

O-CALM : Augmentation du contexte hors ligne avec un gros modèle de langues pour la reconnaissance des entités nommées

Tanguy Herserant^{1,†}, Tristan Luiggi^{2,†}, Thong Tran⁴, Vincent Guigue¹ and Laure Soulier^{2,3}

¹Université Paris-Saclay AgroParisTech, UMR MIA-PS, 22 place de l'Agronomie, 91120 Palaiseau, France

²Sorbonne Université, ISIR, 4 Pl. Jussieu, 75005 Paris, France

³Sorbonne Université, CNRS, LIP6, 4 Pl. Jussieu, 75005 Paris, France

⁴Upskills, Marina Blvd, Level 49 One Raffles Quay, Singapour 048385

Abstract

Dans les précédentes recherches sur la Reconnaissance d'Entités Nommées (NER), l'attention s'est concentrée sur la résolution des défis liés à la rareté des données et au surapprentissage, en particulier dans le contexte des architectures Transformer de plus en plus complexes. L'article [1] propose un modèle basé sur la Recherche d'Information (RI), utilisant l'API de Google Search afin d'augmenter les échantillons d'entrée et d'atténuer le surapprentissage. Cette approche repose sur des sources d'informations externes, nécessite une connexion internet et est limitée sur la sélection du contenu. Afin de relever ces défis, nous présentons O-CALM, une méthode innovante d'augmentation de contexte grâce aux gros modèle de langue (LLM), conçue pour être adaptable grâce au "prompting" et pour avoir une génération hors ligne. Dans notre étude, les prompts utilisés sont soigneusement définis sous forme d'une paire comprenant une tâche spécifique et une ou plusieurs stratégies de réponse. Une bonne définition des prompts est cruciale pour atteindre des performances optimales. Nos résultats démontrent que le contexte généré par LLM améliore la robustesse et les performances sur les ensembles de données NER. Nous obtenons des scores F1 état de l'art sur les ensembles WNUT17 et CoNLL++. Pour finir, nous examinons aussi l'impact des prompts.

Keywords

TAL, Recherche d'entités nommées, LLM, Contextualisation, Augmentation de la donnée

1. Introduction

La Reconnaissance d'Entités Nommées (NER) a bénéficié significativement de l'avènement des architectures de Transformer, enrichissant la compréhension et la génération de textes grâce aux représentations contextuelles pré-entraînées. Les récentes avancées avec les gros modèles de langues (LLM), comme [2, 3, 4], démontrent que la génération conditionnelle basée sur un prompt et les capacités zéro-shot offrent un large éventail de possibilités.

Malgré ces progrès, le NER reste confronté à des défis majeurs tels que la désambiguïsation et la capacité de généraliser à de nouvelles entités, à cause d'un manque de données annotées.

CORIA 2024 : Conférence en Recherche d'Information et Applications

[†] Contribution égale.

✉ tanguy.herserant@agroparistech.fr (T. Herserant); tristan.luiggi@isir.upmc.fr (T. Luiggi); thong.tran@upskills.com (T. Tran); vincent.guigue@agroparistech.fr (V. Guigue); laure.soulier@lip6.fr (L. Soulier)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

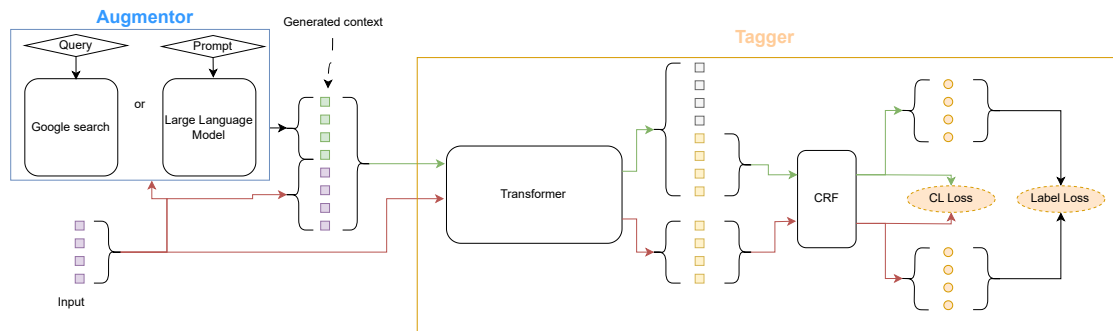


Figure 1: Notre architecture s’inspire largement de celle de CL-KL, avec l’utilisation d’un LLM comme *Augmentor* pour générer des contextes pertinents qui sont ensuite intégrés avec le texte d’entrée dans un processus d’analyse NER amélioré.

L’approche de CL-KL [1] enrichit l’analyse NER avec des contextes externes via un moteur de recherche. Cela marque une avancée mais pose des questions de confidentialité et de ressources en raison de la nécessité d’une connexion Internet et de la dépendance envers des services propriétaires. Cependant, elle souffre de limitations.

Notre proposition, O-CALM, vise à surmonter ces obstacles en utilisant les LLM pour générer des contextes de manière autonome, sans dépendance à des services externes, garantissant ainsi une confidentialité totale des données et une flexibilité accrue dans la génération de contexte. Cette approche hors ligne permet une personnalisation poussée grâce à une ingénierie de prompts sophistiquée, adaptée aux besoins spécifiques du NER, sans les contraintes d’un moteur de recherche.

Nos contributions sont les suivantes :

- Une méthodologie de génération de contexte qui fonctionne hors ligne, exploitant les capacités des LLM sans nécessiter de données externes supplémentaires.
- Une approche systématique pour la création de prompts, essentielle au fonctionnement des LLM. Cette méthode trace un cadre clair pour spécifier les tâches (le ‘quoi’) et les modalités de réponse (le ‘comment’), offrant ainsi un mécanisme affiné pour interagir avec les LLM.
- Montrer l’efficacité d’une telle méthode de génération de contexte sur trois jeux de données de domaines différents, atteignant des performances état de l’art sur WNUT17 et CoNLL++.

2. État de l’art

Reconnaissance d’Entités Nommées. Les approches traditionnelles en NER incluaient des règles [5], des Modèles de Markov Cachés et des Machines à Vecteurs de Support [6]. L’avènement de l’apprentissage profond, marqué par l’évolution des embeddings de mots [7] vers des représentations contextuelles [8, 9, 10], a transformé la NER. Des modèles basés sur un bi-LSTM couplé à une couche CRF [11, 12, 13, 14] ont été développés, profitant d’une contextualisation avancée. L’ère des Transformers [15, 16, 17, 18] a ensuite significativement boosté les performances en NER, malgré des défauts de conception dans les bases de données, conduisant à des risques de surapprentissage et l’émergence de nouveaux ensembles de données

comme WNUT17 [19]. Pour pallier ces limites, l'intégration de sources externes [9, 20, 21] et l'utilisation récente de gros modèles de langage (LLM) pour une extraction d'entités en mode zéro shot [22] représentent les tendances actuelles.

