

Appel à propositions : Traitement automatique du langage naturel (TALN) – 2024

Questions et réponses

Candidature	
Q1	Y a-t-il une limite au nombre de candidatures qu'une même institution peut soumettre ?
R1	Il n'y a pas de limite au nombre de propositions qu'une même institution peut soumettre.
Q2	Existe-t-il un format spécifique de présentation de la proposition que nous devrions suivre ?
R2	Conformément à la section 4 (Informations sur les propositions) du document Appel à propositions : Traitement automatique du langage naturel (TALN) 2024 du Lacuna Fund, il existe effectivement un format spécifique. L'exposé de la proposition devrait être limité à 10 pages, références non incluses, avec des marges de 2,5 cm et une police de caractères de 11 points minimum. L'exposé doit porter sur les qualifications, la définition du problème et la solution proposée, les spécifications et les livrables pour les données et la documentation proposées, les bénéficiaires visés et les cas d'utilisation, la méthodologie et d'autres détails pertinents. Remplissez les informations relatives à votre proposition, y compris les questions relatives à la candidature et les modèles de budget et de livrables sur le portail de candidature SurveyMonkey Apply.
Q3	Quand et comment les propositions retenues seront-elles annoncées ? L'annonce sera-t-elle publique ou seules les équipes lauréates en seront-elles informées ?
R3	Les propositions retenues seront annoncées au printemps 2025. Les équipes lauréates seront informées et, une fois les contrats exécutés, les bénéficiaires seront annoncés publiquement dans la lettre d'information du Lacuna Fund et affichés sur le site web. Pour vous abonner à la lettre d'information, allez au bas de la page d'accueil de lacunafund.org et saisissez vos coordonnées.
Q4	Le portugais est-il une langue éligible ?

R4	<p>Nous acceptons les candidatures en portugais. Pour les candidats soumettant une proposition en portugais, vous pouvez utiliser les portails anglais, espagnol ou français. Nous n'avons pas de portail portugais disponible pour le moment. Toutefois, les propositions soumises en portugais via n'importe quel portail sont acceptées et seront examinées.</p>
Q5	<p>Dans la section 4 « Informations sur les propositions » et la sous-section « Informations sur le candidat », pourriez-vous clarifier les points 5 et 6 ? En ce qui concerne le point 5 en particulier, l'intention est-elle que chaque organisation participante soumette son propre code de conduite/d'éthique ou que chacune soumette le code de conduite établi par le comité d'éthique de son pays hôte ?</p>
R5	<p>Le point 5 fait référence aux informations sur les processus d'examen éthique de l'institution ou des institutions affiliée(s) demandées dans la section « Exposé de la proposition » intitulée « Risques, y compris l'éthique et la confidentialité ». L'objectif est, en tant que candidat, d'« identifier les problèmes et risques potentiels, y compris, mais sans s'y limiter, les problèmes potentiels de confidentialité et d'éthique, et de décrire les mesures que vous prendrez pour les atténuer ». Vous pouvez faire référence au code de conduite/d'éthique de l'organisation chef de file dans le cadre de votre réponse.</p> <p>Le point 6 fait référence à la capacité de l'équipe à obtenir des approbations nationales. Veuillez décrire vos plans pour obtenir l'approbation de votre projet par le comité d'éthique du pays hôte. Si vous n'avez pas l'intention de demander des approbations nationales dans l'un des pays où la recherche sera menée, veuillez expliquer en quoi le projet que vous proposez ne correspond pas à la définition de la recherche donnée par l'autorité réglementaire compétente. Pour plus d'informations, voir les questions relatives aux candidatures disponibles sur la page Soumettre un projet et sur le portail de candidature SurveyMonkey Apply.</p>
Q6	<p>En dehors de l'appel en cours, le Lacuna Fund peut-il accompagner une structure à équiper une ou des langues en ressources numériques ? Si oui, comment procéder ?</p>
R6	<p>Pour l'instant, le Lacuna Fund ne soutient les propositions que dans le cadre d'une procédure d'appel à propositions. Les prochains appels à propositions seront disponibles sur la page Soumettre un projet et seront annoncés dans la lettre d'information (vous pouvez vous abonner en défilant vers le bas de la page d'accueil à l'adresse https://lacunafund.org/).</p>

