# Reconnaissance de mots spécifiques avec Whisper ASR : problèmes et solutions

## Contexte : Whisper et vocabulaire peu commun

Whisper (OpenAI) est très performant en transcription audio, mais comme beaucoup de modèles ASR il **peut échouer sur les mots peu communs, noms propres ou acronymes**. La documentation d’OpenAI souligne que c’est un défi connu : le modèle ne reconnaît souvent pas ces termes rares ou spécialisés[[1]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Improving%20Reliability). Par exemple, sans aide il a pu transcrire *“DALL·E”* en *“DALI”* ou *“GPT-3”* en *“GDP 3”*[[2]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=1,acronyms). Pour améliorer la fiabilité, Whisper offre un paramètre de **prompt** (souvent appelé initial\_prompt en local) permettant de **fournir à l’avance les mots ou orthographes attendus** afin de guider le modèle[[3]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Using%20the%20Prompt%20Parameter)[[4]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=audio_file%20%3D%20open%28,).

**Pourquoi le prompt est-il utile ?** Parce que Whisper a été entraîné à prédire la suite d’une transcription partielle. En lui donnant un contexte textuel contenant vos termes spécifiques, on crée un **biais positif** en faveur de ces mots : le modèle sera plus enclin à les utiliser s’il détecte un son approchant dans l’audio[[5]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=1). En somme, on injecte votre vocabulaire dans la « mémoire » du modèle avant la transcription, ce qui peut corriger des erreurs de reconnaissance (par exemple, inciter *“GDP 3”* à devenir *“GPT-3”* dans l’exemple OpenAI[[2]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=1,acronyms)).

## Bonne utilisation du paramètre initial\_prompt

Pour maximiser l’efficacité du prompt, voici les **bonnes pratiques** dégagées des retours de développeurs et de la documentation :

* **Formater un prompt concis et contextuel** : Plutôt qu’un seul mot isolé, il est conseillé d’énumérer plusieurs termes ou de les inclure dans une phrase cohérente. Par exemple, un contributeur suggère d’écrire une phrase du type *« Let's talk about International Monetary Fund and SDRs. »* pour que Whisper reconnaisse *“SDRs”*[[6]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=1). De même, OpenAI montre qu’une liste de termes séparés par des virgules peut servir de **glossaire** pour guider la transcription[[4]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=audio_file%20%3D%20open%28,). Par exemple : prompt="SWOOD, SolidWorks, EFICAD, nesting" (sans point final) fournit au modèle ces mots inconnus. Dans le code OpenAI, un prompt *"ZyntriQix, Digique Plus, ..."* a permis d’intégrer des noms de produits ésotériques dans la transcription[[4]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=audio_file%20%3D%20open%28,)[[7]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=prompt%3D,). Dans vos tests, n’hésitez pas à inclure le mot spécifique *au sein d’une phrase indicative*, voire avec des termes reliés, pour le mettre en contexte. Des exemples concrets montrent qu’une simple mention en contexte peut suffire : un développeur a réglé la transcription de *“USheep”* en fournissant le prompt *“Welcome to USheep.”*, ce qui a **encouragé le modèle à sortir exactement “USheep”** au lieu d’une interprétation erronée[[5]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=1).
* **Respecter la casse, l’orthographe et la ponctuation souhaitées** : Whisper a tendance à **imiter le style du prompt**. La documentation API note que le modèle suivra la capitalisation et la ponctuation du prompt[[8]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=You%20can%20use%20a%20prompt,can%20help%20in%20different%20scenarios)[[9]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=,Improving%20punctuation). Donc, écrivez vos termes exactement comme vous voulez qu’ils apparaissent. Par exemple, si le mot doit être en majuscules (un acronyme), mettez-le en majuscules dans le prompt. S’il s’agit d’un acronyme épelé, vous pouvez même le **styller avec des points ou majuscules** (ex. *“B.R.I.C.K.”* dans le prompt pour que la sortie contienne *B.R.I.C.K.*[[10]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=file%3Daudio_file%2C%20response_format%3D,)). Dans un cas, fournir *“Aimee, Shawn”* comme prompt a incité le modèle à orthographier correctement ces prénoms dans la transcription finale[[11]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,Aimee%2C%20Shawn) (au lieu de *Amy, Sean* par défaut). En somme, le prompt sert de *guide orthographique* que Whisper va s’efforcer de suivre.
* **Prompt ni trop long ni trop court** : Le prompt textuel est inséré dans le modèle comme contexte ; il est limité à **224 tokens environ** (équivalent à ~150-200 mots maximum)[[12]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=contained%20within)[[3]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Using%20the%20Prompt%20Parameter). Une liste de quelques mots ou une phrase courte suffit largement. **Évitez toutefois un prompt trop court d’un seul mot**, qui pourrait être ignoré. OpenAI note que *« les prompts très courts sont moins fiables »*[[13]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,president%20biden) – un seul mot risque de ne pas suffisamment ancrer le modèle, surtout s’il n’a aucun lien sémantique évident. Si un terme unique ne fonctionne pas, essayez de **l’associer à d’autres mots** liés dans le prompt pour donner plus de contexte. Par exemple, au lieu de juste "SWOOD", un prompt comme *"SWOOD plugin for SolidWorks by EFICAD"* pourrait mieux fonctionner en mettant le mot dans un contexte que le modèle peut interpréter.
* **Forcer la langue et d’autres paramètres** : Puisque votre audio est en anglais (même avec un accent français modéré), spécifiez language="en" dans model.transcribe. Cela évite que Whisper ne se trompe de langue ou n’essaie de traduire le texte détecté[[14]](https://community.openai.com/t/whisper-is-translating-my-audios-for-some-reason/86468#:~:text=API%20community,of%20the%20speaker%27s%20native%20language). D’après des utilisateurs, Whisper peut parfois basculer en mode traduction si l’accent semble non natif – fixer la langue empêche ce comportement. De plus, fixez temperature=0.0 pour rendre la transcription **déterministe** (pas de choix aléatoires) et donc plus cohérente avec le prompt fourni[[15]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=the,3). Enfin, utilisez condition\_on\_previous\_text=False pour **appliquer le prompt à chaque segment audio** indépendamment[[16]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=3,3). Par défaut Whisper utilise le texte déjà transcrit comme contexte pour la suite, ce qui peut diluer l’influence du prompt après le premier segment. En le désactivant (False), chaque nouvelle portion d’audio repart du prompt initial – vos mots-clés restent donc **pris en compte tout au long de la vidéo**, même s’ils ne figurent pas dans le tout début.