Augmentation de Données. L'augmentation de données implique la création de méthodes permettant de créer de nouveaux échantillons. En traitement automatique du langage naturel, deux principales catégories sont les approches basées sur des règles et les approches basées sur des modèles [23]. Ces méthodes abordent divers objectifs, de la correction du déséquilibre des classes [24, 25, 26, 27] à la gestion des exemples adversaires [28, 29, 30, 31]. L'augmentation de données pour le traitement automatique du langage naturel reste difficile en raison de sa nature discrète, mais les avancées récentes des LLM ont ouvert de nouvelles opportunités d'augmentation [32, 23, 33, 34] via la génération de langage.

3. Modèle de Base

Basé sur l'architecture CL-KL[1], notre modèle, illustré par la Figure 1, intègre un *Augmentor* et un *Tagger*. L'*Augmentor* enrichit les données d'entrée avec du contexte additionnel pour clarifier les ambiguïtés et enrichir le *Tagger*, qui identifie et classe les entités du texte selon un schéma d'étiquetage (e.g., IOB [35]).

Augmentor. CL-KL utilise l'API Google Search comme source externe, sélectionnant des contextes via un modèle basé sur le BERTScore pour leur pertinence. Toutefois, cette méthode a ses contraintes : nécessité d'internet, coûts accrus pour des requêtes nombreuses, limites de l'API, et soucis de confidentialité avec la gestion de données sensibles par Google, posant un risque de fuite d'informations.

Tagger. Le *Tagger* classe les tokens en s'appuyant sur le contexte de l'*Augmentor*, avec un Transformer pré-entraîné et un classificateur linéaire, suivi d'un CRF pour optimiser l'étiquetage. CL-KL introduit un apprentissage coopératif pour compenser les coûts d'augmentation minimisant les divergences entre l'entrée originale et l'entrée augmentée via la divergence de Kullback-Leibler. Cela permet de maintenir les performances en l'absence d'augmentation. Nous invitons le lecteur à se référer à l'article original [1] pour une description plus détaillée.

4. Framework

CL-KL utilise un moteur de recherche via API comme *Augmentor*, limité par la confidentialité, la flexibilité et les coûts. Nous proposons un LLM comme *Augmentor* alternatif, permettant une augmentation hors ligne et un contrôle accru via des prompts, essentiels pour optimiser les performances. Selon [36] et [37], l'ingénierie des prompts, ou la conception de prompts pour obtenir des textes pertinents et imaginatifs, est cruciale bien qu'il n'existe pas de méthode universelle.

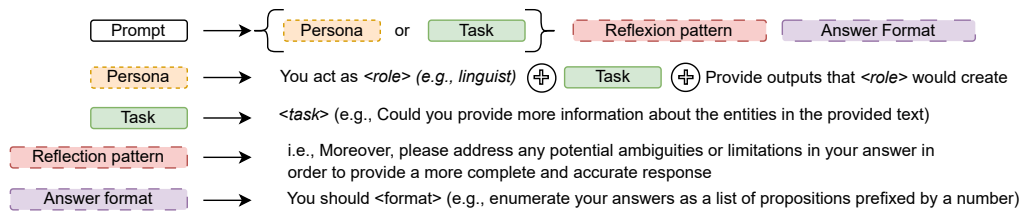


Figure 2: Modèle de création d'un prompt. Le rectangle vert représente le *Tâche* et les autres couleurs représentent les *Variations*. Les *Variations* sont facultatives.

Nous définissons le prompt par une *Tâche* et des *Variations*, où la *Tâche* est l'objectif du prompt et la *Variation*, facultative, enrichit sa sortie. Pour formaliser nos prompts, nous référons à la Figure 2, montrant comment chaque *Tâche* et *Variation* est structurée.

Pour élaborer un prompt efficace, nous posons les questions suivantes : 1. *Que demandons-nous ?* et 2. *Comment un LLM réagit-il à différentes formulations ?*

Des exemples de prompts sont présentés en *Annexe A*.

4.1. Que demandons-nous ?

La tâche cible est l'extraction d'entités, nécessitant des prompts pour contextualiser, reformuler, ou varier le contexte des entités. Chacune de ces approches vise à clarifier, enrichir ou désambiguïser les données d'entrée.

- **Prompt de contextualisation des entités.** Le but est de demander des informations supplémentaires sur les entités identifiées par le LLM dans l'entrée, en explorant leurs significations et les faits connexes.
- **Prompt de reformulation.** Ce prompt vise à changer les mots entourant les entités, en reformulant la phrase tout en conservant sa signification d'origine.
- **Prompt de contexte variation.** L'objectif est de générer des contextes variés dans lesquels les entités peuvent apparaître.

4.2. Comment un LLM réagit-il aux différentes formulations ?

Outre l'instruction principale, les variations du prompt peuvent dicter le style de la réponse, introduire un rôle spécifique, ou préciser le format de sortie. Nous utilisons cinq techniques de variation pour conditionner les réponses du LLM.

- **Classique :** Il s'agit de la variation de base informant uniquement sur la tâche.
- **Persona :** Il introduit un rôle dans le prompt. Cela peut influencer le LLM à se concentrer sur une partie spécifique de l'entrée liée à son rôle associé (linguiste, médecin, etc.) et/ou à conditionner le vocabulaire utilisé pour le processus de génération de la sortie.
- **Patron de réflexion :** Cette méthode met l'accent explicitement sur le fait d'inciter un LLM à analyser l'ambiguïté et à fournir une réponse claire.
- **Format de réponse :** Cette méthode fournit des informations sur le format de sortie que le LLM devrait adopter.
- **Tous :** Combinaison de toutes les variations précédentes.

Base de données	# label	Train	Dev	Test
WNUT17	6	3394	1009	1287
BC5CDR	2	4560	4581	4797
CoNLL ++	4	14987	3466	3466

Table 1

Statistiques de chaque jeux de données.

5. Protocole

Notre évaluation a pour objectif : 1) d'évaluer notre modèle pour la Reconnaissance d'Entités Nommées (NER) et 2) d'analyser la pertinence des contextes augmentés, décrits dans notre protocole d'évaluation.

5.1. Jeux de données

Notre modèle est évalué sur trois jeux de données NER : réseaux sociaux, biomédecine, et actualités, détaillées dans la Table 1.

- WNUT17 cible la détection d'entités rares dans les tweets.
- BC5CDR inclut des articles PubMed annotés sur les produits chimiques et les maladies.
- CoNLL++ est basé sur CoNLL03, et contient des articles d'actualités du corpus Reuters.

5.2. Baselines et mesures d'efficacité

Nos résultats sont comparés à CL-KL avec Google API, et à des approches état de l'art (Table 4), utilisant embeddings contextuels, architectures BiLSTM ou CNN, entraînement par ensemble, ou co-régularisation.

Les performances sont évaluées par extraction d'entités et le score F1 micro, avec une moyenne et un écart-type sur trois exécutions.