	<p>Nous proposons également (en anglais uniquement) des Ressources générales et des Ressources linguistiques sur notre site web, qui sont accessibles à tous. Voir la page Ensemble de données pour consulter les ensembles de données dans les domaines de l'agriculture, des langues et de la santé qui ont été créés avec le soutien du Lacuna Fund.</p>
Éligibilité	
Q7	<p>Je souhaite solliciter le soutien du Lacuna Fund pour développer un ensemble de données sur les images de maladies oculaires. Dans la mesure où il s'agit d'un appel à propositions sur le TALN, suis-je néanmoins éligible pour un financement puisque mon ensemble de données serait principalement composé d'images, étant donné qu'il s'agit d'un défi que j'ai rencontré en raison du peu ou de l'absence d'ensembles de données dans ce domaine ?</p>
R7	<p>Bien que l'appel à propositions soit axé sur le TALN, il reconnaît la nécessité de disposer d'ensembles de données dans des domaines connexes susceptibles de soutenir indirectement les technologies de TALN. Il s'agit avant tout d'expliquer comment l'ensemble de données que vous proposez sur les images de maladies oculaires pourrait contribuer ou compléter la recherche ou les applications relatives au TALN, en particulier dans le contexte des environnements à faibles ressources.</p> <p>Pour déterminer votre éligibilité et la pertinence de votre proposition, nous vous encourageons à lire attentivement la section 3.2.1 et à examiner comment votre projet s'aligne sur les objectifs et les critères décrits dans le document d'appel à propositions. Vous pouvez également envisager des partenariats potentiels ou des approches interdisciplinaires susceptibles de renforcer votre proposition.</p>
Q8	<p>Nous envisageons des partenariats dans d'autres pays de la région. Cependant, notre organisation ne peut pas effectuer de paiements en dehors de l'Argentine. Serait-il possible d'avoir deux institutions principales, ou serait-il possible que vous versiez des fonds directement à plus d'une institution ? Dans le même ordre d'idées, serait-il possible d'avoir deux chefs de projet ou un co-chef de projet ?</p>
R8	<p>Comme précisé dans l'appel à propositions, les partenariats sont fortement encouragés afin de renforcer la collaboration et de maximiser les avantages découlant de l'utilisation des ensembles de données, mais seul le candidat principal recevra des fonds. Il n'est pas possible que plusieurs institutions d'une même équipe reçoivent des fonds directement du Lacuna Fund. Si vous êtes en mesure de trouver un sponsor ou un partenaire financier capable d'effectuer des paiements en dehors du pays, il peut se porter</p>

