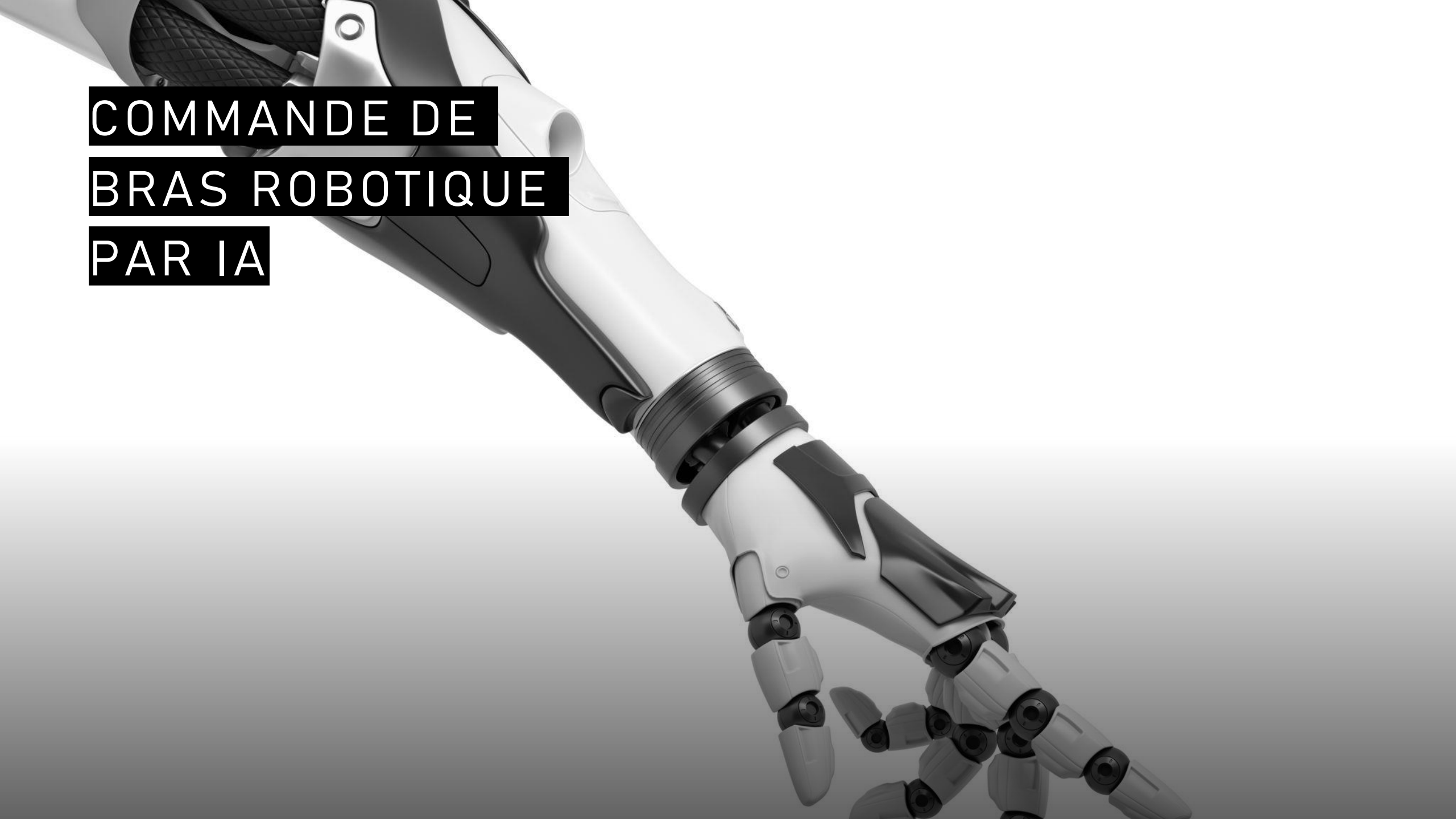
COMMANDE DE BRAS ROBOTIQUE PAR IA



DESCRIPTION DU SCÉNARIO

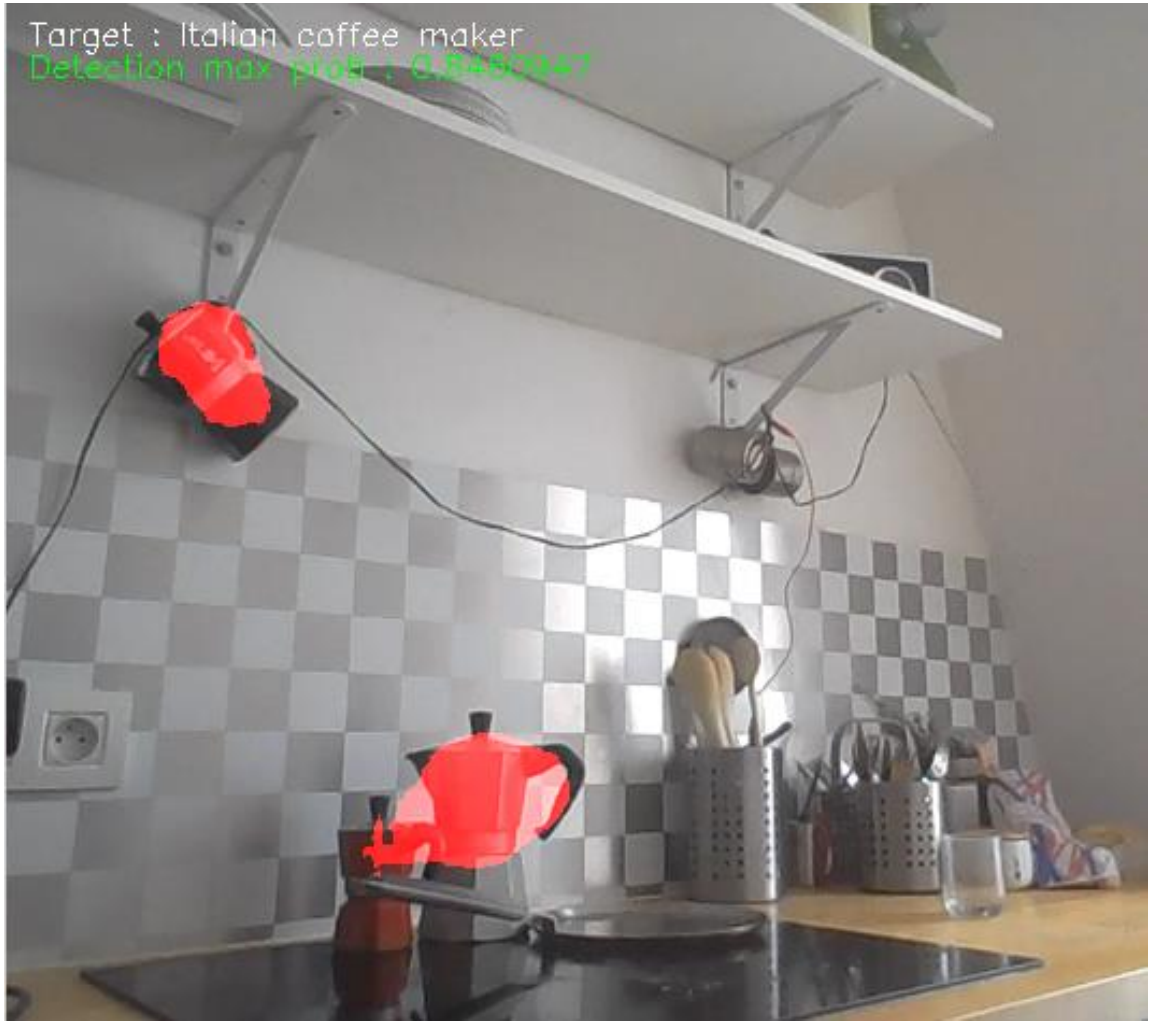
1. Un opérateur décrit au bras robotique les taches qu'il devra exécuter:

2. Le bras exécute les taches en l'absence de l'opérateur

PROBLEMATIQUE 1: DETECTION DES OBJETS

Des modèles existent déjà pour trouver des objets quelconques sur une image:

- CLIP [1]: lien entre texte et image - modèle entraîné sur les images labellisées sur internet