5.3. Mesures qualitatives

Nous évaluons la qualité des contextes générés sous deux aspects :

- **Utilisabilité** Nous catégorisons les sorties de la génération de contexte par les LLM en : Vide, Refusé, Échec, et Correct, basé sur leur format et contenu.
- **Pertinence sémantique** La pertinence sémantique est mesurée via BERTScore entre entrées et contextes générés, traitant les contextes "Vides" comme zéro et les "Refusé" et "Échec" comme peu pertinents. La différence de longueur est également considérée pour assurer la diversité.

5.4. Détails de la phase d'entraînement

Nous utilisons les mêmes paramètres que CL-KL, à savoir AdamW et des taux d'apprentissage spécifiques pour affiner les embeddings contextuels et la couche CRF. L'entraînement s'est fait sur des GPU NVidia v100/a100, totalisant environ 9500 heures.

Nous suivons les paramètres de CL-KL, utilisant l'optimiseur AdamW [38] avec une taille de lot de 4. Nous utilisons un taux d'apprentissage de $5 \cdot 10^{-6}$ pour affiner les embeddings contextuels ainsi qu'un taux d'apprentissage de 0.05 pour les paramètres de la couche CRF. Les

		O-CALM-Variation	Vide	Refusé	Echec	Correct
O-CALM-Tâche	CL-KL	-	202 (5.95%)	0 (0.00%)	0 (0.00%)	3192 (94.05%)
		Classique	214 (6.31%)	374 (11.02%)	441 (12.99%)	2365 (69.68%)
	Reformulation	Persona	215 (6.33%)	257 (7.57%)	262 (7.72%)	2660 (78.37%)
		Pattern de réflexion	209 (6.16%)	433 (12.76%)	216 (6.36%)	2536 (74.72%)
		Format de réponse	222 (6.54%)	350 (10.31%)	281 (8.28%)	2541 (74.87%)
		Tous	118 (3.48%)	310 (9.13%)	103 (3.03%)	2863 (84.35%)
	Contextualisation des entités	Classique	214 (6.31%)	313 (9.22%)	484 (14.26%)	2383 (70.21%)
		Persona	225 (6.63%)	222 (6.54%)	320 (9.43%)	2627 (77.40%)
		Pattern de réflexion	221 (6.51%)	328 (9.66%)	273 (8.04%)	2572 (75.78%)
		Format de réponse	239 (7.04%)	282 (8.31%)	406 (11.96%)	2467 (72.69%)
	Variation de contexte	Tous	134 (3.95%)	258 (7.60%)	109 (3.21%)	2893 (85.24%)
		Classique	237 (6.98%)	347 (10.22%)	415 (12.23%)	2395 (70.57%)
		Persona	221 (6.51%)	285 (8.40%)	256 (7.54%)	2632 (77.55%)
		Pattern de réflexion	209 (6.16%)	338 (9.96%)	215 (6.33%)	2632 (77.55%)
		Format de réponse	212 (6.25%)	372 (10.96%)	289 (8.52%)	2521 (74.28%)
		Tous	136 (4.01%)	292 (8.60%)	91 (2.68%)	2875 (84.71%)

Table 2

Analyse des contextes générés avec Llama2-7B [2] sur le jeu de données WNUT17 [19]. La colonne "tâche" représente la commande générale fournie au modèle linguistique. La colonne variation représente les variantes utilisées pour le conditionnement du format de sortie. Le contexte est ensuite classé en *Vide*, *Refusé*, *Echec* et *Correct*.

modèles NER sont entraînés pendant 10 époques pour chaque jeu de données. Nous utilisons XML-RoBERTa-Large comme contextualisation des jetons pour WNUT17/CoNLL++ et biobert-large-cased pour les jeux de données spécialisés comme BC5CDR. Llama2-7B est utilisé avec les paramètres par défaut. Les entraînements des modèles ont été réalisés sur des GPU Nvidia v100/a100 et a duré environ 9500 heures, y compris les phases de test et de production.

6. Résultats

		O-CALM-Variation	WNUT17		BC5CDR		CoNLL++	
			F1	BERTScore	F1	BERTScore	F1	BERTScore
CL-KL	From paper		<i>0.604</i>	-	<i>0.9099</i>	-	<i>0.9481</i>	-
	Our implementation		0.591 ± 0.027	0.7445	<u>0.9041 ± 0.002</u>	0.7934	0.9495 ± 0.0004	0.7312
Reformulation	Classic		0.577 ± 0.017	0.8029	0.893 ± 0.002	0.8396	0.957 ± 0.002	0.7643
	Persona		0.604 ± 0.007	0.8092	0.890 ± 0.004	0.8374	0.956 ± 0.002	0.7677
	Reflection Pattern		0.594 ± 0.006	0.8007	0.889 ± 0.002	0.8399	0.954 ± 0.002	0.7690
	Answer Format		0.593 ± 0.008	0.8036	0.893 ± 0.001	0.8422	0.956 ± 0.004	0.7664
	Tous		0.590 ± 0.002	0.8074	0.888 ± 0.004	0.8430	0.956 ± 0.001	0.7786
Entities contextualisation	Classic		0.601 ± 0.008	0.7942	0.893 ± 0.001	0.8143	0.956 ± 0.001	0.7605
	Persona		0.600 ± 0.005	0.7856	0.891 ± 0.001	0.8075	0.955 ± 0.002	0.7606
	Reflection Pattern		0.601 ± 0.002	0.7926	0.895 ± 0.003	0.8176	0.957 ± 0.001	0.7647
	Answer Format		0.602 ± 0.006	0.7961	0.893 ± 0.001	0.8258	0.955 ± 0.002	0.7640
Context variation	Tous		0.615 ± 0.003	0.7905	0.890 ± 0.000	0.8174	0.960 ± 0.002	0.7724
	Classic		0.596 ± 0.002	0.7912	0.895 ± 0.001	0.8202	0.955 ± 0.002	0.7636
	Persona		0.593 ± 0.008	0.7899	0.892 ± 0.002	0.8203	0.955 ± 0.001	0.7638
	Reflection Pattern		0.598 ± 0.011	0.7914	0.892 ± 0.000	0.8197	0.956 ± 0.001	0.7666
	Answer Format		0.596 ± 0.005	0.7926	0.892 ± 0.002	0.8277	0.955 ± 0.002	0.7636
	Tous		0.604 ± 0.002	0.7927	0.890 ± 0.000	0.8257	0.957 ± 0.003	0.7741

Table 3

Résultats des expériences menées sur WNUT17, BC5CDR et CoNLL++, en utilisant Llama2-7B. Pour chaque tâche, chaque variation est testée avec le F1 score. Nous ajoutons le BERTScore moyen entre le contexte et l'entrée. Les scores en gras sont nos meilleurs résultats et ceux soulignés sont les meilleurs dans l'ensemble.

6.1. Analyse de la qualité des contextes

Notre analyse se concentre sur la qualité des contextes générés, notamment sur WNUT17, où notre modèle atteint des performances état de l’art. Des observations similaires sont faites sur d’autres jeux de données, détaillées dans l’Annexe B.