	<p>candidat en tant que candidat principal ou institution principale, pour autant qu'il ait son siège dans la zone géographique visée par le présent appel (c'est-à-dire l'Afrique ou l'Amérique latine).</p>
Q9	<p>Pouvez-vous confirmer qu'une université en Colombie peut être une entité éligible au financement, étant donné qu'il s'agit d'un établissement d'enseignement supérieur ?</p>
R9	<p>Comme indiqué dans l'appel à propositions, dans la section « Critères d'éligibilité applicables aux organisations », « Pour pouvoir bénéficier d'un financement, les organisations doivent remplir les critères suivants : être une entité à but non lucratif, un institut de recherche, une entreprise sociale à but lucratif ou une équipe composée de ce type d'organisations ; [...] avoir une mission de soutien du bien sociétal, au sens large ; »</p> <p>Les universités peuvent participer à cet appel à propositions, ainsi que d'autres organisations dont la mission est de soutenir le bien social, au sens large (y compris les entités à but non lucratif, les instituts de recherche, les entreprises sociales à but lucratif, ou une équipe composée de ce type d'organisations).</p>
Q10	<p>Notre organisation est l'institution principale, mais des universités seront également impliquées. Un document attestant d'un accord formel est-il nécessaire en vue de soumettre la proposition ?</p>
R10	<p>Aucun accord formel n'est requis pour soumettre la proposition, à l'exception des réponses aux questions de la candidature et des certifications figurant sur le portail de candidature. Les lettres de soutien sont facultatives. Toutefois, si votre projet est sélectionné, des procédures formelles de diligence raisonnable et d'autres documents contractuels seront nécessaires.</p>
<p>Partenaires</p>	
Q11	<p>Encouragez-vous les partenariats entre les institutions universitaires et les start-ups ?</p>
R11	<p>Comme précisé dans l'appel à propositions, les partenariats sont fortement encouragés afin de renforcer la collaboration et de maximiser les avantages découlant de l'utilisation des ensembles de données. Les institutions universitaires et les start-ups peuvent poser leur candidature si elles ont une mission de soutien du bien sociétal, au sens large.</p>
Q12	<p>Les partenaires commerciaux sont-ils explicitement interdits ? Est-il possible d'affecter une partie du budget à l'achat de services professionnels si nécessaire ?</p>

	<p>Lors du webinaire, il a été souligné que seules les organisations à but non lucratif peuvent poser leur candidature. Cela exclut-il explicitement les organisations commerciales en tant que partenaires ? Je pense en particulier aux entreprises qui pourraient aider à trouver/rémunérer un grand nombre de micro-tâches. Un budget peut-il être consacré à l'achat de services professionnels en cas de besoin ?</p>
R12	<p>Comme indiqué dans l'appel à propositions, dans la section « Critères d'éligibilité applicables aux organisations », « Pour pouvoir bénéficier d'un financement, les organisations doivent remplir les critères suivants : être une entité à but non lucratif, un institut de recherche, une entreprise sociale à but lucratif ou une équipe composée de ce type d'organisations ; [...] avoir une mission de soutien du bien sociétal, au sens large ; » Les partenaires commerciaux sont les bienvenus. Pour cet appel à propositions, les entreprises commerciales/start-ups ayant une mission de soutien du bien sociétal peuvent se porter candidates. Une partie du budget peut être utilisée pour l'achat de services professionnels.</p>
Q13	<p>Si une organisation extérieure à la région éligible pose sa candidature en partenariat avec une institution de la région, qui est considéré comme le candidat principal et le bénéficiaire des fonds ?</p>
R13	<p>Le candidat principal doit être situé dans la région et serait le bénéficiaire des fonds (et pourrait ensuite les distribuer). Les partenaires en dehors de la région cible sont autorisés. Conformément à la section « Critères d'éligibilité applicables aux organisations » du document Appel à propositions : Traitement automatique du langage naturel (TALN) 2024 du Lacuna Fund, les propositions doivent être dirigées par une institution ayant son siège social en Afrique ou en Amérique latine. Si une organisation située en dehors de ces régions pose sa candidature en partenariat avec une institution de la région, l'institution de la région sera considérée comme le candidat principal et le bénéficiaire des fonds. Le candidat principal doit être issu d'une institution éligible d'Afrique ou d'Amérique latine afin de garantir l'alignement sur la cible géographique et les objectifs de la subvention. Les partenariats avec des organisations situées en dehors de ces régions peuvent renforcer les propositions, mais ne modifient pas le statut du candidat principal.</p>
<h2>Budget</h2>	
Q14	<p>Quel est le montant maximal pouvant être sollicité ?</p>
R14	<p>Conformément à la section « Budget de l'appel à propositions » (sous la section 4), le montant total disponible est d'environ 1 million de dollars</p>