## Limitations connues de l’**initial\_prompt**

Malgré ces précautions, il est important de noter que **le prompt n’est pas une garantie absolue**. De nombreux utilisateurs et même des études récentes ont observé des comportements contre-intuitifs de Whisper vis-à-vis des prompts. Par exemple, une recherche de 2024 a montré que *Whisper n’“comprend” pas toujours le prompt textuel comme on l’attendrait humainement*[[17]](https://arxiv.org/html/2406.05806v2#:~:text=performing%20speech%20recognition%20model%2C%20Whisper,Conversely%2C%20we%20discover%20that%20Whisper). Autrement dit, fournir le mot correct ne veut pas forcément dire que le modèle va l’utiliser au bon moment. Parfois même, *un prompt exact peut ne pas améliorer la reconnaissance, voire la dégrader légèrement*[[18]](https://arxiv.org/html/2406.05806v2#:~:text=of%20textual%20prompts%2C%20often%20performing,within%20the%20language%20tokens%2C%20likely). Un développeur a rapporté qu’ajouter un terme unique en prompt avait perturbé la reconnaissance d’autres mots similaires phonétiquement (après avoir ajouté **“Artem”** en prompt, le mot **“startup”** a été mal transcrit)[[19]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=I%20have%20not%20tried%20suppressing%2C,Here%20is%20an%20example). Cela s’explique en partie parce que Whisper a été entraîné avec des **contextes de transcription réels** ; remplacer ce contexte par un prompt artificiel crée un écart qui peut dérouter le modèle dans certains cas[[20]](https://arxiv.org/html/2406.05806v2#:~:text=34%20%2C%20%2060%2C%2036,training%20setting)[[21]](https://arxiv.org/html/2406.05806v2#:~:text=improvements%20do%20not%20necessarily%20imply,facilitating%20further%20development%20and%20application).

En pratique, cela signifie que même en suivant les bonnes pratiques, **il peut arriver que le mot spécifique ne soit toujours pas correctement reconnu** si Whisper n’arrive pas à faire le lien entre le son et le mot. Si vos essais répétés avec le prompt "SWOOD" (ou d’autres) n’aboutissent pas, vous n’êtes pas seul – la communauté signale que *les prompts offrent un contrôle limité* et une fiabilité variable[[22]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=These%20techniques%20are%20not%20especially,be%20useful%20in%20some%20situations)[[23]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,of%20people%2C%20products%2C%20companies%2C%20etc). OpenAI eux-mêmes qualifient ces techniques de **« peu fiables, mais utiles dans certaines situations »**[[24]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,of%20people%2C%20products%2C%20companies%2C%20etc).