Analyse des contextes générés. La Table 2 révèle une génération de contextes corrects par les LLM, avec un taux variant entre 69,68 % et 85,24 %. Les résultats varient selon le prompt, indiquant l’influence significative de la formulation des tâches sur la qualité du contexte généré. Les réponses *Vide* sont plus fréquentes dans la *variation du contexte* sans ajout de variations spécifiques, probablement due à la complexité de la tâche et à l’ambiguïté des instructions. Introduire une variation *Persona* diminue les réponses *Refusé*, et un *patron de réflexion* réduit efficacement les *Échecs*, soulignant l’importance de directives claires et ciblées. L’approche *Tous*, combinant toutes les variations, montre une amélioration notable, minimisant les problèmes tels que les contextes *Vide* et *Échec*.

Par rapport à CL-KL, notre méthode offre une diversité et une richesse de contextes plus importantes, malgré la haute proportion de réponses correctes observées avec CL-KL. La stratégie combinant toutes les possibilités s’avère être la plus efficace pour la génération de contexte de qualité.

Pertinence du Contexte. Notre modèle dépasse CL-KL en BERTScore, fournissant des contextes plus proches sémantiquement du texte d’entrée, sans se limiter à la reproduction du contenu. Les tâches de *Reformulation* affichent des BERTScores supérieurs, indiquant une capacité à choisir des mots sémantiquement proches. Le domaine spécialisé de BC5CDR montre des différences marquées entre les tâches, suggérant que des prompts adaptés pourraient améliorer les performances.

6.2. Résultats du Benchmark

Model	WNUT17	BC5CDR	CONLL++
[39]	58.9	-	-
[40]	58.5	-	-
[41]	57.41	-	-
[18]	-	91.9	-
[42]	-	91.3	-
[43]	-	90.89	-
[44]	-	-	95.088
[45]	-	-	94.28
[8]	-	-	94.04
CL-KL [1]	60.45	90.99	94.81
O-CALM (ours)	61.54	89.5	96.00

Table 4

Comparaison de la meilleure performance de notre modèle par rapport à différents modèles de références. À l’exception du jeu de données BC5CDR, notre approche est plus performante que les modèles précédents.

Model	Vide	Refusé	Echec	Correct
CL-KL	0.6086	0.5656	0.5891	0.6188
O-CALM	0.4571	0.5600	0.5891	0.6262

Table 5

F1 Score mesuré sur le jeu de données de test de WNUT17 selon chaque sous-catégorie définie dans la section 5.3. Le meilleur prompt trouvé (*Contextualisation des entités - Tous*) pour la tâche est utilisée pour le modèle O-CALM.

Nous évaluons l’efficacité de nos variantes de modèle et des modèles de référence sur divers jeux de données, avec des détails supplémentaires dans l’Annexe A. L’analyse des variantes (Table 3) indique qu’une combinaison de toutes les variantes (*Tous*) améliore significativement les performances, suggérant que l’intégration de celles-ci est cruciale pour l’efficacité de la NER. Une analyse plus fine montre que la *Reformulation* est moins performante sur WNUT17, soulignant l’importance d’informations contextuelles additionnelles.

La comparaison de notre meilleure variante avec d’autres modèles (Table 4) révèle un résultat à l’état de l’art sur WNUT17 et CoNLL++, validant l’efficacité de notre approche basée sur l’augmentation de contexte. Pour BC5CDR, une adaptation des prompts pourrait optimiser les résultats, indiquant la nécessité d’une approche ciblée pour des jeux de données spécifiques. Notre méthode surpasse CL-KL, avec une augmentation du score F1 sur WNUT17 et CoNLL++ et confirmant le bénéfice de générer des contextes pertinents hors ligne. 70% des cas fonctionnent avec notre approche de contextualisation (Table 2). Dans le 5, nous décrivons une analyse pour différencier les performances en mode *Correct* de celles en mode *Denied*, *Fail* ou *Empty*. Même si O-CALM semble bien s’adapter au mode *Denied & Fail* (qui correspondent à des exemples difficiles), ses performances sont affectées par les contextes vides.

7. Conclusion

Dans cet article, nous améliorons les performances de la Reconnaissance d’Entités Nommées (NER) en introduisant une technique d’augmentation d’échantillons à l’aide d’un contexte généré hors ligne par un gros modèle de langue (LLM) appelé O-CALM. En tirant parti de la puissance de génération de LLama2-7B, nous démontrons l’efficacité de notre approche en obtenant des résultats état de l’art sur deux jeux de données. De plus, nous fournissons une étude détaillée sur l’ingénierie de création des prompts, mettant en évidence la flexibilité de notre solution pour s’adapter à une multitude de scénarios tout en profitant des capacités hors ligne. Notre code est disponible sur GitHub¹. L’amélioration des performances dans des domaines techniques ciblés est la principale perspective associée à ce travail.

References

- [1] X. Wang, Y. Jiang, N. Bach, T. Wang, Z. Huang, F. Huang, K. Tu, Improving named entity recognition by external context retrieving and cooperative learning, arXiv preprint arXiv:2105.03654 (2021).

¹<https://github.com/Kawatami/O-CALM>

- [2] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, arXiv preprint arXiv:2307.09288 (2023).
- [3] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mistral 7b, 2023. arXiv:2310.06825.
- [4] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.
- [5] S. B. Huffman, Learning information extraction patterns from examples, in: International Joint Conference on Artificial Intelligence, Springer, 1995, pp. 246–260.
- [6] T. D. Singh, K. Nongmeikapam, A. Ekbal, S. Bandyopadhyay, Named entity recognition for manipuri using support vector machine, in: Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 2, 2009, pp. 811–818.
- [7] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, CoRR abs/1310.4546 (2013). URL: <http://arxiv.org/abs/1310.4546>. arXiv:1310.4546.
- [8] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, 2018. arXiv:1802.05365.
- [9] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
- [10] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: a robustly optimized bert pretraining approach (2019), arXiv preprint arXiv:1907.11692 364 (2019).
- [11] J. P. Chiu, E. Nichols, Named entity recognition with bidirectional lstm-cnns, Transactions of the association for computational linguistics 4 (2016) 357–370.
- [12] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, arXiv preprint arXiv:1603.01360 (2016).
- [13] M. Rei, Semi-supervised multitask learning for sequence labeling, CoRR abs/1704.07156 (2017). URL: <http://arxiv.org/abs/1704.07156>. arXiv:1704.07156.
- [14] Z. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging, CoRR abs/1508.01991 (2015). URL: <http://arxiv.org/abs/1508.01991>. arXiv:1508.01991.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
- [16] X. Wang, Y. Jiang, N. Bach, T. Wang, Z. Huang, F. Huang, K. Tu, Automated concatenation of embeddings for structured prediction, CoRR abs/2010.05006 (2020). URL: <https://arxiv.org/abs/2010.05006>. arXiv:2010.05006.
- [17] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, J. Li, Dice loss for data-imbalanced NLP tasks, CoRR abs/1911.02855 (2019). URL: <http://arxiv.org/abs/1911.02855>. arXiv:1911.02855.
- [18] S. Zhang, H. Cheng, J. Gao, H. Poon, Optimizing bi-encoder for named entity recognition via contrastive learning, 2023. arXiv:2208.14565.
- [19] L. Derczynski, E. Nichols, M. Van Erp, N. Limsopatham, Results of the wnut2017 shared task on novel and emerging entity recognition, in: Proceedings of the 3rd Workshop on Noisy User-generated Text, 2017, pp. 140–147.
- [20] I. Yamada, A. Asai, H. Shindo, H. Takeda, Y. Matsumoto, LUKE: Deep contextualized