	<p>(USD). Nous aimerions financer des projets dans chacune des régions ciblées (Afrique et Amérique latine) et nous pensons pouvoir soutenir 6 à 8 petits projets avec des budgets jusqu'à 100 000 USD et 2 à 3 projets plus gros et plus complexes avec des budgets de l'ordre de 100 000 à 250 000 USD. Le Groupe consultatif technique évaluera la faisabilité et la pertinence du budget ainsi que le lien entre le budget et la description de la subvention dans le cadre des critères de sélection.</p> <p>L'appel à propositions n'indique pas le montant maximal de financement qu'un projet peut recevoir. Veuillez solliciter le montant dont vous pensez avoir besoin pour créer l'ensemble de données que vous proposez ; nous vous encourageons toutefois à garder ces fourchettes à l'esprit afin d'augmenter vos chances de recevoir un financement.</p>
Q15	Existe-t-il un modèle officiel pour préparer le budget ?
R15	<p>Oui, vous trouverez des liens vers le modèle de budget officiel et le modèle sur les livrables à fournir sur le site web et sur le portail de candidature (TALN 2024) SurveyMonkey Apply :</p> <ul style="list-style-type: none"> • Modèle de budget (anglais) • Modèle sur les livrables à fournir (anglais) <p>Les traductions de ces modèles sont disponibles sur le portail de candidature SMA. Pour accéder à ces modèles, basculez le portail dans la langue de votre choix.</p> <p><i>Remarque :</i> Nous acceptons les candidatures en portugais. Pour les candidats soumettant une proposition en portugais, vous pouvez utiliser les portails anglais, espagnol ou français. Nous n'avons pas de portail portugais disponible pour le moment. Toutefois, les propositions soumises en portugais via n'importe quel portail sont acceptées et seront examinées.</p>
Q16	Quelles sont les exigences et les méthodes prévues pour le transfert des fonds ?
R16	<p>Les paiements sont effectués par transfert électronique de fonds (TEF) ou par virement bancaire à l'aide d'un formulaire d'autorisation de dépôt direct et d'un formulaire de virement bancaire international.</p> <p>Pour les équipes sélectionnées comme finalistes pour recevoir des fonds, l'organisation principale doit soumettre ses documents de diligence raisonnable pour examen avant d'être approuvée pour l'attribution des fonds. Nous vous demandons de procéder à une vérification préalable de toutes les organisations partenaires qui recevront des subventions.</p>

Exigences en matière de spécificité et de portée des ensembles de données

Langues

Q17

Combien de langues faut-il prendre en considération au minimum ou au maximum par projet ?

Qu'en est-il si notre proposition ne concerne qu'une seule langue africaine ? Est-elle tout de même éligible ?

R17

Conformément au document de l'appel à propositions TALN 2024 du Lacuna Fund, il n'y a pas de mention spécifique d'un nombre minimal ou maximal de langues à prendre en considération par projet. Toutefois, les propositions doivent se concentrer sur les défis à relever dans les langues et cultures à faibles ressources en Afrique et en Amérique latine.

Malgré l'absence d'exigence stricte quant au nombre de langues (ce qui implique qu'une proposition spécifique à une seule langue pourrait être éligible), les candidats doivent justifier le choix de la ou des variantes des langues et leur incidence sur les cultures sous-représentées en fonction de l'impact potentiel et de la pertinence par rapport aux objectifs de l'appel à propositions.

Le projet peut se concentrer sur une seule langue et inclure différentes variantes de la langue en fonction de la tâche à accomplir. Par exemple, si le projet est axé sur la création d'ensembles de données relatifs à des discours haineux spécifiques à une culture, il pourrait inclure différentes variantes des langues utilisées par différents groupes sociaux.

Q18

Étant donné que l'ensemble de données de Google Translation contient 785 millions d'enregistrements couvrant 548 langues, je pense qu'il serait utile de clarifier davantage les langues africaines et latino-américaines qui sont sous-représentées.

Le portugais est-il une langue éligible ?

Quelles sont les caractéristiques d'une langue éligible pour un projet ?

