When Teachers are Lazy

(Tentative Title)

MT4H International Workshop, Valencia, Spain
January 14, 2024

Diogo Nuno Freitas

PhD Student at the University of Madeira, Portugal
diogo.freitas@staff.uma.pt

Joint collaboration between Brigt Håvardstun, Cèsar Ferri, Dario Garigliotti, Jan Arne Telle and José Hernández-Orallo.

How did it Begin?

Research Question:

How effectively can GPT models identify hand-drawn concepts by analyzing stroke coordinate data?

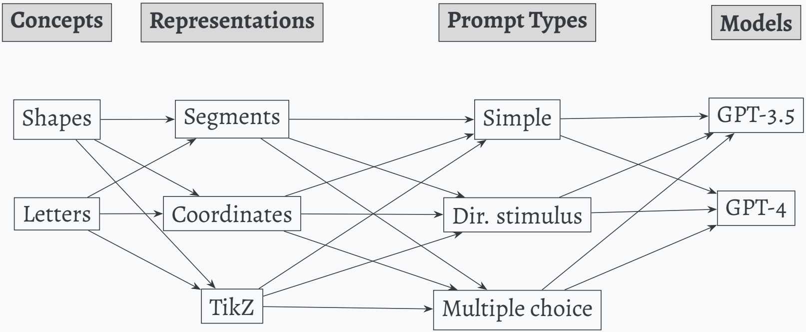
The hand-drawn concepts were to be extracted from the Google *Quick, Draw!*[1] dataset.

- Publicly available.
- 345 concepts (e.g., *apple*, *The Mona Lisa*, *pizza*).
- Stroke coordinates for 40M+ moderated drawings.

[1] quickdraw.withgoogle.com

How did it Begin? (cont.)

First, we conducted a basic experiment:



How did it Begin? (cont.)

The prompt structure employed for **segments**, incorporating a **directional stimulus**, is as follows:

```
You will be provided with a set of line segments of a shape.

Each line segment is represented as [ (x0, y0), (x1, y1) ], where (x0,
y0) is the starting coordinate, and (x1, y1) is the final coordinate.

The line segments are given below, delimited by triple backticks:
```{segments}```

Your task is to identify the polygon or letter represented by the figure
based on the hint.

Hint: Possible polygons are: Triangle, Square, Rectangle, Pentagon,
Hexagon, Octagon, Parallelogram, Right arrow, Diamond, Trapezoid or Star.
Possible letters are: A, E, I, O, U.
```
```

How did it Begin? (cont.)

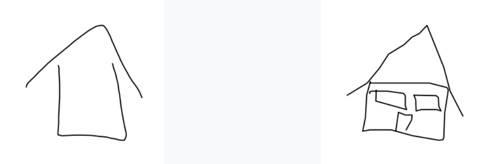
Table: Accuracy of the GPT models in identifying "easy" concepts.

Concept	**GPT-4**	**GPT-3.5**
Square	100%	100%
Triangle	94%	100%
Pentagon	89%	89%
Hexagon	89%	83%
⋮	⋮	⋮
Parallelogram	0%	0%
Right arrow	0%	0%
A	79%	57%
E	43%	7%
I	36%	14%
O	7%	0%
U	0%	0%

The most effective method involved using either **segments or TikZ** with the prompting technique that presents **multiple choices** (72% $\leqslant$ avg. acc. $\leqslant$ 88%).

Concept's Complexity

We focused on the *Quick, Draw!* dataset. In this dataset, we assume that the **complexity of a drawing is related to the number of hand-drawn strokes it contains**.



(a) 2 strokes (b) 5 strokes

Figure: Two hand-drawn representations of the concept *house*.

Using the number of strokes data, we can **sort concepts** and their hand-drawn images by their level of complexity.

A (Potential) Machine Teaching Framework

Research Question:

How many strokes are minimally required for GPT to identify the concept in a hand-drawn representation?

We thus define the teaching size (TS) of a given concept c as

$$\mathrm{TS}(c) \approx \min_{w \in Q : L_m(R(w)) = c} size(w), \tag{1}$$

where R is a representation of w (which could be either stroke coordinates [text-based] or an image[visual-based]), and *size* is a function that, e.g., returns the number of strokes of a given hand-drawn representation.

