Les enceintes Echo, vitrine IA d'Amazon

Amazon et le *cloud gaming*, c'est désormais du sérieux. Le groupe américain a abattu ses cartes ce 24 septembre, en dévoilant un service nommé Luna. Il a aussi renouvelé sa gamme hardware.

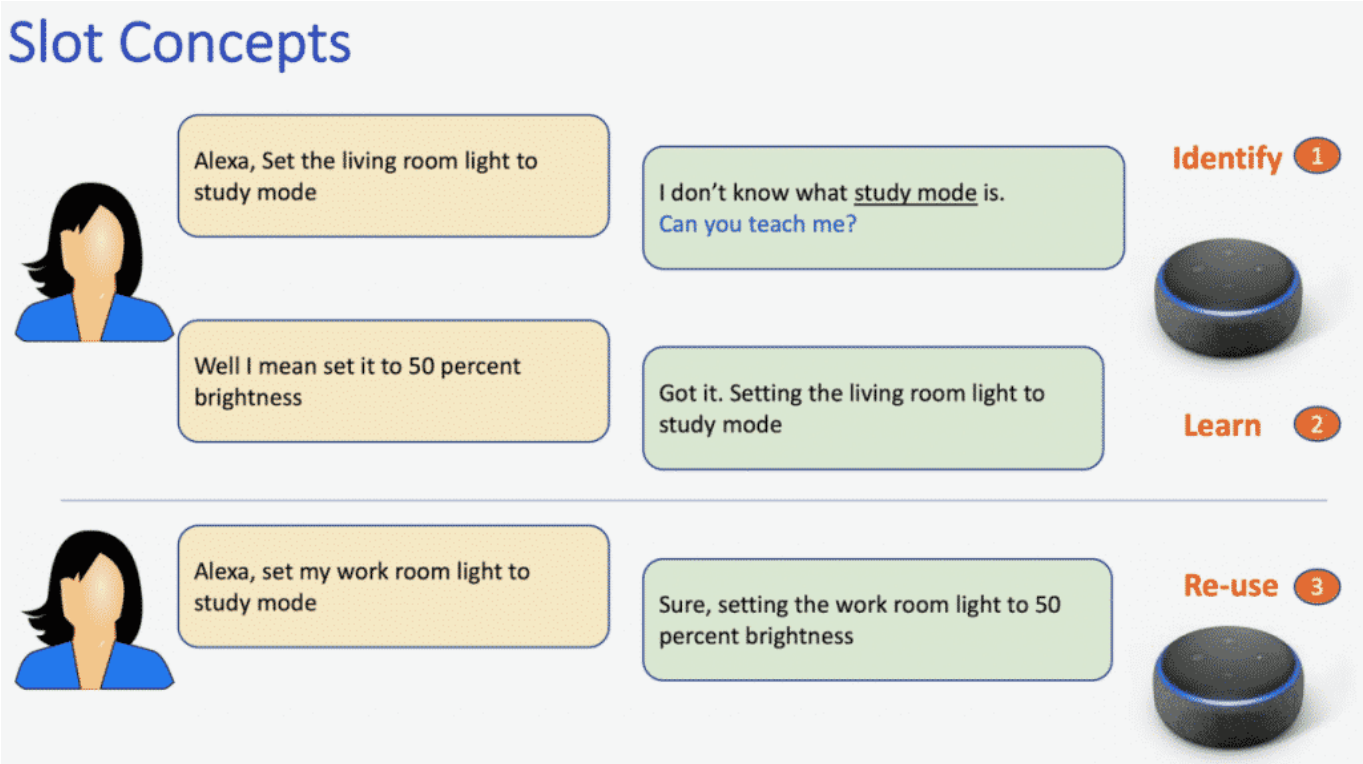
Avec, en première ligne, la 4e génération des Echo.

