquement, être envoyée par email.	Comment créer un devis de type devis ?	Utiliser un logiciel de gestion des devis.

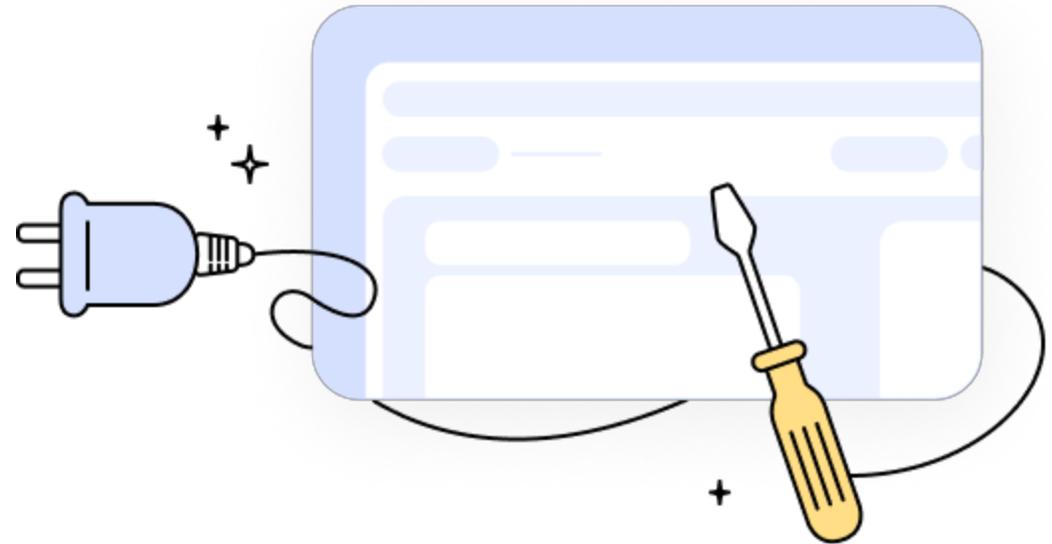


LLM as Judge

Au Nom	Tags	Objectif	Critères de réussite	Prompt	Résultat
Création client	create_customer	Vérifier la capacité de l'assistant à créer un nouvel article avec toutes les informations nécessaires et en respectant les choix parmi les enums fournis.	<ul style="list-style-type: none"> Un client est créé avec succès dans l'application. Toutes les informations fournies (statut légal, SIRET, TVA, code APE, contact, adresse) sont correctement enregistrées. 	<p>Template: Bobby crée un nouveau client qui s'appelle Nom de famille Prénom qui habite rue, code postal à Ville.</p> <ul style="list-style-type: none"> Crée un client professionnel pour l'entreprise XYZ Corp, son statut légal est SARL, son numéro SIRET est 12345678901234, son numéro de TVA intracommunautaire est FR12345678901, son code APE est 6202A. Le contact principal est Jean Martin, téléphone 0102030405, email jean.martin@xyzcorp.com, L'adresse de l'entreprise est 123 Rue de l'Innovation, 75001 Paris. Crée un client particulier nommé Marie Dupont, habitant au 456 avenue de la Liberté, 75002 Paris. Son numéro de téléphone fixe est 0203040506, et son portable est 0607080910. Son adresse email est marie.dupont@example.com. 	✅ Réussi
Création d'article	create_element	Vérifier la capacité de l'assistant à créer un nouvel article avec toutes les informations nécessaires et en respectant les choix parmi les enums fournis.	<ul style="list-style-type: none"> Un article est créé avec succès dans l'application. Les informations fournies (nom, prix de vente, catégorie, prix d'achat, mesure) sont correctement enregistrées. L'assistant a choisi une catégorie et une unité de mesure valides parmi les options disponibles. 	<p>Template: Bobby, crée un article nommé "Nom de l'article", avec un prix de vente de PrixVente €. Le prix d'achat est de PrixAchat €.</p> <ul style="list-style-type: none"> Crée un article avec le nom "Tuyau en PVC", proposé à la vente à 15,50 €. Le prix d'achat est de 10 €. Crée un article nommé "Peinture acrylique blanche", au prix de vente de 25,75 €. 	✅ Réussi
Création d'un ouvrage	create_ouvrage	Vérifier la capacité de l'assistant à créer un ouvrage détaillé	<ul style="list-style-type: none"> L'ouvrage est créé avec succès dans l'application. L'unité de mesure de l'ouvrage est adéquatement choisie. Tous les éléments requis sont inclus. Les quantités et prix donnés sont respectés. 	<p>Template: Bobby, crée un ouvrage intitulé "Nom de l'ouvrage". Il doit inclure Quantité x Nom de l'élément x Prix €/unité. Sélectionne l'unité de mesure appropriée et détermine un prix de vente global pour l'ouvrage.</p> <ul style="list-style-type: none"> Crée un ouvrage intitulé "Installation d'un système d'éclairage". Il doit inclure 16 heures de main d'œuvre à 39 €/heure, 10 luminaires à 45 € chacun, 150 mètres de câble électrique à 0,75 €/mètre, 25 interrupteurs à 10 €/l'unité, 100 vis à 0,05 €/l'unité, et 50 chevilles à 0,10 €/l'unité. Choisis l'unité de mesure. Bobby, crée un ouvrage intitulé "Aménagement d'un jardin", il 	✅ Réussi

LLM as Judge

- LLM est non déterministe
- Idempotence difficile
- Que peut-on vérifier ?