- CLIPSeg [2]: segmentation par un texte quelconque - modèle dérivé de CLIP



Le modèle CLIP [1]:

- La plupart des modèles comme d'analyse d'image comme ResNet procède par classification:

Retour du modèle:
Chat

- Ces modèles ne sont pas adaptés à notre cas car l'opérateur peut désigner des objets qui ne sont pas parmi les classes d'objets reconnus par le modèle
- Le modèle CLIP est adapté car on lui fournit l'image et le texte et il répond par une probilité d'alignement entre les deux medias:

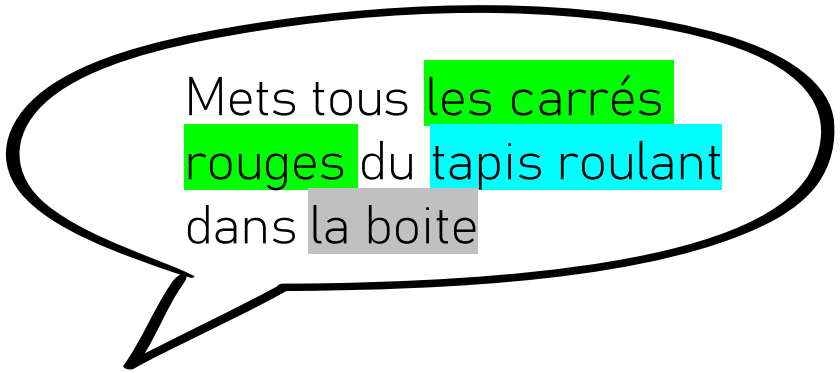
et le texte "chat assis"