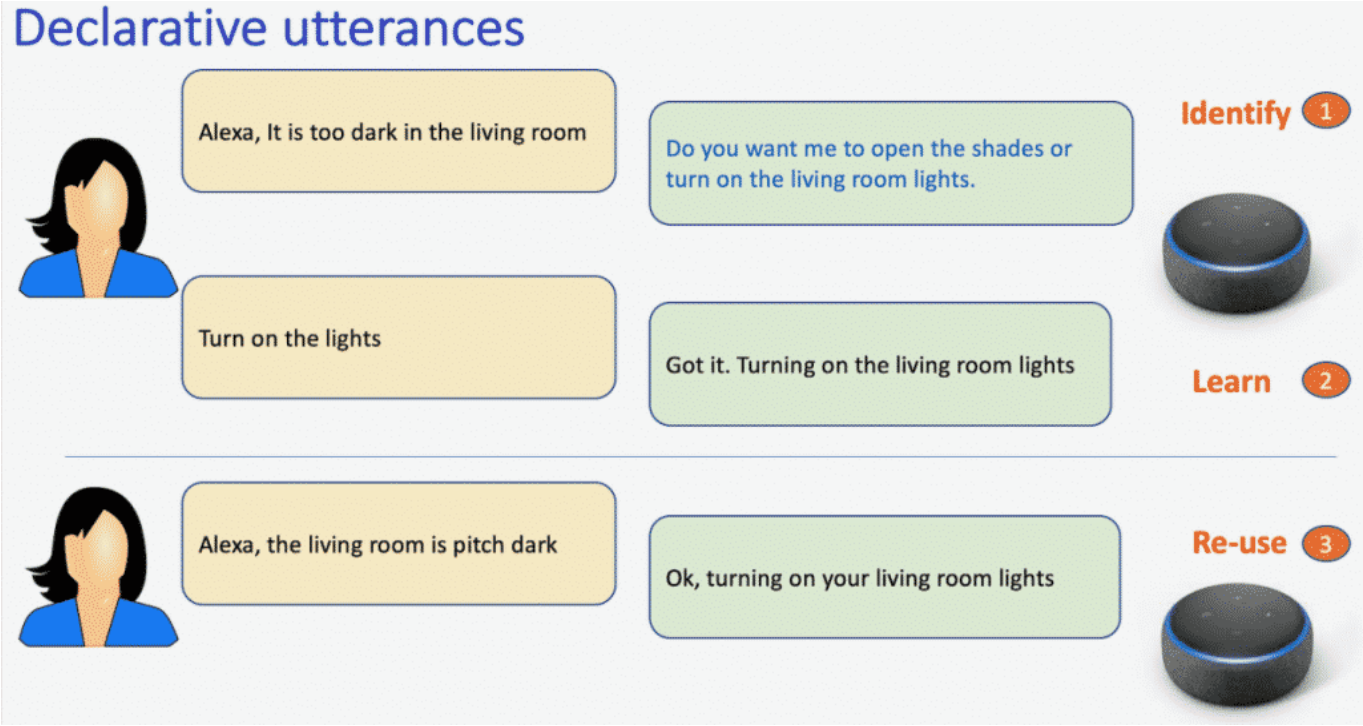
Il ne s'est pas attardé sur la fiche technique. Sauf pour un composant : la puce AZ1, destinée à exécuter des modèles IA à même les enceintes. À commencer par ceux qui interprètent et génèrent du langage.

Dans cet exercice, l'assistante Alexa est capable, depuis l'an dernier, de détecter des « signes d'insatisfaction » chez les utilisateurs – par exemple, lorsque ces derniers reformulent une question. Et de corriger ses réponses en conséquence.

Cette capacité d'autoapprentissage s'enrichira dans les prochains d'une composante « humaine ». Alexa pourra en l'occurrence demander des précisions pour apprendre des concepts « à la volée ». L'ensemble du processus se fera par voie conversationnelle. Il différera en cela des « routines », qui permettent d'utilisation l'application Alexa pour paramétrer des séries d'actions en réponse à des mots-clés.

On pourra enseigner à Alexa deux types de concepts. D'une part, ceux qui représentent des entités explicites (titre d'une chanson, nom d'un appareil connecté…). De l'autre, ceux dits « déclaratifs ». C'est-à-dire des instructions indirectes de type « Il fait trop sombre dans cette pièce ».