- entity representations with entity-aware self-attention, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 6442–6454. URL: <https://aclanthology.org/2020.emnlp-main.523>. doi:10.18653/v1/2020.emnlp-main.523.
- [21] D. Seyler, T. Dembelova, L. Del Corro, J. Hoffart, G. Weikum, A study of the importance of external knowledge in the named entity recognition task, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 241–246. URL: <https://aclanthology.org/P18-2039>. doi:10.18653/v1/P18-2039.
- [22] S. Wang, X. Sun, X. Li, R. Ouyang, F. Wu, T. Zhang, J. Li, G. Wang, Gpt-ner: Named entity recognition via large language models, arXiv preprint arXiv:2304.10428 (2023).
- [23] S. Y. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi, T. Mitamura, E. Hovy, A survey of data augmentation approaches for nlp, arXiv preprint arXiv:2105.03075 (2021).
- [24] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, Smote: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002) 321–357.
- [25] A. Fernández, S. García, F. Herrera, N. Chawla, Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary, J. Artif. Intell. Res. 61 (2018) 863–905. URL: <https://api.semanticscholar.org/CorpusID:3373087>.
- [26] F. Charte, A. J. Rivera, M. J. del Jesus, F. Herrera, Mlsmote: Approaching imbalanced multilabel learning through synthetic instance generation, Knowledge-Based Systems 89 (2015) 385–397. URL: <https://www.sciencedirect.com/science/article/pii/S0950705115002737>. doi:<https://doi.org/10.1016/j.knosys.2015.07.019>.
- [27] J. Wei, K. Zou, EDA: Easy data augmentation techniques for boosting performance on text classification tasks, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6382–6388. URL: <https://aclanthology.org/D19-1670>. doi:10.18653/v1/D19-1670.
- [28] R. Jia, A. Raghunathan, K. Göksel, P. Liang, Certified robustness to adversarial word substitutions, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 4129–4142. URL: <https://aclanthology.org/D19-1423>. doi:10.18653/v1/D19-1423.
- [29] Y. Zhang, J. Baldridge, L. He, PAWS: Paraphrase adversaries from word scrambling, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 1298–1308. URL: <https://aclanthology.org/N19-1131>. doi:10.18653/v1/N19-1131.
- [30] D. Kang, T. Khot, A. Sabharwal, E. Hovy, AdvEntuRe: Adversarial training for textual entailment with knowledge-guided examples, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1:

- Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 2418–2428. URL: <https://aclanthology.org/P18-1225>. doi:10.18653/v1/P18-1225.
- [31] M. Glockner, V. Shwartz, Y. Goldberg, Breaking NLI systems with sentences that require simple lexical inferences, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 650–655. URL: <https://aclanthology.org/P18-2103>. doi:10.18653/v1/P18-2103.
- [32] Y. Belinkov, Y. Bisk, Synthetic and natural noise both break neural machine translation, arXiv preprint arXiv:1711.02173 (2017).
- [33] K. M. Yoo, D. Park, J. Kang, S.-W. Lee, W. Park, Gpt3mix: Leveraging large-scale language models for text augmentation, arXiv preprint arXiv:2104.08826 (2021).
- [34] H. Dai, Z. Liu, W. Liao, X. Huang, Y. Cao, Z. Wu, L. Zhao, S. Xu, W. Liu, N. Liu, S. Li, D. Zhu, H. Cai, L. Sun, Q. Li, D. Shen, T. Liu, X. Li, Auggpt: Leveraging chatgpt for text data augmentation, 2023. arXiv:2302.13007.
- [35] L. A. Ramshaw, M. P. Marcus, Text chunking using transformation-based learning, in: Natural language processing using very large corpora, 1999, pp. 157–176.
- [36] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing, ACM Computing Surveys 55 (2023) 1–35.
- [37] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, D. C. Schmidt, A prompt pattern catalog to enhance prompt engineering with chatgpt, arXiv preprint arXiv:2302.11382 (2023).
- [38] I. Loshchilov, F. Hutter, Fixing weight decay regularization in adam (2018).
- [39] M. Jeong, J. Kang, Regularizing models via pointwise mutual information for named entity recognition, CoRR abs/2104.07249 (2021). URL: <https://arxiv.org/abs/2104.07249>. arXiv:2104.07249.
- [40] A. Ushio, J. Camacho-Collados, T-ner: An all-round python library for transformer-based named entity recognition, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, 2021. URL: <http://dx.doi.org/10.18653/v1/2021.eacl-demos.7>. doi:10.18653/v1/2021.eacl-demos.7.
- [41] J. Hu, Y. Shen, Y. Liu, X. Wan, T.-H. Chang, Hero-gang neural model for named entity recognition, 2022. arXiv:2205.07177.
- [42] M. Jeong, J. Kang, Enhancing label consistency on document-level named entity recognition, 2022. arXiv:2210.12949.
- [43] V. Kocaman, D. Talby, Biomedical named entity recognition at scale, 2020. arXiv:2011.06315.
- [44] W. Zhou, M. Chen, Learning from noisy labels for entity-centric information extraction, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2021. URL: <http://dx.doi.org/10.18653/v1/2021.emnlp-main.437>. doi:10.18653/v1/2021.emnlp-main.437.
- [45] Z. Wang, J. Shang, L. Liu, L. Lu, J. Liu, J. Han, Crossweigh: Training named entity tagger from imperfect annotations, CoRR abs/1909.01441 (2019). URL: <http://arxiv.org/abs/1909.01441>. arXiv:1909.01441.

Tâche / Variation	Prompt
Contextualisation des entités / Classique	Could you provide more information about the entities in the provided text.
Reformulation / <i>Persona</i>	You act as an expert linguist, could you provide reformulations of the provided input text while keeping the same entities, you can provide extra information. Provide outputs that an expert linguist would create.
Variation du contexte / <i>Format de réponse</i>	Could you please present diverse situations in which the mentioned entities are encountered in the provided text. You should enumerate your answers as a list of propositions prefixed by a number.
Contextualisation des entités / <i>Pattern de réflexion</i>	Could you provide more information about the entities in the provided text. Moreover, please address any potential ambiguities or limitations in your answer in order to provide a more complete and accurate response.
Entités contextualisation / Tous	You act as an expert linguist, could you provide more information about the entities in the provided text. Provide outputs that an expert linguist would create. Moreover, please address any potential ambiguities or limitations in your answer in order to provide a more complete and accurate response.

