

Évaluation de cinq outils de genèse de voix par intelligence artificielle afin de soutenir la narration de vidéos

Mathis Laverdière¹, Nicolas Martel-Côté¹, Thaina Jean-Baptiste¹, Denis Lebel¹, Jean-François Bussièrès^{1,2}

¹Unité de recherche en pratique pharmaceutique, Département de pharmacie, CHU Sainte-Justine, Montréal, Québec, Canada;

²Faculté de pharmacie, Université de Montréal, Montréal, Québec, Canada

Introduction

- La narration de vidéo peut être générée à partir d'outils peu coûteux dotés d'intelligence artificielle. La qualité des voix et la facilité d'utilisation varient selon les outils. Dans le processus de production de vidéos formatives, nous sommes rendus à la troisième étape (Fig.1).

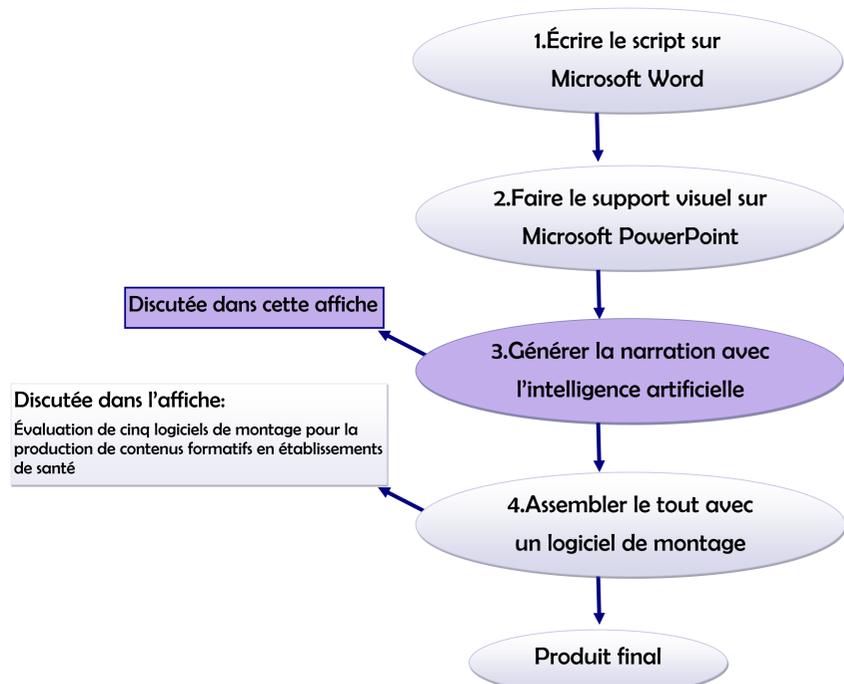


Figure.1 Exemple de processus de production de vidéos formatives avec proposition de logiciels/outils qui seraient utilisés.

Objectif

- Évaluer cinq outils de genèse de voix par intelligence artificielle dans le soutien de narration.

Méthode

- Étude descriptive transversale
- Cinq outils** (ElevenLabs™, MurfAI™, Alteredstudio™, FineVoice™ et Listnr™) ont été sélectionnés à partir d'une recherche dans le web.
- Un scénario (486 mots) de formation en pharmacovigilance a été développé et enregistré avec chaque outil deux fois (scénario original, ensuite après cinq minutes d'ajustement texte et voix)
 - Les ajustements de texte incluent :
 - la mise en lettre des chiffres, l'ajout de virgule et points dans les phrases, le retrait de caractères spéciaux, la mise en majuscule de lettres dans certains mots.
 - Les ajustements de voix incluent :
 - modification de la prononciation, du débit, de la stabilité, ajout de pauses.
- Cinq caractéristiques** (guide intégré, plusieurs langues disponibles, la prise en compte des caractères spéciaux/chiffres, le contrôle sur la stabilité et la clarté) ont été notées sur la présence (1) ou l'absence (0).
- Un groupe de discussion de 10 personnes a écouté les voix générées par les outils pour les classer du meilleur (1^{er} place) au pire (5^e place) selon **quatre critères** (fluidité, clarté, tonalité comparable à l'humain et accent en contexte québécois).
 - Pour le classement des outils dans chaque critère, 5 points étaient attribués à l'outil classé en 1^{er} place, 4 points en 2^e place, ..., 1 point en 5^e place.
 - Un maximum de 50 points pouvaient être attribués à un outil pour chaque critère.
 - Un outil « parfait » obtiendrait au plus 200 points.
- Après manipulation, les outils ont été classés en ordre croissant de difficulté d'utilisation selon l'expérience générale de l'utilisateur.

Résultats

- MurfAi™ était l'outil le plus cher (\$29) et ElevenLabs™ le moins cher (\$5) (Tableau I).
- MurfAi™ et Fine Voice™ étaient les deux outils avec toutes les caractéristiques présentes. Le reste des outils avait tous une caractéristique absente (Tableau.II).
- Après attribution des points selon leur classement, Alteredstudio™ avait accumulé le plus de points et Fine Voice™ le moins de points (Tableau.III).
- Avec le plus de points accumulés (171 points), Alteredstudio était le favori du groupe de discussion (Fig.2).
- ElevenLabs™ était le plus facile à utiliser selon l'utilisateur (Fig.3).

