Turtle-like Geometry Learning: How Humans and Machines Differ in Learning Turtle Geometry

Sina Rismanchian, Shayan Doroudi, Yasaman Razeghi

University of California, Irvine
srismanc@uci.edu, doroudis@uci.edu, yrazeghi@uci.edu

Abstract

While object recognition is one of the prevalent affordances of humans' perceptual systems, even human infants can prioritize a place system over the object recognition system, that is used when navigating. This ability, combined with active learning strategies can make humans fast learners of Turtle Geometry, a notion introduced about four decades ago. We contrast humans' performances and learning strategies with large visual language models (LVLMs) and as we show, LVLMs fall short of humans in solving Turtle Geometry tasks. We outline different characteristics of human-like learning in the domain of Turtle Geometry that are fundamentally unparalleled in state-of-the-art deep neural networks and can inform future research directions in the field of artificial intelligence.

Introduction

The remarkable progress made by large language models (LLMs)—including large visual language models (LVLMs)—and, more broadly, cutting-edge generative artificial intelligence (AI) has prompted researchers to evaluate these models in various fields where humans typically excel, like mathematics and science question answering (Team et al. 2023). The extent to which the reasoning and learning processes in LVLMs parallel human learning remains a contentious topic among researchers. Here, we aim to examine the strengths and shortcomings of these models in performing visual programming tasks, specifically in the domain of Turtle Geometry (Abelson and DiSessa 1986).

Recent work in the field of psychology shows that infants possess two different systems in the domain of geometry: a *form* system which is used for object detection and a *place* system that is used when navigating (Spelke and Kinzler 2007; Dillon 2023). Furthermore, researchers have shown that humans tend to prioritize the place system over the form system when perceiving abstract geometric shapes (Lin and Dillon 2023). We believe such findings resonate with the ability of young children (and adults) to learn Turtle geometry, a form of geometry that can be explored through programming, in contrast to Euclidean geometry as traditionally taught in schools. Indeed, recent work even suggests the human mind might recognize shapes in terms of procedural programs in a Turtle geometry-like language (Sablé-Meyer et al. 2022). Solving Turtle geometry tasks requires a niche human ability that combines both perception and problem-solving. Hence, we believe it could be an interesting task for evaluating the performance and behavior of LVLMs in order to enhance our knowledge of the fundamental characteristics of these models.

We hypothesize that state-of-the-art generative AI multimodal models such as GPT4-V, lack the human ability to visualize and procedurally generate abstract shapes and patterns—and to quickly learn how to do this. To test this hypothesis, we (1) curated a dataset of images generated with Turtle programming and subsequently tested a large-language model in its ability to generate programs for those shapes and (2) ran preliminary human subject experiments with graduate students who have a background in programming and mathematics but are not familiar with Turtle geometry, to see how quickly they learn to create some of the shapes in our dataset.

Experiments with Human Participants

We recruited 7 graduate students—4 female and 3 male—none of whom were previously familiar with Turtle Geometry but all of whom had some experience in Python programming. During each experiment, participants were first introduced to a manual that described specific commands needed in the Python Turtle library. They did not need to memorize the commands as they could return to the manual during the experiment. After reading the manual, we asked them to recreate a star using Python Turtle. Three of the participants were successful in creating the star in five minutes. Others who were not successful in the given period were provided with some example code in Python Turtle that created a square that they could run to see the turtle's path and gain insights into how it works. After this step, we wanted all of the participants to recreate four other specific shapes, and interestingly, 6 out of 7 participants were successful in recreating all the shapes.

During these experiments, three specific strategies were identified:

1. 6 participants employed the place system over the form system. In particular, they embodied themselves as turtles and tried to navigate the shapes to find out how to

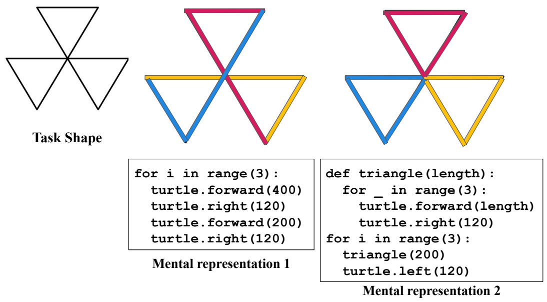

Figure 1: Two different decompositions of a single shape and their corresponding pieces of codes

program them. They usually used their hands or heads to simulate turtle heading and calculate turning angles.