Table 6

Exemples de définition des prompts. Une invite est définie par une paire de *Tâche* et *Variation*. *Tâche* (vert) peut être : *Contextualisation des entités*, *Reformulation* et *Variation du contexte*. *Variations* peut être : *Persona* (orange), *Pattern de réflexion* (rouge), *Format de réponse* (violet) et *Tous*.

A. Exemples de Prompts

Dans la Table 6, nous pouvons voir 5 exemples de prompts.

De plus, comme mentionné dans la section 6, le jeu de données BC5CDR montre une diminution du score F1 dans toutes les variations. L'hypothèse émise est que le prompt utilisé pour la génération de contexte est trop générique et peu adapté à ce type de domaine spécifique. À cet égard, nous proposons le prompt suivant en modifiant la variation de persona en "médecin" :

"Vous agissez en tant que **médecin**. Pourriez-vous fournir plus d'informations sur **le composé chimique et les maladies** dans le texte fourni ? Fournissez des résultats qu'**un médecin** pourrait créer. De plus, veuillez aborder toute ambiguïté potentielle ou limitation dans votre réponse afin de fournir une réponse plus complète et précise."

La Table 7 fournit des exemples de générations.

Entrée	Google API	LLama2-7B
<p>A high percentage of kanamycin - colistin and povidone - iodine irrigations were associated with erosive cystitis and suggested a possible complication with human usage .</p>	<p>Bladder irrigation with povidone-iodine in the prevention of urinary-tract infections after single or intermittent urethral catheterisation was investigated in a controlled study . In the control group (36 patients Because aging is accompanied by a steady decline in resistance to infectious diseases , the diagnosis and treatment of these diseases in the elderly is not only much more complex</p>	<p>As a doctor, I can provide more information on the chemical compounds mentioned in the text and their potential associations with diseases. However, I must also acknowledge the limitations of the information provided and the need for further research in some areas...Kanamycin and colistin are both antibiotics that are commonly used to treat bacterial infections. Kanamycin is a aminoglycoside antibiotic that is effective against a wide range of bacteria, including Gram-negative and Gram-positive bacteria. Colistin, on the other hand, is a polymyxin antibiotic that is effective against many types of Gram-negative bacteria, including some that are resistant to other antibiotics...Povidone-iodine is a topical antiseptic that is commonly used to disinfect the skin and mucous membranes. It is effective against a wide range of bacteria, viruses, and fungi, and is often used to prevent infections in surgical wounds, burns, and other skin injuries...The text suggests that a high percentage of kanam</p>

Entrée	Google API	LLama2-7B
<p>The possibilities that these cardiovascular findings might be the result of non - selective inhibition of monoamine oxidase or of amphetamine and metamphetamine are discussed .</p>	<p>Cardiovascular responses to combined treatment with selective monoamine oxidase type B inhibitors and L-DOPA in the rat Cardiovascular activity of rasagiline , a selective and potent inhibitor of mitochondrial monoamine oxidase B : comparison with selegiline Amphetamine and its derivatives exhibit a wide range of pharmacological activities , including psychostimulant , hallucinogenic , entactogenic , anorectic or antidepressant effects . The mechanisms of action underlying these Monoamine oxidase : isoforms and inhibitors in Parkinson 's disease and depressive illness Selegiline is a specific MAO-B inhibitor . As MAO-B has been shown to be significantly involved in the metabolism of dopamine in certain regions of the primate brain , selegiline Monoamine oxidase B (MAO-B) inhibitors have an established role in the treatment of Parkinsons disease as monotherapy or adjuvant to levodopa . Two m Inhibitors of MAO-A and MAO-B in Psychiatry and Neurology</p>	<p>As a doctor, I can provide you with more information on the chemical compounds and diseases mentioned in the text...Monoamine oxidase (MAO) is an enzyme that plays a crucial role in the degradation of certain neurotransmitters, including serotonin, dopamine, and norepinephrine. Non-selective inhibition of MAO can lead to an accumulation of these neurotransmitters in the brain, which can result in various adverse effects, including cardiovascular problems...Amphetamine and metamphetamine are both stimulants that can non-selectively inhibit MAO. Long-term use of these drugs can lead to increased levels of dopamine, norepinephrine, and serotonin in the brain, which can contribute to cardiovascular problems, such as hypertension, myocardial infarction, and arrhythmias...The text suggests that the cardiovascular findings in the patient might be the result of non-selective inhibition of MAO or of amphet</p>

Entrée	Google API	LLama2-7B
<p>The present study was undertaken to investigate the effects of TET and FAN on the experimental thrombosis induced by collagen plus epinephrine (EP) in mice , and platelet aggregation and blood coagulation in vitro .</p>	<p>Regulation of platelet activation and coagulation and its role in vascular injury and arterial thrombosis Action of Nanoparticles on Platelet Activation and Plasmatic Coagulation In Vitro impairment of whole blood coagulation and platelet function by hypertonic saline hydroxyethyl starch Mechanism Action of Platelets and Crucial Blood Coagulation Pathways in Hemostasis Request PDF In Vitro and Ex Vivo Tests of Coagulation and Platelet Function Thrombosis could be disposed by either local or global factors . Global factors are associated Whole-Blood Platelet Aggregation Predicts In Vitro and In Vivo Primary Hemostatic Function in the Elderly Arteriosclerosis , Thrombosis , and Vascular Biology Pentamethylquercetin (PMQ) reduces thrombus formation by inhibiting platelet function</p>	<p>As a doctor, I can provide more information on the chemical compounds and diseases mentioned in the text...TET and FAN are likely to be chemical compounds that have been synthesized for the purpose of studying their thrombolytic activity in vivo. Thrombosis is the formation of a blood clot within a blood vessel, which can lead to serious health problems such as heart attack, stroke, and pulmonary embolism. Inhibition of thrombosis refers to the ability of a compound to prevent or reduce the formation of blood clots...The text mentions that TET and FAN showed inhibition of thrombosis by 55% and 35%, respectively, in mice. This suggests that these compounds have a moderate to strong ability to prevent blood clots in mice. It is important to note, however, that the results of in vivo studies in mice may not always translate to the same effects in humans, and further studies are needed to determine the safety and efficacy of these compounds in humans...The text also mentions acetylsalicylic acid (ASA) as a positive control. ASA</p>

Table 7: Comparaison des contextes générés à partir de l'API de Google et de la méthode O-CALM.