## Solutions en cas d’échec du prompt

Si l’ajout du mot au prompt ne suffit pas à le faire reconnaître, plusieurs **stratégies complémentaires** peuvent être envisagées :

* **Segmentation de l’audio** : Si votre vidéo est longue (plus de ~10 minutes), songez à la découper en segments plus courts (5–10 min) puis à transcrire chaque partie séparément en réutilisant le prompt à chaque fois. En effet, la performance de Whisper peut baisser sur de très longues durées continues[[16]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=3,3). Des utilisateurs indiquent avoir obtenu *100% de précision sur des noms propres et acronymes* en découpant de longs fichiers et en appliquant le prompt segment par segment[[25]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=I%20suggest%20the%20above%20order,could%20build%20the%20prompts%20dynamically). L’API OpenAI limite d’ailleurs les fichiers à 25 Mo, encourageant implicitement un traitement par morceaux[[26]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Longer%20Inputs). Veillez à couper de préférence aux pauses naturelles pour ne pas perdre de contexte[[27]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=By%20default%2C%20the%20Whisper%20API,some%20context%20to%20be%20lost).
* **Post-traitement avec un modèle de langage** : Une autre approche pratique consiste à **corriger a posteriori** les erreurs de vocabulaire spécifique. Par exemple, vous pouvez faire passer la transcription brute dans GPT-4 en lui fournissant la liste de vos termes-cibles, afin qu’il remplace les occurrences mal orthographiées. OpenAI suggère ce procédé : fournir à GPT une consigne du style *« Corrige les noms des produits suivants dans ce texte : SWOOD, SolidWorks, EFICAD… »* et lui donner la transcription à réviser[[28]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Post)[[29]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=names%20of%20the%20following%20products,capitalization%2C%20and%20use%20only%20the). Ce post-traitement permet de rattraper les termes que Whisper aurait manqués. Bien sûr, cela suppose que les mots *apparaissent au moins sous une forme reconnaissable* dans le texte (même déformée) – si Whisper a complètement ignoré un mot, c’est plus délicat, mais s’il l’a converti en un mot erroné récurrent, GPT peut aisément le corriger sur commande.
* **Suppression de tokens indésirables (approche avancée)** : Whisper offre, via son API bas niveau, la possibilité de **bloquer certains tokens** pendant le décodage. Des développeurs s’en sont servis pour forcer la reconnaissance d’un mot spécifique. L’idée est de *supprimer du vocabulaire du modèle le faux mot qu’il sort à la place du mot cible*. Par exemple, si au lieu de *“SWOOD”* le modèle transcrit *“wood”*, on pourrait tenter de *supprimer le token correspondant à "wood"* de la sortie possible. Ainsi, lors de la transcription suivante, Whisper devra choisir autre chose – potentiellement le terme correct s’il est dans le prompt. Un membre de la communauté explique cette astuce avec un exemple : Whisper transcrivait un nom latin *“Esthear”* au lieu de l’acronyme *“SDR”* attendu ; en **supprimant le token “Esthe”**, le modèle a dû proposer autre chose et a fini par deviner *“SDR”* correctement[[30]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=that%20token). Cette méthode nécessite d’identifier précisément le ou les tokens à bannir (via l’ID du tokenizer) et doit être maniée avec prudence. En effet, *supprimer un token trop courant peut dégrader d’autres mots* de la langue[[31]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=If%20you%20suppress%20the%20token,And%20hopefully%20guessing%20%E2%80%9CSDR%E2%80%9D%20correctly). Mais sur un mot très spécifique qui parasite votre résultat, cela peut orienter le modèle vers la bonne solution[[32]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=2). À noter que cette suppression se fait en paramétrant l’option suppress\_tokens dans l’appel Whisper (non disponible via l’API OpenAI standard, mais possible avec la librairie open source).
* **Fine-tuning du modèle** : En dernier recours, la solution la plus fiable est de **réentraîner (fine-tuner) Whisper sur des données incluant votre vocabulaire**. Si vous disposez d’enregistrements audio où ces termes sont prononcés avec leur transcription exacte, un fine-tuning même léger peut apprendre au modèle à les reconnaître[[33]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=If%20you%20have%20a%20corpus,For%20Multilingual%20ASR%20with%20Transformers)[[34]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=The%20tokenizer%20has%20an%20extensive,such%20as%20by%20adding%20extra). Hugging Face propose un tutoriel pour fine-tuner Whisper sur des corpus spécifiques[[33]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=If%20you%20have%20a%20corpus,For%20Multilingual%20ASR%20with%20Transformers). Dans votre cas, fine-tuner une version anglaise du modèle sur quelques exemples contenant “SWOOD”, “SolidWorks”, etc., améliorerait grandement leur détection. Cependant, cette approche est lourde : OpenAI n’a pas publié d’outil officiel pour le fine-tuning de Whisper, il faut passer par des scripts communautaires, et cela consomme du temps de calcul[[35]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=tschechische)[[36]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=Fine,hard%20task). À réserver si les autres méthodes échouent et que l’enjeu le justifie.