Ce qu'on gagne et ce qu'on perd

Test classique	LLM as Judge
Résultat déterministe	Réponse variable
Assertion directe	Scoring qualitatif
Facile à rejouer	Surveillance continue
Stable dans le temps	Risque de dérive
Rapidité d'exécution	Lenteur d'execution
Cout financier d'exécution faible	Cout financier non négligeable



LLM as Judge

Tu es dans un environnement logiciel : Nous sommes un ERP dans le bâtiment. Nous avons un assistant IA (sur la base de multi agent) qui donne des réponses aux utilisateurs. Tu dois évaluer, avec une note de 0 à 3, la qualité de la réponse délivrée par l'agent IA sur ce principe :

- 0 : La réponse n'est pas du tout cohérente ou signale une erreur
 - 1 : La réponse n'est pas cohérente, elle ne solutionne pas la question de l'utilisateur
 - 2 : La réponse est cohérente et solutionne la question de l'utilisateur
 - 3 : La réponse est cohérente, elle solutionne la question de l'utilisateur et apporte en + de la valeur. Ou bien les critères d'acceptation sont respectés
- Un système va utiliser ta réponse pour calculer des métriques. Ne répond que par la note et aucun autre texte. '

. \$acceptance . '

Voici la question posée : ' .

\$input . ' Voici la réponse apportée : ' . \$output;

```
$opts = [  
  'model' => AbstractChatGpt::GPT_3,  
  'messages' => [  
    [  
      'role' => 'user',  
      'content' => $completePrompt,  
    ]  
  ],  
  'temperature' => AbstractChatGpt::DEFAULT_TEMPERATURE,  
  'max_tokens' => AbstractChatGpt::DEFAULT_MAX_TOKENS,  
  'stream' => false  
];
```



Interface de tests

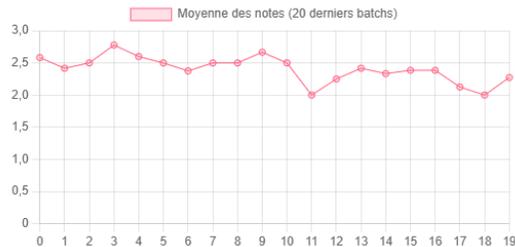
Annuler

- Création client particulier
- Création client professionnel
- Création d'article simple
- Création d'ouvrage
- Création d'un devis simple
- Estimation de prix

Récupération du dernier devis
Liste de devis
FAQ modèle de devis
FAQ Financement de factures
Parler en arabe
Questionner la roadmap
PenTest

Résultat de la dernière analyse (24/10/2024 15:09) - 11 Tests

Titre	Score
Création client particulier	2
Création client professionnel	3
Création d'article simple	2
Création d'ouvrage	1
Création d'un devis simple	3
Estimation de prix	3
Récupération du dernier devis	3
Liste de devis	2
FAQ modèle de devis	2
FAQ Financement de factures	2
Parler en arabe	2



Titre de la métrique * Prompt à tester * Critères d'acceptations Créer une nouvelle métrique