B. Analyse des Résultats

Analyse de l'extraction des entités

Afin de comprendre le rôle du BERTScore dans les performances d'extraction, nous avons réalisé une étude sur l'ensemble de test de WNUT17. Nous avons collecté les extractions d'entités associées, plus spécifiquement les faux positifs et les faux négatifs. Le nombre de ces cas est ensuite réparti en groupes en fonction du BERTScore de leurs échantillons associés. La Figure 3 montre les distributions obtenues. Nous pouvons observer une réduction significative des cas de faux positifs à un BERTScore équivalent dans le cas de la génération grâce au LLM par rapport à celle de l'API Google. Cela pourrait être une indication d'une meilleure utilité du premier dans le contexte de l'extraction NER. Plus de faux négatifs sont observés dans les catégories de faible BERTScore, comme le montre la Figure 4.

Analyse de la longueur

La longueur du contexte généré a été mesurée et comparée au contexte de référence. La Figure 5 représente les deux distributions. Nous pouvons observer une grande variation de la longueur du contexte dans le cas d'un moteur de recherche, ce qui pourrait limiter les informations disponibles pour l'entraînement du *Tagger*. La génération de contexte par le LLM ne souffre pas de ce problème, car la longueur du contexte peut être influencée et maîtrisée, contrairement au moteur de recherche. Notez que le contrôle de la longueur du contexte peut être utile pour répondre aux exigences de certains modèles et équilibrer le compromis entre la longueur de génération et l'hallucination couramment observée dans les modèles de génération.

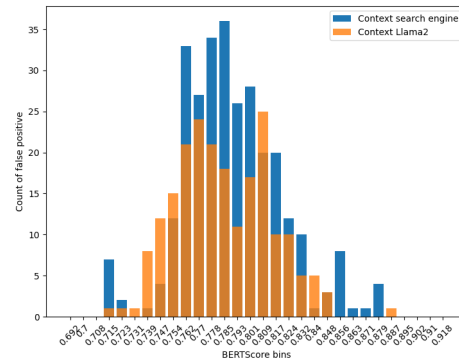


Figure 3: Taux de faux positifs en fonction du BERTScore associé.

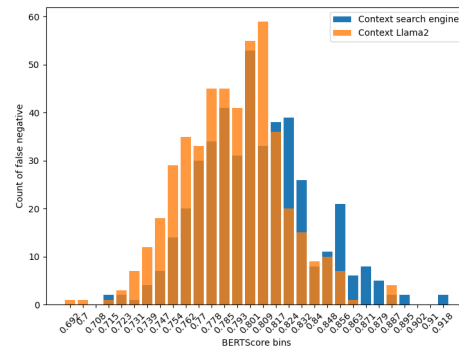


Figure 4: Taux de faux négatifs en fonction du BERTScore associé.

Analyse sans contexte

L'utilisation de LLM est coûteuse en termes de GPUs, rendant la solution peu pratique dans un environnement limité en ressources. Une solution proposée pour ce problème est l'utilisation de l'apprentissage coopératif, en contraignant les décisions de sortie à être proches les unes des autres en cas de contexte supplémentaire par rapport à l'entrée d'origine. La Table 8 compare les performances entre notre meilleur modèle et le modèle CL-KL. Globalement, nous pouvons observer une diminution des performances, mais elles restent proches avec les performances grâce au contexte dans le cas de WNUT17. Notre approche sans contexte

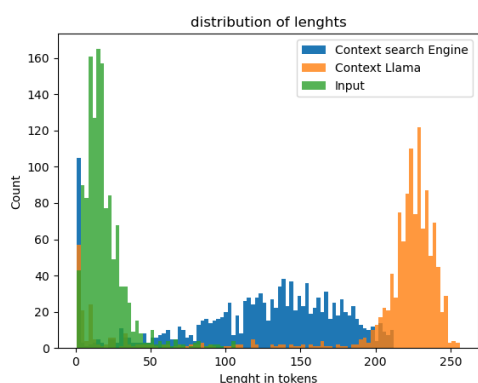


Figure 5: Distribution du nombre de mots présents dans les contextes de l'ensemble de test de WNUT17. Taille de l'entrée en vert, taille du contexte de l'API Google Search en bleu, et taille du contexte de Llama2-7B en orange.

Modèle	WNUT17	BC5CDR	CONLL++
Baseline AVEC CONTEXTE [1]	60.45	90.99	94.81
Baseline SANS CONTEXTE [1]	59.33	89.24	94.55
O-CALM AVEC CONTEXTE	61.15	89.5	96.00
O-CALM SANS CONTEXTE	60.13	88.97	95.90

Table 8

Comparaison des performances avec et sans contexte.

surpasse toujours le modèle CL-KL dans tous les cas, sauf pour les ensembles de données BC5CDR.

Analyse des prompts sur BC5CDR et CONLL++

La Table 9 et la Table 10 présentent les résultats de l'analyse des prompts pour les ensembles de données CoNLL++ et BC5CDR, respectivement.

Exemple de prompts générés

Voir Table 11.

Tâche	Variation	Vide	Refusé	Echec	Correct
CL-KL	-	1263 (8.43%)	0 (0.00%)	1521 (10.15%)	12203 (81.42%)
Reformulation	Classique	396 (2.64%)	279 (1.86%)	4795 (31.99%)	9517 (63.50%)
	Persona	439 (2.93%)	223 (1.49%)	4631 (30.90%)	9694 (64.68%)
	Pattern de réflexion	544 (3.63%)	274 (1.83%)	4717 (31.47%)	9452 (63.07%)
	Format de réponse	476 (3.18%)	212 (1.41%)	4852 (32.37%)	9447 (63.03%)
	Tous	580 (3.87%)	387 (2.58%)	4450 (29.69%)	9570 (63.86%)
Contextualisation d'entités	Classique	351 (2.34%)	151 (1.01%)	5014 (33.46%)	9471 (63.19%)
	Persona	435 (2.90%)	162 (1.08%)	4620 (30.83%)	9770 (65.19%)
	Pattern de réflexion	490 (3.27%)	249 (1.66%)	4771 (31.83%)	9477 (63.23%)
	Format de réponse	458 (3.06%)	178 (1.19%)	5275 (35.20%)	9076 (60.56%)
	Tous	610 (4.07%)	270 (1.80%)	4460 (29.76%)	9647 (64.37%)
Variation de contextes	Classique	332 (2.22%)	187 (1.25%)	4691 (31.30%)	9777 (65.24%)
	Persona	486 (3.24%)	168 (1.12%)	4555 (30.39%)	9778 (65.24%)
	Pattern de réflexion	494 (3.30%)	283 (1.89%)	4694 (31.32%)	9516 (63.50%)
	Format de réponse	485 (3.24%)	238 (1.59%)	4970 (33.16%)	9294 (62.01%)
	Tous	620 (4.14%)	428 (2.86%)	4398 (29.35%)	9541 (63.66%)