Probabilité: 0.14315

PROBLÉMATIQUE 2: COMPRÉHENSION ET EXÉCUTION DES ORDRES DE L'OPERATEUR

3 procédés possibles:

1) **Extraction des relations**: Dans chaque phrase rechercher les relations de contenance



(les carrés rouges, tapis roulant) Model IA --> emplacement

(les carrés rouges, la boite) Model IA --> contenu/contenant

(tapis roulant, la boite) Model IA --> pas de relation

Ensuite on exécutera toujours le même programme:
`takeAndDrop(contenu,contenant)`

Pour cette approche les modèles sont des modèles d'**extraction des relations**:

- Le modèle [3] "Matching the blanks" (implémenter en [4]) **classifie** les pairs de groupes nominaux d'une phrase dans un nombre prédéterminé de catégories de relations.

  Exemple: dans la phrase "Mets tous les carrés rouges du tapis roulant dans la boite"
  paire (carré rouge, boite) -> contenu/contenant

- Le modèle [5] "FewRel – Zero shot" va donner une **probabilité** qu'une relation donnée (quelconque) correspondante à une paire de groupe nominaux.

  Exemple: dans la phrase "Mets tous les carrés rouges du tapis roulant dans la boite"
  paire (carré rouge, boite) et relation "contenu/contenant" -> probabilité 0.97321

PROBLÉMATIQUE 2: COMPRÉHENSION ET EXÉCUTION DES ORDRES DE L'OPERATEUR

3 procédés possibles:

2) Utiliser un **LLM Multimodal** (eg GPT4) pour commander le robot directement (cf [6])

STATUS> Etant donné <  > et l'ordre "mettre tous les carrés rouges sur la plaque blanche", que doit-on faire?

Model IA --> `takeAndDrop("carré rouge","plaque blanche")`

STATUS> Etant donné .....

PROBLÉMATIQUE 2: COMPRÉHENSION ET EXÉCUTION DES ORDRES DE L'OPERATEUR

3 procédés possibles:

3) **Génération de code**: à partir des phrases de l'opérateur, générer un code pour commander le robot (cf [7] et les modèles libres [8] et [9]) :

"mettre tous les carrés rouges sur la plaque blanche"

Model IA -->
```
While(1):
        TakeAndDrop("carré rouge","plaque blanche")
```

COMPARAISON DES 3 PROCÉDÉS

- Procédé 1 – Extraction des relations: modèles déjà existants mais ordres simples

- Procédé 2 - LLM: modèles uniquement accessibles en ligne, coût élevé, non protection des données, imprévisibilité des réactions du modèle

- Procédé 3 - Génération de code: modèles existants mais nécessitent beaucoup de puissance ou bien utilisation de modèles plus simples mais nécessitent de nombreuses données pour ajuster leur apprentissage (cf par exemple [10]). Comment obtenir des données d'entraînement?

RÉFÉRENCES

- [1]: Alec Radford, et al. "Learning Transferable Visual Models From Natural Language Supervision." (2021).
- [2]: Timo Lüddecke, et al. "Image Segmentation Using Text and Image Prompts." (2022).
- [3]: Baldini Soares, Livio et al. "Matching the Blanks: Distributional Similarity for Relation Learning." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019.
- [4]: Wee Tee Soh, BERT(S) for Relation Extraction, Github repo: github.com/jpablou/Matching-The-Blanks-Ths, 2019
- [5]: Cetoli, Alberto. "Exploring the zero-shot limit of FewRel." *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, 2020.
- [6]: Danny Driess, et al. "PaLM-E: An Embodied Multimodal Language Model." (2023).
- [7]: Jacky Liang, et al. "Code as Policies: Language Model Programs for Embodied Control." (2023).
- [8]: Raymond Li, et al. "StarCoder: may the source be with you!." (2023).
- [9]: Allal, Loubna Ben et al. "SantaCoder: don't reach for the stars!". *arXiv preprint arXiv:2301.03988*. (2023).
- [10]: Philipp Schmid, Fine-tune a non-English GPT-2 Model with Huggingface, Google Colab notebook: colab.research.google.com/github/philschmid/fine-tune-GPT-2/blob/master/Fine_tune_a_non_English_GPT_2_Model_with_Huggingface.ipynb, 2020