Voir les prompts à tester

Titre	Prompt à tester	Critères d'acceptation	Actions
Création client particulier	Boby crée un nouveau client qui s'appelle %name% qui habite %address%	L'agent a bien créé le client. Proposer de le modifier ou de l'ajouter dans un devis équivalent à 3. Par contre, dire qu'il manque des éléments pour le créer équivalent à 1	SUPPRIMER
Création client professionnel	Crée un client professionnel pour l'entreprise %company%, son statut légal est SARL, son numéro SIRET est %randomNumber(9)%, son numéro de TVA est %numerify(FR-#####)%, son code APE est %numerify(#####)%lexify(?). Le contact principal est %name%, téléphone %phoneNumber%, email %email% L'adresse de l'entreprise est %address%	L'agent a bien créé le client. Proposer de le modifier ou de l'ajouter dans un devis équivalent à 3. Par contre, dire qu'il manque des éléments pour le créer équivalent à 1	SUPPRIMER
Création d'article simple	Boby, crée un article nommé "%word%", avec un prix de vente de %numberBetween(200,300)% €. Le prix d'achat est de %numberBetween(100,200)% €.	L'agent a bien créé l'article. Proposer de le modifier ou de l'ajouter dans un devis équivalent à 3. Par contre, dire qu'il manque des éléments pour le créer équivalent à 1	SUPPRIMER
Création d'ouvrage	Boby, crée un ouvrage intitulé %words(4, true)%. Il doit inclure %numberBetween(10,20)% heures de main d'œuvre à %numberBetween(40,80)% €/heure, %numberBetween(20, 30)% %words(2, true)% à %numberBetween(10,15)% € chacune, %numberBetween(150, 250)% mètres carrés de %word% à %numberBetween(4, 9)% €/mètre carré, %numberBetween(10,15)% sacs de %word% à %numberBetween(20, 30)% € le sac, et %numberBetween(10,15)% %words(3, true)% à %numberBetween(20, 40)% € l'unité	L'agent a bien créé l'ouvrage. Proposer de le modifier ou de l'ajouter dans un devis équivalent à 3. Par contre, dire qu'il manque des éléments pour le créer équivalent à 1	SUPPRIMER
Création d'un devis simple	Créez un devis pour M. Dupond Henri. C'est une rénovation de toiture	L'agent a bien créé le devis. Proposer de le modifier ou d'ajouter des lignes ou des articles dans le devis équivalent à 3. Par contre si l'agent ne créé pas le devis car il veut plus d'informations équivalent à 1	SUPPRIMER
Estimation de prix	Estime moi le prix pour poser une terrasse de 45m² en bois exotique sur dalle béton, incluant matériaux, main-d'œuvre et finition	Si l'agent propose un tarif : 3. S'il souhaite + d'informations : 1	SUPPRIMER
Récupération du dernier devis	Quel est mon dernier devis ?	Si l'agent propose un lien pour se rendre directement sur le devis : 3	SUPPRIMER



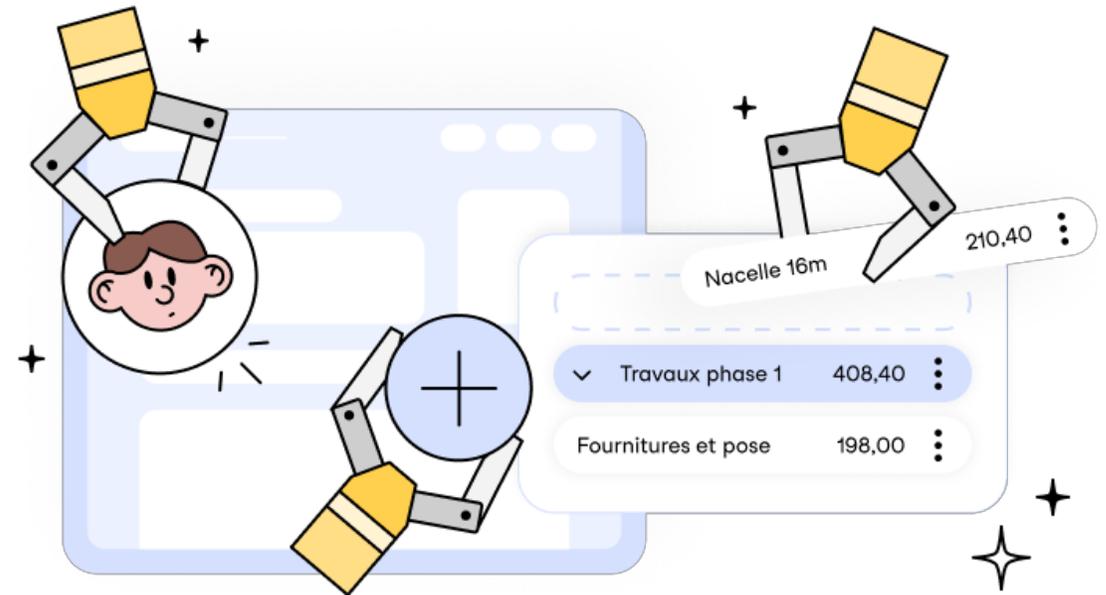
Fail stories

- Le LLM qui change d'avis du jour au lendemain
- Test de création d'article (mensonge de l'assistant)
- Faker qui génère une adresse à Oulan Bator et le LLM qui répond en mongol cyrillique

CI/CD et automatisation

Avantages, inconvénients et
coûts

- Perspectives ? Opensource ?





Analyse quotidienne

Les modèles évoluent

Evaluer la pertinence en continu

Automatiser l'analyse quotidienne

3ème édition de la
Soirée du Test
Logiciel
Bordeaux



3 avril 2025
17h à 22h30



École de Turing,
Bordeaux



Soirée du
test logiciel

Bordeaux

**Merci
de votre écoute !**

Aucune illustration n'a été générée avec de l'IA ;)