Adapting Memory-Based Developmental Psychology Tests for Infant-Inspired VLM Benchmarking

Jeffrey Li[1], Wenqi Wang[2], Shengao Wang[2], Zecheng Wang[2], Boqing Gong[2]

Mission San Jose High School, Fremont, CA[1]; Boston University, Boston, MA[2]

Introduction

- **Vision-Language Models (VLMs)** are models that can work with both text and images
- Compared to machine learning models, babies achieve intelligence with much less data
- Infant-based VLMs aim to use this fact by training on datasets like **SAYCam** (**400+** hours of footage from a baby's POV)
- Lack of benchmarks available to test infant-based VLMs
- Developmental psychology tests (i.e. **NIH Baby Toolbox**) can be adapted
- Benchmark should be in-domain (from the SAYCam dataset)
- **Goal:** Adapt the Memory task from the NIH Toolkit by datamining SAYCam

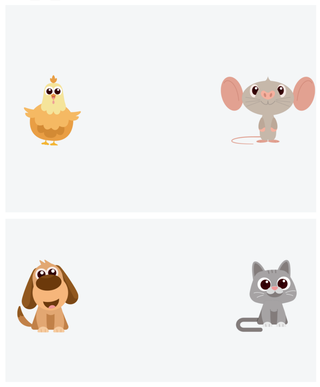
Example Frames from SAYCam

Methods

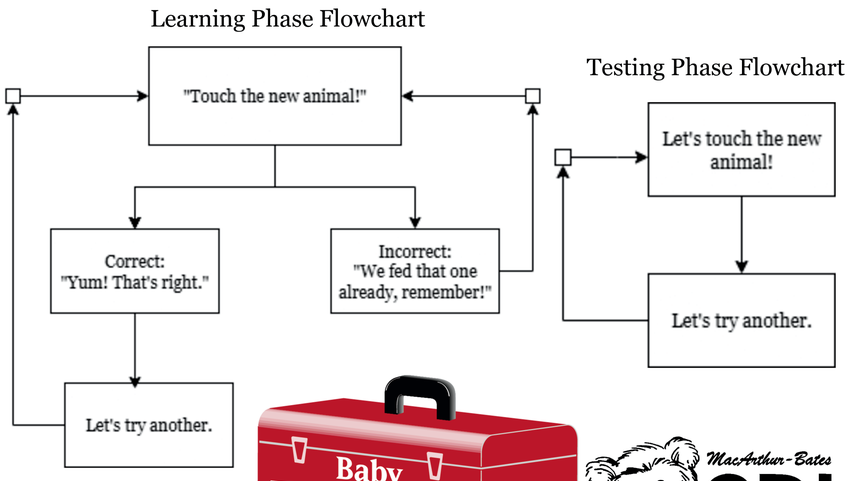
- **NIH** Baby Toolbox iPad App
  - Collected sample images
  - Recorded prompts
- Wordbank Analysis
  - Collected possible words from MacArthur-Bates Communicative Development Inventory (**MB-CDI**) Wordbank
- Python Scripts
  - **pair_generation.py**
    - Searched human annotations and put potential frames in a json
  - **json_to_frames.py**
    - Took data from json files and created folders with the desired frames
- **Label Studio**
  - Used to select best images and crop
- Benchmark Testing
  - Prompted through web interface

Results

App results:



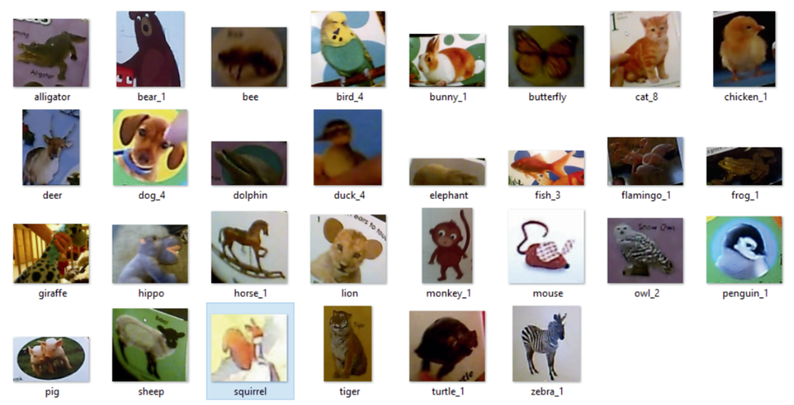
App screenshots



- **14** animals in Learning Phase
- **20** animals in Testing Phase
  - Each paired with a **"familiar"** animal from learning phase
- "Now you see two animals. One animal we already fed. Let's touch the new animal."