## Conclusion et recommandations

En résumé, **introduire vos mots spécialisés via initial\_prompt reste la première chose à faire** – c’est une fonctionnalité prévue pour *“biaiser” Whisper vers le jargon souhaité* sans modification du modèle[[5]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=1). Vous avez déjà appliqué les principaux réglages conseillés (modèle large, prompt pertinent, langue fixée, température basse, etc.), ce qui est idéal. Si malgré tout le mot n’apparaît pas correctement dans la transcription, souvenez-vous que ce comportement n’est pas inhabituel : le *prompt engineering* audio avec Whisper a ses limites, comme le confirment des chercheurs[[37]](https://arxiv.org/html/2406.05806v2#:~:text=information,In%20sum%2C%20We%20raise)[[18]](https://arxiv.org/html/2406.05806v2#:~:text=of%20textual%20prompts%2C%20often%20performing,within%20the%20language%20tokens%2C%20likely). Il faudra alors combiner d’autres astuces.

**Que faire concrètement dans votre projet ?** D’après les retours de développeurs, on peut procéder ainsi :

1. **Affiner le prompt** : tenter une phrase plus naturelle incluant le mot cible et éventuellement des synonymes ou contextes liés. Assurez-vous que le mot du prompt est **exactement** sous la forme souhaitée dans la sortie (majuscules, etc.)[[8]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=You%20can%20use%20a%20prompt,can%20help%20in%20different%20scenarios). Parfois, ajouter un ou deux mots connexes autour peut suffire à le faire émerger correctement dans la transcription[[6]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=1).
2. **Vérifier l’audio** : si possible, isoler un extrait où le mot est prononcé et tester Whisper dessus avec le prompt pour voir si le problème persiste. Un léger accent ou une prononciation ambiguë peut tromper le modèle. Vous pourriez comparer la reconnaissance en mode *greedy* (temperature=0) versus en mode *beam search* (paramètre beam\_size plus grand) pour voir si le mot apparaît dans les alternatives.
3. **Post-traiter les erreurs** : si Whisper continue d’écrire un mot incorrect à la place (par ex. *“Sword”* au lieu de *“SWOOD”*), envisagez un correctif automatique : un script qui remplace *“Sword”* par *“SWOOD”* dans le texte final, ou mieux, utilisez GPT-4 pour valider/corriger comme décrit plus haut. Un développeur indique qu’avec ce genre de workflow (transcription Whisper + révision GPT avec liste de vocabulaire), il a pu obtenir des transcriptions impeccables sur des termes techniques pointus[[25]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=I%20suggest%20the%20above%20order,could%20build%20the%20prompts%20dynamically).

En dernier lieu, gardez à l’esprit que Whisper, sans entraînement supplémentaire, **ne connaîtra pas toujours vos mots custom** – on peut l’orienter avec un prompt (et c’est généralement efficace dans 70–80% des cas d’après des témoignages[[36]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=Fine,hard%20task)[[38]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=Pretend%20you%27re%20talking%20about%20a,ewe%20sheep)), mais pas le forcer à 100%. Si votre application exige une perfection sur ces termes, investir du temps dans un fine-tuning ou dans un filtre/correction après coup sera peut-être nécessaire.

**Sources :** Les conseils ci-dessus proviennent de la documentation officielle d’OpenAI[[2]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=1,acronyms)[[4]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=audio_file%20%3D%20open%28,), du guide OpenAI Cookbook sur le prompting de Whisper[[39]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,Aimee%2C%20Shawn)[[40]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,Shawn%2C%20BBQ%2C%20Whisky%2C%20Doughnuts%2C%20Omelet), ainsi que de retours d’expérience de développeurs sur des forums (Hugging Face[[6]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=1)[[30]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=that%20token), Reddit[[5]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=1)[[32]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=2), StackOverflow[[41]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=1,3)). Ces sources confirment les meilleures pratiques et rappellent les limites actuelles de Whisper pour les vocabulaires spécialisés. En les suivant et en expérimentant prudemment, vous maximiserez vos chances d’obtenir la transcription correcte de vos termes spécifiques. Bonne continuation dans votre projet !