R18

Cet appel à propositions porte sur les « langues et cultures à faibles ressources en Afrique et en Amérique latine ». Il existe des milliers de langues africaines, qui peuvent toutes être considérées comme sous-représentées. De même, les langues latino-américaines, en particulier les langues indigènes, peuvent également être considérées comme telles. Nous nous concentrons tout particulièrement sur les langues disposant de peu de ressources.

	<p>Toutefois, des langues aux ressources plus importantes, telles que le portugais, pourraient être incluses si l'ensemble de données est susceptible d'avoir un impact sur une culture sous-représentée dans la technologie de TALN. Cela peut être le cas, par exemple, dans un contexte tel que l'alternance codique avec une communauté d'Amérique latine de langue/discours différent, comme les immigrants pauvres piémontais, ou les créoles portugais (c'est-à-dire le créole cap-verdien, le créole bissau-guinéen, le créole des Antilles néerlandaises ou papiamentu). Un autre exemple est lorsque la variante de la langue comprend un argot ou un vocabulaire péjoratif propre à une région, comme c'est souvent le cas dans les ensembles de données d'alignement.</p> <p>Nous avons conscience que c'est vous qui connaissez le mieux vos communautés et que vous savez quelles langues disposent de ressources ou ont besoin de ressources. Prenons l'exemple d'une langue supposée avoir de nombreuses ressources mais qui présenterait de faibles ressources dans un domaine spécifique, tel que la réponse à des questions sur des images pour des personnes aveugles dans un pays d'Amérique latine. Dans ce contexte, les personnes aveugles prennent des photos et posent des questions sur les produits qu'elles utilisent quotidiennement, qui pourraient ne pas être reconnues par les modèles linguistiques multimodaux les plus modernes, car les marques, les emballages et d'autres détails n'existent pas dans l'hémisphère nord.</p> <p>Si le français, l'espagnol et le portugais sont mieux représentés, il peut encore y avoir des cas (par exemple dans des domaines tels que la violence numérique et la santé) où les technologies linguistiques manquent de ressources, selon la situation. Ainsi, une langue peut être considérée comme manquant de ressources même si elle est parlée par des centaines de millions de personnes.</p> <p>Pour plus d'informations, voir les ressources linguistiques sur le site du Lacuna Fund. https://lacunafund.org/language-resources/.</p>
Q19	Les livres sacrés et les livres de langue peuvent-ils être classés comme des ensembles de données existants en Afrique et en Amérique latine ?
R19	Conformément au document Appel à propositions : Traitement automatique du langage naturel (TALN) 2024, les ensembles de données existants sont définis comme ceux qui sont librement accessibles et qui peuvent soutenir le développement de modèles d'apprentissage machine pour les langues et les cultures à faibles ressources en Afrique et en Amérique latine. Les livres sacrés et les livres de langue peuvent en effet être considérés comme des ensembles de données existants s'ils satisfont à ces critères.