2. 6 participants solved at least one task by using arbitrary subtasks. For instance, to recreate a shape including three rotated triangles, one first tried to create a single triangle and generalize their program to create the whole shape.
3. 4 participants, at least once showed the ability to hold different representations of the task at hand and pick the one that eases the programming as depicted in Figure 1.
4. 2 participants, at least once used background knowledge in Euclidean geometry to calculate the turning angles

Experiments with GPT-4V

We curated a set of 60 different tasks and asked the GPT-4V model (as a widely used LVLM) to come up with code using the Python Turtle library that could recreate the shapes. We used different prompting techniques to account for a variety of ways that could possibly enhance the model's performance. We found that while the model can produce coherent Python code that runs without errors, it is often unable to write programs that can recreate shapes that are found to be easily learned by human learners. In fact, the model is mostly able to recreate single shapes that only include a specific shape (e.g., square, circle, triangle, etc.), and for the shapes that include a composition of single shapes, these models lag behind participants in our experiment. Further analysis of GPT-4V's performance and shortcomings along with our benchmark will be published in future work.

Conclusion

Our experiments with students show three significant differences between how humans and machines learn in the domain of visual programming: (1) humans can adopt the place (i.e., navigation) system, while transformer architectures mostly rely on object recognition, (2) humans can generate subgoals to engage in active learning strategies such as trial and error that help them learn from feedback while on the other hand, current deep neural networks are passively fed by input-output pairs, and (3) humans can hold different forms of visual abstractions from a single shape and choose the one that eases their programming experience while to best of our knowledge, these models lag behind humans in abstraction tasks (Moskvichev, Odouard, and Mitchell 2023) and we have not seen any reports on the flexibility of these models in holding different abstractions. We see several research questions motivated by human characteristics that need to be answered in the future. For instance, it is not clear how LVLMs connect their visual and language models, while previous work has shown that language has a significant impact on humans' visual perception (Dillon 2023). Furthermore, it is noteworthy that the notion of Turtle Geometry is different from traditional mathematics such as Euclidean geometry in the way it is learned (Abelson and DiSessa 1986):

> it is the rare student who gets [a] chance to approach mathematics by doing it rather than only learning about it, by investigating new phenomena, by formulating original hypotheses, or by proving original theorems.

Therefore, unlike learning an axiomatic deductive form of geometry, Turtle geometry may require active learning or discovery learning. However, the problem of having an AI generate its subgoals or engage in active learning is perhaps still ambiguous and needs to be studied, while some promising work has been done recently (Ellis et al. 2023). We believe studying the tasks that humans excel at and studying their learning characteristics can enhance the interdisciplinary study of intelligence and in turn, inform AI research to eventually build machines that learn like humans.

References

Abelson, H.; and DiSessa, A. 1986. *Turtle geometry: The computer as a medium for exploring mathematics*. MIT press.

Dillon, M. R. 2023. Divisive language. 10.31234:1706.03762.

Ellis, K.; Wong, L.; Nye, M.; Sable-Meyer, M.; Cary, L.; Anaya Pozo, L.; Hewitt, L.; Solar-Lezama, A.; and Tenenbaum, J. B. 2023. DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning. *Philosophical Transactions of the Royal Society A*, 381(2251): 20220050.

Lin, Y.; and Dillon, M. R. 2023. We Are Wanderers: Abstract geometry reflects spatial navigation. *Proceedings of the National Academy of Sciences of the United States of America*, 110: 35.

Moskvichev, A.; Odouard, V. V.; and Mitchell, M. 2023. The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain. *arXiv preprint arXiv:2305.07141*.

Sablé-Meyer, M.; Ellis, K.; Tenenbaum, J.; and Dehaene, S. 2022. A language of thought for the mental representation of geometric shapes. *Cognitive Psychology*, 139: 101527.

Spelke, E. S.; and Kinzler, K. D. 2007. Core knowledge. *Developmental science*, 10(1): 89–96.

Team, G.; Anil, R.; Borgeaud, S.; Wu, Y.; Alayrac, J.-B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A. M.; Hauth, A.; et al. 2023. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.