[[1]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Improving%20Reliability) [[2]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=1,acronyms) [[3]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Using%20the%20Prompt%20Parameter) [[4]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=audio_file%20%3D%20open%28,) [[7]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=prompt%3D,) [[8]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=You%20can%20use%20a%20prompt,can%20help%20in%20different%20scenarios) [[9]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=,Improving%20punctuation) [[10]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=file%3Daudio_file%2C%20response_format%3D,) [[26]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Longer%20Inputs) [[27]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=By%20default%2C%20the%20Whisper%20API,some%20context%20to%20be%20lost) [[28]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=Post) [[29]](https://www.rockapi.ru/docs/en/openai-speechtotext/#:~:text=names%20of%20the%20following%20products,capitalization%2C%20and%20use%20only%20the) Speech to Text | RockAPI Docs

<https://www.rockapi.ru/docs/en/openai-speechtotext/>

[[5]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=1) [[32]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=2) [[35]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=tschechische) [[36]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=Fine,hard%20task) [[38]](https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/#:~:text=Pretend%20you%27re%20talking%20about%20a,ewe%20sheep) [D] Custom vocabulary for Whisper? : r/MachineLearning

<https://www.reddit.com/r/MachineLearning/comments/xnzjpu/d_custom_vocabulary_for_whisper/>

[[6]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=1) [[19]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=I%20have%20not%20tried%20suppressing%2C,Here%20is%20an%20example) [[30]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=that%20token) [[31]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=If%20you%20suppress%20the%20token,And%20hopefully%20guessing%20%E2%80%9CSDR%E2%80%9D%20correctly) [[33]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=If%20you%20have%20a%20corpus,For%20Multilingual%20ASR%20with%20Transformers) [[34]](https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311#:~:text=The%20tokenizer%20has%20an%20extensive,such%20as%20by%20adding%20extra) Adding custom vocabularies on Whisper - Beginners - Hugging Face Forums

<https://discuss.huggingface.co/t/adding-custom-vocabularies-on-whisper/29311>

[[11]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,Aimee%2C%20Shawn) [[12]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=contained%20within) [[13]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,president%20biden) [[22]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=These%20techniques%20are%20not%20especially,be%20useful%20in%20some%20situations) [[23]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,of%20people%2C%20products%2C%20companies%2C%20etc) [[24]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,of%20people%2C%20products%2C%20companies%2C%20etc) [[39]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,Aimee%2C%20Shawn) [[40]](https://cookbook.openai.com/examples/whisper_prompting_guide#:~:text=,Shawn%2C%20BBQ%2C%20Whisky%2C%20Doughnuts%2C%20Omelet) Whisper prompting guide

<https://cookbook.openai.com/examples/whisper_prompting_guide>

[[14]](https://community.openai.com/t/whisper-is-translating-my-audios-for-some-reason/86468#:~:text=API%20community,of%20the%20speaker%27s%20native%20language) Whisper is translating my audios for some reason - API

<https://community.openai.com/t/whisper-is-translating-my-audios-for-some-reason/86468>

[[15]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=the,3) [[16]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=3,3) [[25]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=I%20suggest%20the%20above%20order,could%20build%20the%20prompts%20dynamically) [[41]](https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text#:~:text=1,3) transcription - How to improve Whisper speech to text - Stack Overflow

<https://stackoverflow.com/questions/78246902/how-to-improve-whisper-speech-to-text>

[[17]](https://arxiv.org/html/2406.05806v2#:~:text=performing%20speech%20recognition%20model%2C%20Whisper,Conversely%2C%20we%20discover%20that%20Whisper) [[18]](https://arxiv.org/html/2406.05806v2#:~:text=of%20textual%20prompts%2C%20often%20performing,within%20the%20language%20tokens%2C%20likely) [[20]](https://arxiv.org/html/2406.05806v2#:~:text=34%20%2C%20%2060%2C%2036,training%20setting) [[21]](https://arxiv.org/html/2406.05806v2#:~:text=improvements%20do%20not%20necessarily%20imply,facilitating%20further%20development%20and%20application) [[37]](https://arxiv.org/html/2406.05806v2#:~:text=information,In%20sum%2C%20We%20raise) Do Prompts Really Prompt? Exploring the Prompt Understanding Capability of Whisper

<https://arxiv.org/html/2406.05806v2>