The Experiment

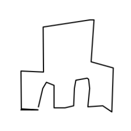
We started by categorizing each hand-drawn image from the *Quick, Draw!* dataset into a **bin according to its level of complexity**.

For every bin, we then **randomly select 50 hand-drawn representations** from the dataset.

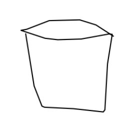
For every hand-drawn image ($\approx 345 \times 10 \times 50 = 172\,500$), we evaluated whether the given representation was **adequate for the learner (i.e., GPT) to identify and learn the concept**.

In addition to getting the TS for each concept, we can examine how **changes in complexity impact the learning accuracy**.

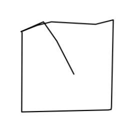
The Results



(a) Car

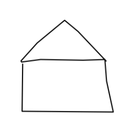


(b) Cup

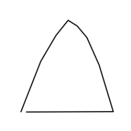


(c) Envelope

(d) Golf club



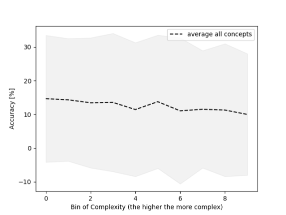
(e) House



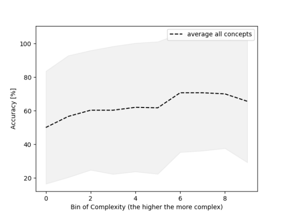
(f) Triangle

Figure: Minimal hand-drawn representations of a subset of concepts learned by the learner. (Representation as strokes coordinates.)

The Results (cont.)



(a) Text-based



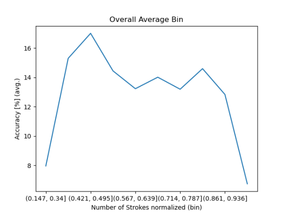
(b) Visual-based

Figure: Comparison, in terms of complexity, between the two representations.

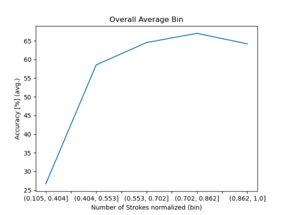
Concept complexity: *line > banana > triangle > square > envelope = house > … > car > guitar > butterfly > piano*

Concept complexity: *line > stairs > triangle > golf club > square > banana > … > candle > airplane > cup > apple*

The Results (cont.)



(a) Text-based



(b) Visual-based

Figure: Comparison, in terms of complexity, between the two representations.

This behavior can be, to some extent, **similar to human identification capabilities**.

The Results (cont.)

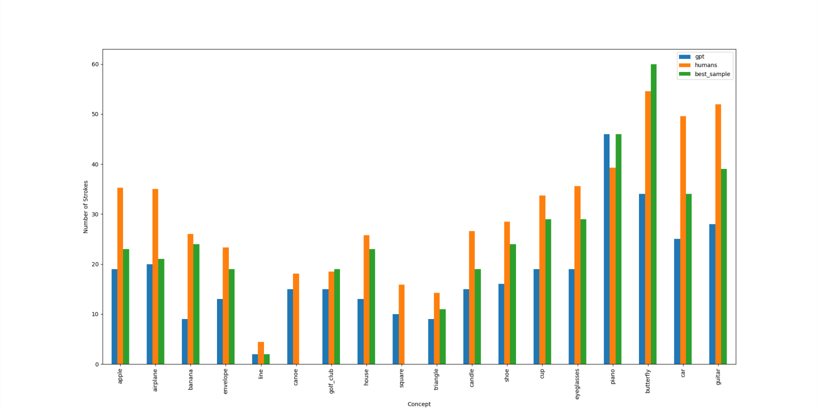


Figure: Comparison between the number of strokes used by humans versus the number of strokes the learner needed to identify a concept (text-based).

The Final Research Question

The previous results pose the following question: "**How can GPT be used to understand fundamental teaching questions?**".

(Final) Research Question:

How intrinsically difficult is teaching a concept based solely on its shape?

The Final Machine Teaching Framework (cont.)