	<p>Ces textes peuvent être considérés comme des ensembles de données existants pour certaines langues. Par ailleurs, il existe des langues à faibles ressources que les communautés tentent de revitaliser et qui ne disposent que d'enregistrements réalisés par des personnalités telles que des prêtres, à des fins d'évangélisation. Cependant, ce type de documentation nécessite des méthodes d'interprétation rigoureuses afin de réutiliser ces enregistrements et ces analyses au service du TALN et d'autres applications dépassant leur contexte religieux d'origine. En outre, il est important de noter que les textes religieux doivent être utilisés avec une extrême prudence, car ils peuvent comporter des risques éthiques importants.</p> <p>Pour plus d'informations, voir les ressources linguistiques sur le site du Lacuna Fund. https://lacunafund.org/language-resources/.</p>
Considérations techniques	
Q20	Pour les ensembles de données vocales, existe-t-il un nombre minimal d'heures d'enregistrement ?
R20	<p>Conformément au document de l'appel à propositions TALN 2024 du Lacuna Fund, il n'y a aucune mention spécifique d'un nombre minimal d'heures d'enregistrement requis pour les ensembles de données vocales. Toutefois, les propositions doivent fournir une justification suffisante de l'adéquation de l'ensemble de données pour soutenir les technologies de traitement automatique du langage naturel (TALN) pour les langues et les cultures à faibles ressources en Afrique et en Amérique latine.</p> <p>Les propositions doivent démontrer l'impact potentiel de l'ensemble de données sur l'amélioration des modèles d'apprentissage machine et la résolution des problèmes spécifiques au traitement de la parole.</p> <p>N'hésitez pas à consulter les précédents ensembles de données soutenus par le Lacuna Fund ici : https://lacunafund.org/fr/ensemble-de-donnees/.</p>
Q21	Y a-t-il des logiciels spécifiques pour étiqueter les données ?
R21	<p>Le document de l'appel à propositions TALN 2024 n'impose pas l'utilisation d'un logiciel spécifique pour l'étiquetage des données. Toutefois, il souligne l'importance d'utiliser des outils robustes et fiables pour garantir la qualité et la précision des ensembles de données étiquetées. Le choix du logiciel doit s'aligner sur les objectifs du projet et les exigences spécifiques de l'ensemble de données en cours d'élaboration, tout en respectant les principes des données FAIR.</p> <p>Si vous utilisez des grands modèles de langage (LLM) ou d'autres modèles de fondation à un stade quelconque de votre annotation, cette information doit être divulguée et précisée dans votre méthodologie. Si vous utilisez</p>

	des plateformes de travail collaboratif à distance, vous devez respecter les pratiques éthiques consistant à payer un salaire équitable et à demander un consentement éclairé, comme vous le feriez pour les méthodes de collecte en personne.
Q22	L'hébergement de l'ensemble de données produit peut-il également se faire sur les plateformes publiques telles que HuggingFace, Kaggle, etc. ?
R22	<p>Conformément à la section « Exposé de la proposition » de l'appel à propositions, le Lacuna Fund insiste sur l'importance de rendre les ensembles de données librement accessibles à l'ensemble de la communauté des chercheurs afin d'en maximiser l'impact. Les plateformes d'hébergement doivent attribuer un identifiant numérique d'objet (DOI) à l'ensemble de données, quantifier les téléchargements de l'ensemble de données et collecter les informations de contact pour les téléchargements de l'ensemble de données. Pour obtenir des orientations sur les plateformes d'hébergement suggérées, voir le document Dataset Hosting and Documentation Guidance [Orientations en matière d'hébergement des données et de documentation] du Lacuna Fund.</p> <p>Bien que l'appel à propositions ne mentionne pas spécifiquement des plateformes telles que HuggingFace ou Kaggle, il encourage l'utilisation de plateformes réputées et largement répandues afin de garantir une vaste accessibilité et une grande facilité d'utilisation. L'hébergement sur HuggingFace est judicieux si l'ensemble de données ne contient pas d'informations confidentielles ou sensibles. Il est essentiel d'expliquer comment le consentement éclairé sera recueilli pour toutes les données. Cela est particulièrement important pour les données qui peuvent être culturellement sensibles, et la candidature doit aborder les mesures, telles que l'anonymisation, qui peuvent être prises pour atténuer le risque. Il est important de noter que ce qui ne constitue pas une information sensible dans une culture peut l'être dans une autre.</p>
Qualité des données	
Q23	<p>Pouvez-vous clarifier la question de la fourniture d'un échantillon de modèle et de données avant la soumission ; devrions-nous créer un échantillon avant le début du contrat ?</p> <p>Une phase pilote est-elle prévue dans le projet ?</p>
R23	Le projet peut inclure des expériences de référence ainsi qu'un échantillon de modèle à des fins d'évaluation. Il peut s'agir de tester des modèles SOTA (State of the Art) sur votre ensemble de données et vos tâches, comme cela se fait couramment dans le domaine du TALN. Il n'est pas nécessaire de fournir un modèle optimisé, il suffit de démontrer son utilité et sa facilité d'utilisation avec vos données et pour votre tâche. La méthode