Table 9
Résultats de l'analyse des prompts menée sur CoNLL++

Tâche	Variation	Vide	Refusé	Echec	Correct
CL-KL	-	153 (3.36%)	0 (0.00%)	93 (2.04%)	4314 (94.61%)
Reformulation	Classique	94 (2.06%)	33 (0.72%)	575 (12.61%)	3858 (84.61%)
	Persona	90 (1.97%)	20 (0.44%)	443 (9.71%)	4007 (87.87%)
	Pattern de réflexion	86 (1.89%)	40 (0.88%)	397 (8.71%)	4037 (88.53%)
	Format de réponse	85 (1.86%)	20 (0.44%)	447 (9.80%)	4008 (87.89%)
	Tous	43 (0.94%)	31 (0.68%)	277 (6.07%)	4209 (92.30%)
Contextualisation d'entités	Classique	93 (2.04%)	27 (0.59%)	597 (13.09%)	3843 (84.28%)
	Persona	99 (2.17%)	12 (0.26%)	471 (10.33%)	3978 (87.24%)
	Pattern de réflexion	99 (2.17%)	34 (0.75%)	461 (10.11%)	3966 (86.97%)
	Format de réponse	100 (2.19%)	14 (0.31%)	626 (13.73%)	3820 (83.77%)
	Tous	54 (1.18%)	19 (0.42%)	291 (6.38%)	4196 (92.02%)
Variation de contexte	Classique	76 (1.67%)	21 (0.46%)	539 (11.82%)	3924 (86.05%)
	Persona	84 (1.84%)	21 (0.46%)	408 (8.95%)	4047 (88.75%)
	Pattern de réflexion	94 (2.06%)	47 (1.03%)	380 (8.33%)	4039 (88.57%)
	Format de réponse	86 (1.89%)	23 (0.50%)	496 (10.88%)	3955 (86.73%)
	Tous	42 (0.92%)	34 (0.75%)	247 (5.42%)	4237 (92.92%)

Table 10
Résultats de l'analyse des prompts menée sur BC5CDR

Entrée	Google API	LLama2-7B
<p>Replying to another question , Bhamre said the jawans deployed at places such as Siachen Glacier are provided with the best-quality winter clothing.</p>	<p>New Delhi : The Army has provided latest Swiss winter gear including trousers , jackets and gloves to jawans deployed in places such as Siachen Glacier Comptroller and Auditor General said that there is a massive shortage and inadequate reserves of specialised winter clothing , snow goggles , multi-purpose boots for Army troops posted areas like “ Presently , the DRDO is working on four projects to develop new technologies for more accurate prediction of avalanches , ” he said in a written response . Experts project the high-altitude deployment to adversely impact the physical and psychological well-being of the jawans , as well as their weaponry . Subhash Bhamre told the Lok Sabha that DRDO has 56 high altitude observatories , 55 weather automatic stations and five avalanche mitigation centres in HP , J & K and The Defence Research Development Organisation (DRDO) is working on four projects to develop new technologies for more accurate prediction of avalanches , the gov The Defence Research Development Organisation (DRDO) is working on four projects to develop new technologies for more accurate prediction of avalanches , the gov</p>	<p>As an expert linguist , I can provide more information about the entities mentioned in the text . Here are some possible outputs that an expert linguist might create : . 1 . Entities : . * "Jawans"- This term refers to soldiers or troopers in the Indian Army . It is a Hindi word that is commonly used in India . . * "Siachen Glacier"- This is a geographical location in the Himalayas , located in the disputed Kashmir region . It is a glacier that is considered one of the highest battlefields in the world , with temperatures often dropping to - 60C . . * "Winter clothing "- This term refers to clothing designed to keep people warm in cold weather conditions . In the context of the text , it is likely that the best - quality winter clothing is being provided to soldiers deployed at Siachen Glacier to protect them from the extreme cold . . 2 . Ambiguities or Limitations : . * "Best - quality "- The term "best - quality" is subjective and can be interpreted in different ways . It is possible that the term is being used to imply that the winter clothing provided to</p>

Entrée	Google API	LLama2-7B
Amazon Echo / Dot on your network	<p>You can connect your Amazon Echo to a Wi-Fi network in minutes using the Amazon Alexa app on your phone or tablet . Amazon.com Help : Set Up Your Echo Dot Amazon.com Help : Echo Dot Support Amazon.com Help : Update the Wi-Fi Settings for Your Echo Device Here are simple instructions for reconnecting your Amazon Echo , or connecting the smart speaker to a new Wi-Fi network if you see the light has turned orange , and The Amazon Echo is the body and Alexa is the brains , but you 're going to need to connect to a Wi-Fi network in order to receive the full Amazon.com Help : Echo Device Is Having Wi-Fi Issues</p>	<p>As an expert linguist , I can provide you with information on the entities mentioned in the provided text . Here are the entities I identified : . 1 . Amazon Echo : This is a smart speaker developed by Amazon that can perform various tasks such as playing music , setting alarms , and controlling smart home devices . . 2 . Dot : This is a smaller version of the Amazon Echo , a compact smart speaker that can perform similar tasks as the Echo but in a more compact form factor . . 3 . Network : This refers to a system of interconnected devices that can communicate with each other to share resources and exchange data . In the context of the text , it is likely that the author is referring to the internet or a local network . . 4 . Smart home devices : These are devices that can be controlled using a smart speaker like the Amazon Echo or Dot , such as lights , thermostats , and security cameras . . 5 . Alarms : These are sounds or alerts that can be set on a device to notify the user of an event or condition , such as a timer or a security breach . . 6 . Music : This refers to audio content that can be played on a smart speaker like the Amazon</p>

Entrée	Google API	LLama2-7B
<p>Retweeting for @ BadlandsNPS . # ClimateChange is real ! # TheResistance # DefyTrump # TrumpLeaks # DworkinReport ... https://t.co/1b7qIhtqwH</p>	<p>Days after a social media ban on the National Park Service , one park began tweeting facts about climate change Several parks tweet messages highlighting climate change fears or apparently opposing immigration plans . A US national park posts a series of tweets about climate change science that are later removed . The national park offers a few facts about our role in driving climate change before the posts are removed from Twitter . A former employee was responsible for the tweets In today ' s pass notes : the Twitter feed of the South Dakota park defied the Trump administration by posting facts about global warming . Was it an ex-employee Sometimes dissent takes surprising forms . The tweets — seen as a veiled shot at Donald Trump — were soon deleted .</p>	<p>As an expert linguist , I can provide more information about the entities mentioned in the provided text . Here are some potential entities and their possible meanings : . 1 . @ BadlandsNPS : This entity is likely a Twitter handle belonging to the Badlands National Park in South Dakota , USA . The tweet is likely a message from the park ' s official account . . 2 . # ClimateChange : This entity is a hashtag used to categorize tweets related to climate change . It is a popular hashtag used by individuals and organizations to raise awareness about the issue and to advocate for action to address it . . 3 . # TheResistance : This entity is a hashtag used to categorize tweets related to the political resistance movement against the Trump administration in the United States . The movement emerged in response to the election of Donald Trump as President in 2016 and has continued to grow and evolve since then . . 4 . # DefyTrump : This entity is a hashtag used to categorize tweets that are critical of or opposed to the policies and actions of Donald Trump , the 45th President of the United States . . 5 . # TrumpLe</p>

Table 11: Comparaison des contextes générés pour l'ensemble de tests WNUT17.