

IAC - Symbole Grounding Problem - Synthèse

CONRAD Chloé - DAKHLI Sonia - MARTINEZ Christophe - SALAZAR Nathan

I-Présentation

Les objectifs principaux de la présentation étaient dans un premier temps de présenter et de définir le problème de l'ancrage des symboles pour ensuite aborder son importance en IA et faire une revue des avancées de la recherche dans ce domaine. Pour bien présenter ce problème, il était nécessaire d'introduire le triangle sémiotique et de définir les éléments qui le composent (i.e. : Objet, symbole et concept). Ainsi, il est possible de se servir de ces définitions pour définir le problème étudié. En effet, on peut dire qu'un système a ses symboles ancrés s'il possède une méthode pour les relier à leur concept respectif. Le problème de l'ancrage des symboles a été défini pour la première fois en 1990, mais connaît un regain d'intérêt au cours de la dernière décennie. Il est aussi important de souligner que cet ancrage dépend également du contexte, car un même symbole peut désigner plusieurs concepts.

La notion d'intrinsèque est centrale dans le problème de l'ancrage des symboles. En effet, un ancrage codé en dur ne pourrait être exhaustif pour un système devant interagir avec un environnement complexe. De plus, cela introduit par ailleurs les biais de perception des programmeurs dans le système.

Ce problème fait donc l'objet de nombreuses recherches en intelligence artificielle, ce qui aboutit à un plusieurs propositions pour le résoudre, dont celle de Taddeo et Floridi (2007) qui a longtemps été considérée comme la solution au problème. Cependant, elle a depuis été de nombreuses fois contredite et aucune solution ne fait désormais consensus.

II-Débat

1) *Les symboles sont-ils ancrés chez l'humain ? La représentation via le triangle sémiotique est-elle tout le temps représentative dans notre cas ?*

Différents exemples ont été cités pour répondre à cette question, notamment celui des mathématiques. L'infini ou le nombre "i" possèdent un symbole, mais ils n'ont pas d'objet associé et bien que leur concept soit compréhensible, il reste compliqué à définir. Vient ensuite la question du triangle en mathématiques, le symbole et le concept sont bien définis mais l'objet du triangle est plus difficile à définir. Le dessin d'un triangle est-il l'objet ? Ou alors s'agit-il plus d'une caractéristique géométrique d'objets physiques ?

Il s'en est poursuivi une réflexion sur la pertinence de chercher à ancrer tous les symboles dans le contexte de l'intelligence artificielle. Certains participants se sont demandés si cet ancrage était nécessaire pour le développement de l'IA. Un participant a comparé le fonctionnement des réseaux de neurones artificiels à celui d'un enfant confronté à quelque chose qu'il ne connaît pas. Alors que l'enfant poserait la question "C'est quoi ?", le réseau de neurones classe l'objet ou le concept sans chercher à comprendre en profondeur.

Le débat a ensuite abordé la capacité humaine à apprendre un nombre pratiquement infini de mots et de concepts. Cela a soulevé la question de savoir si la recherche de l'ancrage pour chaque symbole était nécessaire ou s'il était acceptable que certains symboles ne soient pas intrinsèques à l'IA.

C'est ensuite la nécessité d'oublier certaines informations qui pourraient devenir obsolètes qui concentra le débat. Il a été suggéré que le système devrait non seulement chercher à accumuler de nouvelles connaissances, mais également à réviser ses connaissances existantes, en supprimant ce qui n'est plus valide. Cette capacité à réviser la connaissance a été identifiée comme une composante importante du développement d'une intelligence artificielle capable de contextualiser et d'interagir de manière flexible avec un monde en constante évolution.

2) "Peut-on parler de symbol grounding dans le cas de ChatGPT ?"

Il a d'abord été évoqué que ChatGPT répond de manière statistique. Cependant, il est capable de fournir des descriptions détaillées d'objets, comme une chaise, ce qui peut parfois donner l'impression qu'il a compris la signification du mot. Cette habileté pose la question de la distinction entre la simple réponse statistique et une véritable compréhension sémantique. Il est aussi important de noter que ChatGPT peut également produire des textes qui semblent cohérents à la lecture, mais qui en réalité sont le résultat d'une forme d'hallucination linguistique. Ces "hallucinations" se produisent lorsque le modèle génère du contenu qui paraît sensé et pertinent, mais qui n'est pas basé sur une compréhension authentique du sujet. En effet, dans certains contextes, lorsque ChatGPT produit une explication simple et précise, il peut donner l'impression d'avoir compris le sujet en question, notamment lorsqu'il fournit une explication claire de ce qu'est une chaise. Il est même en capacité de reformuler.

Un point de vue partagé dans le débat est que, malgré ses dispositions à fournir des réponses précises, ChatGPT ne possède pas une véritable compréhension sémantique au sens humain. Certains participants ont fait référence à la comparaison avec les êtres humains qui peuvent mémoriser des informations sans nécessairement les comprendre en profondeur. Un participant a souligné que la capacité de ChatGPT à reformuler les informations ne signifie pas sa compréhension du sujet. En d'autres termes, la reformulation n'implique pas une réelle compréhension conceptuelle.

Cela a soulevé une nouvelle question en lien avec la première de ce débat : si nous ne sommes pas capables de faire la différence entre ce qu'il a compris ou non, est-ce que la partie "objet" du triangle sémiotique est si importante que cela ? En effet, tous les symboles n'ont pas forcément d'objets physiques qui leur sont liés. Alors, pour ces symboles, comment différencier l'objet du concept ?

Il a été évoqué que ChatGPT était en capacité d'apprendre de l'interaction avec les utilisateurs. Que pour être capable de maintenir l'interaction il se crée une représentation vectorielle des nouvelles connaissances. Mais comme il ne fait pas de cognition sur ces connaissances, on ne peut pas définir ce mécanisme comme de l'ancrage de symbole.

En conclusion, il n'y a pas de consensus sur la définition et l'aspect résolu ou non du symbole grounding problem dans le milieu de la recherche. Et si ce sujet revient en vogue aujourd'hui, c'est en partie suite au déploiement des modèles conversationnels tels que ChatGPT et aussi en conséquence de la quête de l'explicabilité de nos modèles d'IA.