	utilisée et les coûts associés à cette petite phase pilote doivent être inclus dans votre proposition, car il n’y aura pas d’autre appel à ce sujet.
Q24	Quelles sont les erreurs les plus courantes que vous avez constatées dans les réponses aux appels à propositions et qui compromettent leurs chances de succès ?
R24	<p>Veillez à ce que votre projet réponde à tous les critères d’éligibilité et complétez toutes les sections de la proposition, en respectant chaque exigence de l’exposé (voir la section « Exposé de la proposition » de l’appel à propositions), afin que votre proposition ne soit pas jugée inéligible. Veillez à inclure des cas d’utilisation clairs et des plans sur la durabilité.</p> <p>Parmi les exemples de problèmes, citons les licences incorrectes, l’absence d’une méthodologie claire pour préserver la qualité des données ou l’absence de politiques d’hébergement bien définies (c’est-à-dire précisant l’endroit où les données seront stockées).</p>
Cas d’utilisation	
Q25	Quel est le niveau de spécificité requis pour expliquer comment l’ensemble de données peut aider à résoudre les cas d’utilisation ? Souhaitez-vous que cet aspect soit quantifié ? Dans l’affirmative, peut-il s’agir d’estimations générales ou vous faut-il une quantification minutieuse ?
R25	Conformément à la section « Exposé de la proposition » de l’appel à propositions TALN 2024 du Lacuna Fund, il est important d’expliquer clairement comment l’ensemble de données proposé permettra de relever des défis spécifiques en matière de traitement automatique du langage naturel. La section « Processus de sélection et critères d’évaluation » précise que nous recherchons des propositions qui démontrent une bonne compréhension du problème et de la manière dont l’ensemble de données contribuera à le résoudre. Nous serons ravis d’obtenir des estimations approximatives et, si possible, une quantification détaillée des impacts, des bénéfices et des risques attendus.
Q26	Étant donné que ces ensembles de données peuvent aider à résoudre toute une série de problèmes, quel est le nombre idéal de cas d’utilisation à mettre en évidence en tant que cibles potentielles pour l’ensemble de données ?
R26	Conformément au document de l’appel à propositions TALN 2024 du Lacuna Fund, il n’y a pas de mention spécifique d’un nombre idéal de cas d’utilisation à mettre en évidence pour l’ensemble de données. Toutefois, les candidats sont encouragés à se concentrer sur un nombre raisonnable de cas d’utilisation qui démontrent l’impact potentiel et la pertinence de l’ensemble de données pour relever les défis du traitement automatique