Tableau.I Informations générales des outils

Variable	ElevenLabs™ (1)	MurfAI™ (2)	Alteredstudio™ (3)	Fine Voice™ (4)	Listnr™ (5)
Logo					
Pays d'origine	Royaume-Uni	Inde	Royaume-Uni	Chine	Australie
Abonnement mensuel (\$USD)	\$5	\$29	\$6	\$8,99	\$9

Tableau.II Caractéristiques présentes dans les outils

Variable (Présente/Absente)	ElevenLabs™	MurfAI™	Alteredstudio™	Fine Voice™	Listnr™
Guide intégré	Présente	Présente	Présente	Présente	Présente
Plusieurs langues disponibles	Présente	Présente	Présente	Présente	Présente
Contrôle sur la stabilité	Présente	Présente	Présente	Présente	Absente
Contrôle sur la clarté	Présente	Présente	Absente	Présente	Présente
Prise en compte des caractères spéciaux/chiffre	Absente	Présente	Présente	Présente	Présente
Nombre total de critères présents sur le nombre de critères évalués	4/5	5/5	4/5	5/5	4/5

Tableau.III Points attribué aux outils selon les classements effectués par le groupe de discussion

Variable (n points)	ElevenLabs™	MurfAI™	Alteredstudio™	Fine Voice™	Listnr™
Fluidité: c'est agréable à écouter	44	36	40	15	15
Clarté: je comprends bien	44	32	44	14	16
Tonalité: comparable à la voix humaine ?	37	32	48	11	22
Accent: compatible au contexte québécois ?	35	34	39	12	15
Le nombre de points accumulés sur le nombre de points maximal	160/200	134/200	171/200	52/200	68/200



Figure.2 Classement de préférence, selon le nombre de points accumulés, de la voix générée par les outils avec les commentaires du groupe de discussion.

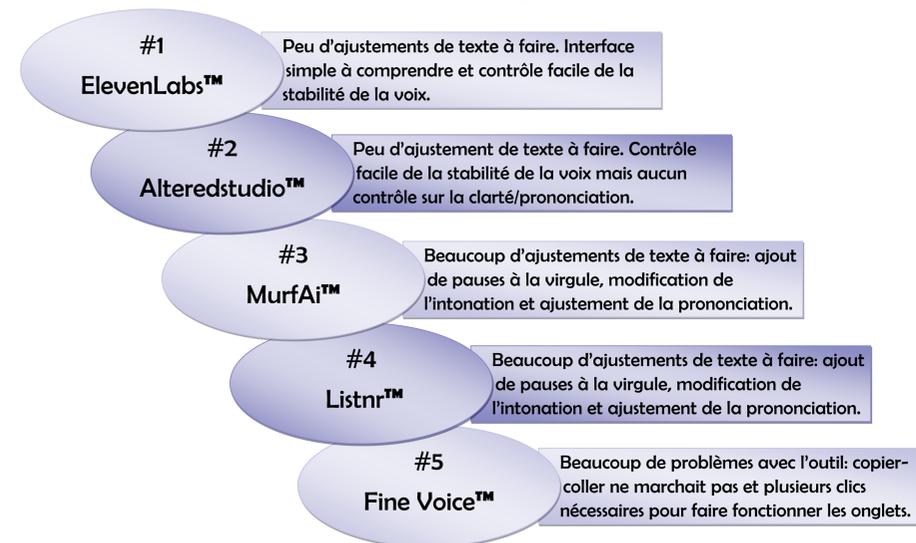


Figure.3 Ordre du plus facile au plus difficile d'utilisation avec les commentaires sur l'expérience générale de l'utilisateur.

Discussion/conclusion

- Selon le groupe de discussion:
 - les outils Listnr™ et Fine Voice™, à raison de voix robotique et de débit trop rapide, ne seraient pas adéquats dans le cas d'une narration de formation.
 - La prononciation de Alteredstudio™, permettant une bonne compréhension et intégration de l'information émise par la voix, était la raison pour laquelle cet outil était le favori.
- Les outils Alteredstudio™ et ElevenLabs™ demandaient de faire peu de modifications pour optimiser la voix, ce qui augmentait le nombre d'ajustements possible dans les 5 minutes allouées. Toutefois, une interface plus simple à comprendre et un meilleur contrôle de la stabilité de voix étaient remarqués dans ElevenLabs™.
- MurfAi™ et Listnr™ étaient les outils qui demandaient beaucoup d'ajustements de texte, limitant ainsi l'optimisation de la voix en 5 minutes. Des ajustements de prononciation, d'intonation et l'ajout de pauses dans la narration devaient être effectués.
- Fine Voice™ était l'outil le plus difficile à utiliser. Plusieurs problèmes ont été remarqués dans l'outil: copier-coller qui ne marchait pas et plusieurs clics nécessaires pour faire fonctionner les onglets. Cela rendait la genèse de voix difficile et limitait les possibilités de l'outil.
- Alteredstudio™ était l'outil préféré du groupe et ElevenLabs™ était plus facile à utiliser. Ces outils peuvent facilement être utilisés pour la narration de formations en établissement de santé.