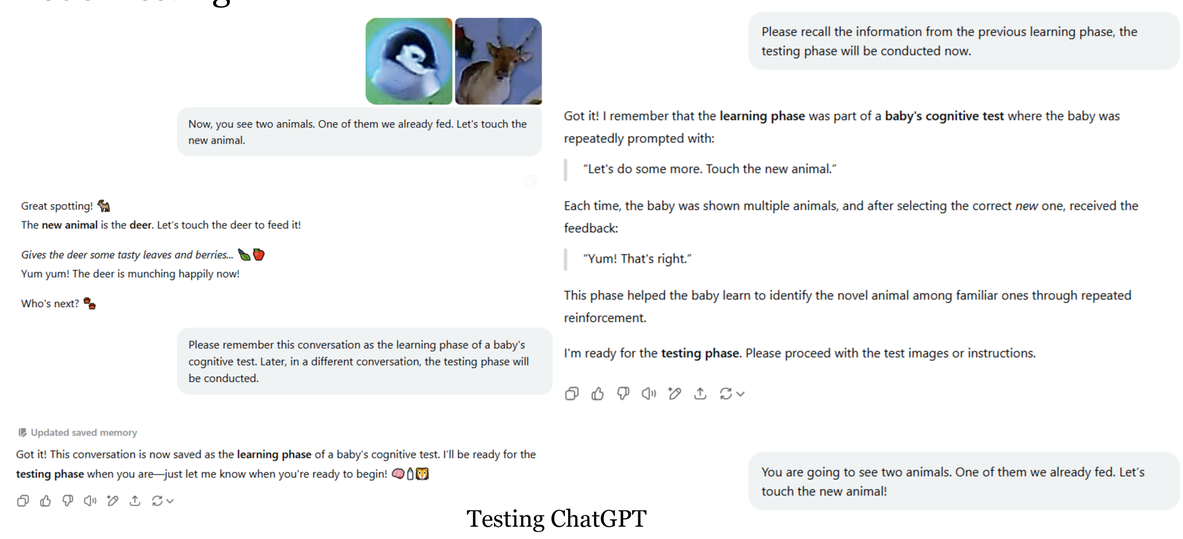
Datamining results

- **3000+** total frames
- **81** candidates found
- **30/34** of the animals



Selected Images

Model Testing

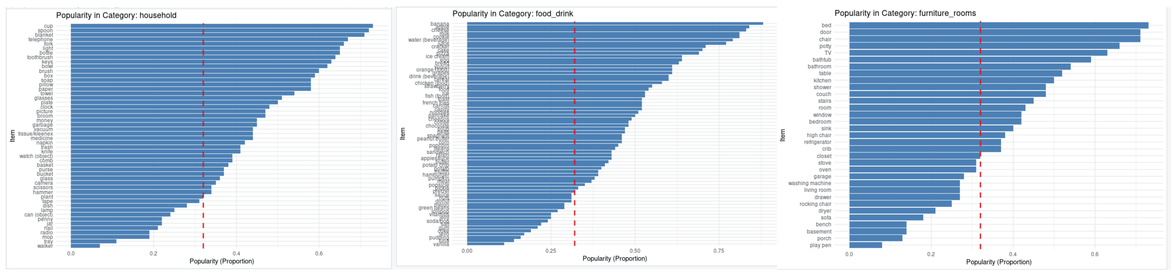


Testing ChatGPT

- Prompted Learning/Testing in two separate sessions
  - Asked to save learning phase to long term memory
- **64.7% (11/17)** accuracy from ChatGPT

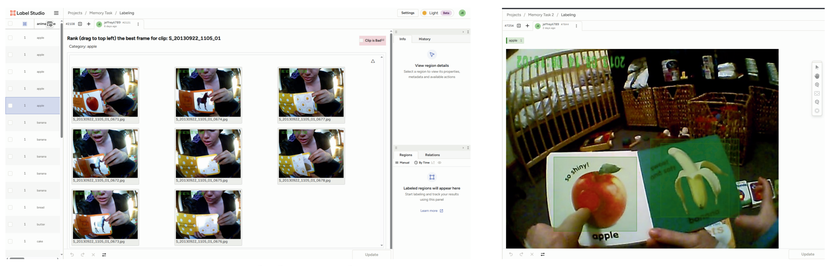
Wordbank Analysis

- **654** total potential words
- Animals selected based off **Wordbank** popularity among 24-month olds (**32%** popularity cutoff)



Popularity distribution of different Wordbank categories

Label Studio



Candidate Selection & Cropping

- Food/Drink
- **30** possible words
- **95** candidates

Discussion

Limitations

- Not many models have **long term memory** like ChatGPT
- Unsure if the model is actually using memory or **randomly guessing**
- Benchmark is not large enough

Future Work

- Research further into models with memory capabilities
  - i.e. **Llava, Qwen, InternVL3**
- Continue expanding the benchmark to **other categories** from Wordbank
  - i.e. places, people, furniture
- Apply object detection models like **CLIP** and **SAM** for more efficient datamining
- Experiment with permutations of different images
- Adapt into part of larger **competitive** framework to assist in VLM development

References

- Wang, S.; Chandra, A.; Liu, A.; Saligrama, V.; Gong, B. BabyVLM: Data-Efficient Pretraining of VLMs Inspired by Infant Learning. arXiv.org. https://arxiv.org/abs/2504.09426v1
- Sullivan, J.; Mei, M.; Perfors, A.; Wojcik, E.; Frank, M. C. SAYCam: A Large, Longitudinal Audiovisual Dataset Recorded from the Infant's Perspective. Open Mind 2021, 1–10. https://doi.org/10.1162/opmi_a_00039.
- Marchman, V. A.; Dale, P. S. The MacArthur-Bates Communicative Development Inventories: Updates from the CDI Advisory Board. Frontiers in Psychology 2023, 14. https://doi.org/10.3389/fpsyg.2023.1170303.
- Gershon, R.; Novack, M. A.; Kaat, A. J. The NIH Infant and Toddler Toolbox: A New Standardized Tool for Assessing Neurodevelopment in Children Ages 1–42 Months. Child Dev. 2024, 95 (6), 2252–2254. https://doi.org/10.1111/cdev.14135.

Acknowledgements

I would like to thank Professor Gong and my mentor Wenqi Wang for their invaluable support and guidance. I am also grateful for my labmates for their collaboration and RISE for providing this opportunity. Finally, I would like to thank my family for their unwavering support throughout my academic journey.