	<p>du langage naturel pour les langues et les cultures à faibles ressources en Afrique et en Amérique latine.</p> <p>L'appel à propositions insiste sur l'importance de la qualité plutôt que de la quantité dans la présentation des cas d'utilisation. Les propositions devront expliquer clairement comment l'ensemble de données contribuera à résoudre des problèmes spécifiques dans les applications de TALN, en donnant un aperçu détaillé des avantages et des résultats attendus.</p>
Politiques	
Q27	<p>L'appel à propositions semble dire que les données et les outils utilisés doivent être en source ouverte. Cela semble exclure l'utilisation d'outils qui ont été créés antérieurement et dont les propriétaires ne veulent pas les mettre en source ouverte ? Là encore, je pense à une entreprise partenaire qui propose l'utilisation d'une plateforme pour la collecte de données.</p> <p>Si les partenaires disposent d'outils qu'ils nous autorisent à utiliser, est-il acceptable qu'ils ne partagent pas la propriété intellectuelle de ces outils ?</p>
R27	<p>C'est l'ensemble de données lui-même qui est important. Vous pouvez utiliser des outils commerciaux pour le créer, mais l'ensemble de données lui-même, les métadonnées associées et l'ensemble des modèles ou publications élaborés grâce aux fonds de la subvention doivent être mis en source ouverte en utilisant l'une des licences énumérées dans notre politique en matière de propriété intellectuelle. Consultez notre Politique de propriété intellectuelle.</p>
Q28	<p>Existe-t-il des sources de données potentielles spécifiques qui ne devraient pas être utilisées pour des raisons de confidentialité ou d'éthique ?</p>
R28	<p>Les bénéficiaires doivent s'assurer que les ensembles de données sont élaborés dans le respect de l'éthique, par exemple en évitant tout matériel protégé par des droits d'auteur. En outre, si vous travaillez avec des données médicales, par exemple, celles-ci peuvent contenir des informations sensibles ou confidentielles. Vous devez faire preuve de prudence avec ce type de données et expliquer votre méthode d'anonymisation, le cas échéant. Pour de nombreuses communautés indigènes, il existe des types de connaissances, telles que les connaissances médicales ou religieuses, qu'il peut être judicieux d'archiver à des fins de préservation, mais qui ne doivent pas être rendues librement accessibles au grand public. Ce type de documentation peut être utile et même être utilisé pour le TALN, mais il doit être sauvegardé, et cette sauvegarde doit être précisée dans le projet. Si des informations sensibles liées au savoir autochtone ou à la religion sont incluses, les candidats devront demander une dérogation à la politique de propriété intellectuelle du Lacuna Fund en ce qui concerne la licence. Consultez notre Politique de propriété intellectuelle pour plus de détails sur les licences lorsque la confidentialité</p>

	<p>est en jeu. Veuillez également vous référer aux principes du Lacuna Fund en matière d'éthique (page 5 de l'appel à propositions) et aux critères d'évaluation en matière d'éthique à la page 5 de l'appel à propositions, ainsi qu'aux conseils en matière de gestion des données et licences, d'éthique et de confidentialité aux pages 11 et 12, section 4 « Exposé de la proposition » de l'appel à propositions. Vous pouvez également consulter nos ressources sur le TALN à l'adresse suivante : https://lacunafund.org/language-resources/</p>
<h2 style="color: green;">Mentorat</h2>	
Q29	Est-il obligatoire d'avoir un mentor ?
R29	Non, ce n'est pas une obligation. Il s'agit d'une possibilité gratuite qui vous est offerte.
Q30	<p>Quel est le processus pour bénéficier du mentorat ?</p> <p>En quoi consisterait ce mentorat de Masakhane ?</p>
R30	<p>Le principal avantage du mentorat est d'avoir un mentor qui examinera votre projet de proposition et discutera des moyens de l'améliorer. Les candidats au mentorat peuvent demander différentes formes d'assistance telles que : discuter des lacunes du TALN dans les langues à faibles ressources, rédiger une proposition de recherche et préparer des budgets.</p> <p>Pour plus d'informations, voir les enregistrements et les diapositives du webinaire destiné aux candidats disponibles sur le site web du Lacuna Fund à l'adresse suivante :</p> <ul style="list-style-type: none"> • Enregistrement (anglais) : https://vimeo.com/982223215 • Enregistrement (français) : https://vimeo.com/982223384 • Enregistrement (portugais) : https://vimeo.com/982223670 • Enregistrement (espagnol) : https://vimeo.com/982223853 <p>Diapositives :</p> <ul style="list-style-type: none"> • Anglais • Français • Portugais • Espagnol <p>Le programme de mentorat de Masakhane est disponible via ce Google Form ; la date limite initiale était le 15 juillet 2024.</p>
Q31	Le mentorat de Masakhane se déroule-t-il en anglais ou dans une autre langue ?

R31	<p>Le mentorat sera disponible dans chaque langue (anglais, français, espagnol, portugais). Lorsque vous postulez pour un mentorat, n'hésitez pas à indiquer la langue dans laquelle vous souhaitez être encadré, et l'équipe de Masakhane vous fera savoir si elle est en mesure de répondre à votre demande. Pour toute question, vous pouvez également contacter Masakhane à l'adresse : masakhane_leadership@googlegroups.com.</p>
-----	---