Les modèles de *deep learning* sur lesquels se fonde cette brique d'apprentissage remplissent quatre fonctions principales :

- Identifier, dans le discours, les parties problématiques pour Alexa
- Extraire la définition d'un concept
- Entretenir la conversation jusqu'à comprendre le concept
- Évaluer les actions réalisables en réponse à une instruction déclarative

Amazon vise le *concept-to-speech*

« Plus tard cette année », Alexa progressera sur un autre point : l'adaptation de son intonation et de son rythme de diction au contexte. Les modèles IA qui sous-tendront cette capacité devront aussi permettre de varier la formulation des questions que posera l'assistant.

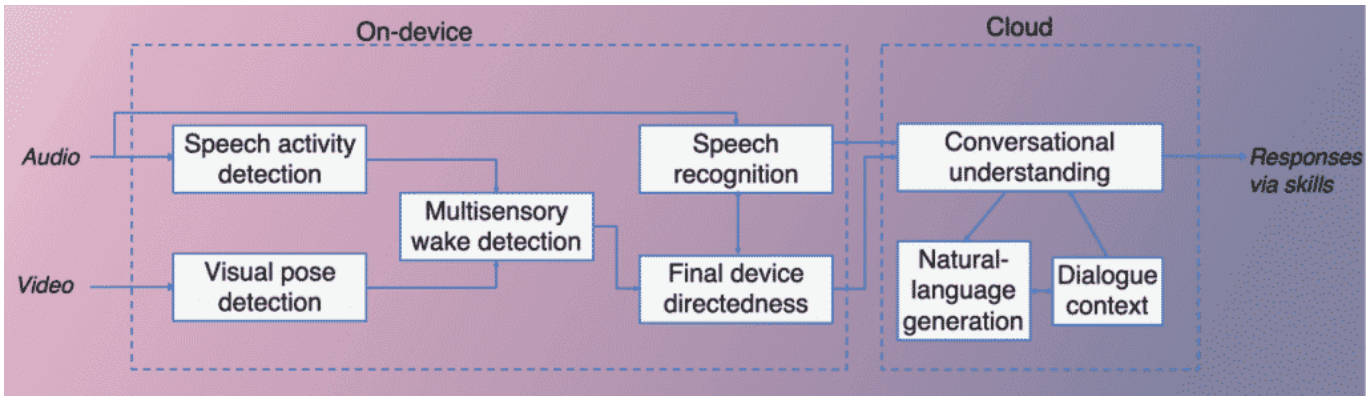
Amazon entend aller, dans ce cadre, du *text-to-speech* vers le *concept-to-speech*. Avec quatre éléments de référence : l'intention (action à réaliser), les entités impliquées, l'avancement de la conversation et le niveau de confiance d'Alexa quant à sa compréhension. La combinaison de ces paramètres sera transmise successivement à trois réseaux de neurones : le premier pour reformuler ; le deuxième pour identifier les mots-clés ; le troisième pour produire l'intonation et la diction correctes.

Autre extension à prévoir, mais pas avant l'an prochain : celle du mode Follow-Up, lancé voilà deux ans pour permettre de poursuivre une conversation sans avoir à prononcer à chaque fois le mot-clé « Alexa ».
Amazon y ajoutera un système dit de « tour de parole » (*natural turn taking*). Objectif : gérer les conversations qui impliquent plusieurs personnes. Avec des questions de type :

- Ces personnes parlent-elles entre elles ou s'adressent-elles à Alexa ?
- L'assistant doit-il rejoindre la conversation ?
- Dans l'affirmative, à qui doit-il répondre ?

Indices langagiers

Tandis que le mode Follow-Up utilise des indices acoustiques, le « tour de parole » y greffera, sur les appareils dotés d'une caméra, des indices visuels, pour déterminer si un utilisateur s'adresse à Alexa. En l'absence de caméra, des indices linguistiques (syntaxe et sémantique) seront pris en compte. L'ensemble fera l'objet d'un traitement en local, avant transmission vers les serveurs d'Amazon pour la compréhension du contexte et la conception de la réponse.



Ce processus induit, pour Alexa, une aptitude à gérer les interruptions. Autrement dit à savoir, quand on la coupe, où elle en était. Mais aussi à déterminer si un silence signifie que l'utilisateur a fini de parler ou s'il faut lui laisser davantage de temps. Principaux signaux exploités à ces fins : les onomatopées (hum, euh, etc.), les voyelles allongées et les phrases semblant incomplètes.

Illustrations © Amazon