

L'IA fait parler l'ordinateur

Posted on 7 octobre 2024 by Patrick Rouillier

Chacun connaît l'ordinateur qui joue de la musique ou restitue une conférence ou un discours. C'est le principe du fichier « Audio ». Mais l'ordinateur qui lit un texte qu'il ne connaît pas, à haute voix, c'est plus complexe.

Revenons un peu en arrière.

1) Dans une langue le langage parlé précède toujours le langage écrit. Les humains ont commencé à s'exprimer par des borborysmes avant de les structurer en langage parlé, avant de les transcrire sur la pierre ou le papier... ou sur un écran.

2) Le langage parlé est constitué de « phonèmes », les sons élémentaires qui combinés entre eux peuvent reproduire le son de tous les mots de la langue. Le Français d'aujourd'hui est constitué de 37 phonèmes (16 voyelles, 20 consonnes, 1 silence) pour exprimer les syllabes, les mots, les phrases. Pour transcrire ces phonèmes sur le papier on a associé à chaque phonème un « graphème », petite image graphique symbolisant le phonème. Ce que nous appelons une lettre n'est qu'une image symbolisant un son. C'est un jeu de briques Lego.

3) L'ordinateur dans sa mémoire ne connaît que deux entités 0 et 1. Il raisonne donc en langage binaire. On a défini une codification qui permet à la machine de comprendre les symboles des graphèmes. A chaque caractère symbolisant un graphème on associe un numéro (code ASCII)

Par exemple 2 est n° 50, A est numéro 65, a est numéro 97 etc. etc. Pour que l'ordinateur puisse comprendre cette numérotation elle doit être transcrite en numération binaire à l'aide de 0 ou 1.

2, ascii 50 devient 11 00 10

A, ascii 65 devient 1 00 00 01

a, ascii 97 devient 1 10 00 01

etc. etc.

L'ASCII est évidemment plus facile à retenir pour l'humain que le binaire

4) Lorsque vous tapez un 2 sur le clavier de l'ordinateur, celui-ci ne reçoit pas 2 (qu'il ne pourra comprendre) mais 11 00 10 qui est beaucoup plus clair pour lui. Inversement lorsque l'ordinateur doit écrire 11 00 10 sur l'écran il affichera un symbole 2 afin qu'un humain puisse comprendre.

Revenons maintenant à notre problème. Fournissons à notre machine un texte (fichier PDF par exemple pour les initiés) nous pensons fournir une suite de lettres alors que dans les faits nous fournissons une suite de nombres binaires représentant les lettres (totalement incompréhensible pour nous, mais que l'ordinateur comprend parfaitement)

Si à chacun de ces nombres binaires on associe le graphème correspondant

obtiendra un texte parfaitement lisible pour nous.

Si à chacun de ces nombres binaires on associe le phonème associé au graphème on reproduira le son de la voix qui lit le texte fourni initialement.

Procédé

- 1) On connaît les équivalences entre les lettres (graphèmes) compréhensible par l'humain et les nombres binaires compréhensibles par la machine.
- 2) On connaît l'équivalence entre les graphèmes et les phonèmes
- 3) On fait enregistrer par une voix humaine les 37 phonèmes de la langue Française
- 4) IL peut s'agir d'une voix masculine ou féminine, sans accent ou avec un accent Marseillais, Bourguignon, Alsacien, Sénégalais ou Vietnamien, avec un débit lent ou rapide, etc. etc. En choisissant la bon ensemble de phonèmes on pourra faire lire notre texte par un homme, un femme, un Marseillais ...
- 5) On fournit à la machine un texte littéral qu'il comprend en termes binaires
- 6) Soit il associe chaque nombre binaire au graphème correspondant et il reproduit le texte à l'écran
- 7) Soit il associe chaque nombre binaire au phonème correspondant et jouera ces phonèmes sur un haut-parleur, ce qui donnera l'illusion qu'il lit le texte. Banco, c'est gagné !

Conclusion

Cette approche (hyper simplifiée) permet à un ordinateur de simuler un locuteur qui lit le texte. A ce stade on obtient le genre de parler saccadé, mono corde qu'on attribuait jadis aux robots de science-fiction. Cela s'appelle la « Synthèse vocale ». L'ordinateur récite mais ne comprend rien à ce qu'il dit. C'est à ce stade que l'IA va jouer son rôle. Elle va simuler la compréhension du texte, y apporter les intonations voulues, un peu de vie et de chaleur, restituer le « climat » du texte. Il interprétera les ponctuations (point final, virgule, point d'exclamation ou d'interrogation) pour créer des silences ou moduler le ton de la voix, il comprendra si on est dans une phase de colère, de douceur, de persuasion ... pour modifier l'intonation du phonème.. Bref on aura vraiment l'impression d'un humain qui lit un texte