To answer this question, we can use the teaching size that we discussed earlier:

$$\mathrm{TS}(c) \approx \min_{w \in Q : L_m(R(w)) = c} size(w), \tag{2}$$

where R is a representation of w, either an image $\mathrm{IMG}(w)$ or the segments given by $\mathrm{RDP}_\epsilon(w)$[2].

[2] Ramer–Douglas–Peucker algorithm.

A Note on the Use of GPT

Assume the teaching size would be given as follows:

$$\mathrm{TS}(c) = \min_{w:L(w)=c} \mathit{size}(w) \tag{3}$$

and that the learner would be Bayesian posterior:

$$\begin{aligned} L_p(w) &= \arg\max_c p(c|w) = \arg\max_c \frac{p(w|c)p(c)}{p(w)} \\ &= \arg\max_c p(w|c)p(c) = \arg\max_c p(w,c) \end{aligned} \tag{4}$$

or a Bayesian likelihood estimator:

$$L_l(w) = \arg\max_c p(w|c). \tag{5}$$

Since $p(w|c)$, $p(c)$, and $\mathrm{TS}(c)$ are unknown, and we have a poor estimation of $p(w,c)$, we must use a proxy for L (thus, L_m).

The Experiment Algorithm

procedure LAZYTEACHER(c, n), where c is a given concept and n the number of samples
    $D_{\text{raw}} \leftarrow \text{DownloadRawData}(c)$
    $D_{\text{filtered}} \leftarrow \{d \in D_{\text{raw}} \mid d.\text{recognized} = \text{True}\}$
    $D \leftarrow \text{Sample}(D_{\text{filtered}}, n)$
    $D_{\text{simple}} \leftarrow \{\text{RDP}(d, 2) \mid d \in D\}$
    $P \leftarrow \text{ObtainPrototypes}(D_{\text{simple}})$
    $\text{TS}_{\text{coord}} \leftarrow \infty$
    $\text{TS}_{\text{img}} \leftarrow \infty$
    for each prototype $p \in P$ **do**
        $\epsilon \leftarrow 2$
        $p_{\text{simple}} \leftarrow p$
        repeat
            $\hat{c}_{\text{coord}} \leftarrow \text{GPTPrompt}(p_{\text{simple}}.\text{coordinates})$
            if $\text{match}(\hat{c}_{\text{coord}}, c)$ **then**
                $\text{TS}_{\text{coord}} \leftarrow \min(\text{TS}_{\text{coord}}, |\text{Segments}(p_{\text{simple}})|)$
            end if
            $\hat{c}_{\text{img}} \leftarrow \text{GPTPrompt}(p_{\text{simple}}.\text{image})$
            if $\text{match}(\hat{c}_{\text{img}}, c)$ **then**
                $\text{TS}_{\text{img}} \leftarrow \min(\text{TS}_{\text{img}}, |\text{Segments}(p_{\text{simple}})|)$
            end if
            $\epsilon \leftarrow \epsilon + 1$
            $p_{\text{simple}} \leftarrow \text{RDP}(p, \epsilon)$
        until $\text{CannotSimplifyFurther}(p_{\text{simple}}, \epsilon)$ ▷ (i.e., when all segments only have two coordinates)
    end for
    return $\text{TS}_{\text{coord}}, \text{TS}_{\text{img}}$
end procedure

Prototypes (Possible Approaches)

The aim is to obtain minimal hand-drawn representations that are still sufficiently detailed to be **representative of the concepts they illustrate**.

We pretend to explore **different methods**, such as:

- **Mean shift clustering**[3] using the latent representation obtained from a convolutional autoencoder.
- **CLIP (Contrastive Language–Image Pre-training)**[4], and select the top-n highest CLIP scores.

[3] Georgescu, Shimshoni, and Meer, "Mean shift based clustering in high dimensions: A texture classification example".

[4] Radford et al., *Learning transferable visual models from natural language supervision*.

Conclusion

- **Machine Teaching Framework:** We established the Teaching Size as the minimal number of strokes necessary for a learner to recognize a given concept.
- **Algorithm:** We developed the algorithm to minimize the number of strokes (RDP) within a multimodal learning environment.
- **Research Direction:** (1) How can GPT be used to understand fundamental teaching questions? (2) How intrinsically difficult is teaching a concept based solely on its shape?