LOGO-ISTICS: Computer Generated Logos

A Capstone Project

by

Samuel Parsons

Submitted to the Faculty of the

Department of Software Technology

in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE

in

DATA SCIENCE

Champlain College

2023

ABSTRACT

The logo is a visual representation of a company's values and products. Through imagery/stylized text a company, like Nike, can make a logo that will invoke emotion in a consumer to become a lifelong customer. The design process of creating the perfect logo for a business is a long and arduous task. A strong well-designed logo can make all the difference in the world for a business and yet many do the bare minimum in designing their logo. This project aims to automate the process of designing a logo by using stable diffusion to create an AI-generated image.

Table of Contents

CHAPTER 1: INTRODUCTION (2-3 pages) 8

Problem Statement 8

Purpose Statement 8

Context 9

Significance of Project 11

CHAPTER 2: RELATED WORK (3-5 pages) 12

Literature Overview 12

Software 17

CHAPTER 3: METHODS (5-7 pages) 18

Design 18

Frameworks 19

Algorithms 19

Analytical Methods 20

Features 20

Test Plan 22

Criteria and Constraints 22

CHAPTER 4: RESULTS (2-4 pages) 24

Final User Interface 24

Analytical Results 25

Testing Results 26

Creative Results 26

CHAPTER 5: CONCLUSION (2-4 pages) 28

Challenges and Solutions 28

Future Work 29

Project Importance 30

REFERENCES 31

List of Figures

Figures

1. Example of early DALL-E images 9
2. Example of a DALL-E 2 take on Van Gogh 10
3. LOGO-istics Flow Chart 18
4. Example Google Drive output directory 24
5. Previous search method results for the query, "Chauncey the Beaver baseball" 25
6. Current search method results 25
7. Champlain College logo in different artist styles 27
8. Chauncey the Beaver in various promotional images 27

CHAPTER 1: INTRODUCTION (2-3 PAGES)

Problem Statement

In recent years there has been an effort to automate the design process of a logo using "artificial intelligence." These applications are often lacking in depth of design and really any usage of true artificial intelligence. With computer vision software like stable diffusion, real artificial intelligence can be used to design logos. This is where our project hopes to combine the artistic vision of graphic designers with artificial intelligence to create a model that generates visually pleasing logos.

Purpose Statement

Logos are a major part of a brand's identity and designing them is an incredibly difficult task for any graphic designer. The design must reflect the spirit of the company while accounting for the specific requirements that a non-artistic executive is demanding. The perfect logo is a combination of artistic vision and company image/values.

A well-designed logo can make a company or a brand stand out from its competitors. There is no better example of this in action than the Nike Swoosh. The Swoosh was designed by a graphic design student in 1971 for 35$ and has set the standard for what all logos should look like and the purpose that they should serve. What makes the Swoosh so special is that it's so incredibly simple and yet invokes such a strong sense of speed and athleticism which is what

Nike is all about. However, not every company is going to be handed a Swoosh of its own on a silver platter like Nike.

The goal of this project is to automate the long and arduous process of re-designing a logo by using neural networks to generate logos. This will be done using stable diffusion to generate alternative variations of sample images and then use those to train our model on.

Context

On January 5, 2021, DALL-E was released by OpenAI which at the time was the best artificial intelligence image generator. The DALL-E model quickly became popular both inside and outside of the machine-learning world as social media users began posting the images that they generated. This sparked a conversation on what is art and can AI-generated images be considered art. Even though the images produced by the early DALL-E model were not the best people were still flocking to use it while others were rushing to improve on it.

Fig 1.1 Example of early DALL-E images.

Fig 1.2 Example of a DALL-E 2 take on Van Gogh

The grounds for the research of DALL-E and this project are based on the work of Elman Mansimov and Emilio Parisotto who in 2015 published a paper titled GENERATING IMAGES FROM CAPTIONS WITH ATTENTION that would kickstart the age of AI-generated images. The key advancement in this paper is that their model was able to generate images from captions that were not in the training set meaning the model was able to produce images of things that it had not previously seen. From there, researchers have honed in on several ways to improve this idea with better data, more powerful GPUs, transformers, and better ways to associate words from images.

There is so much in the field of computer vision and image generation to explore and we are only at the beginning stage of what can be accomplished. However, with this kind of technology, there are obvious ethical concerns that researchers need to take into account when pushing the envelope.

Significance of Project

The intersection of artificial intelligence and art is currently one of the hottest topics in machine learning. Most areas of traditional image generation have been explored and heavily documented with source code. This project strays from the traditional scope of AI art and looks at how a business could use this technology to automate graphic design processes. Graphic design provides more traditional value for a company than regular image generation can. Every large company uses graphic design in some way making it a very significant area for image generation to explore. The value that this project brings to the machine learning community is huge as it lays the groundwork for other design work to be automated. The hope of this project is that the processes/methodologies used here can be built upon and used in projects with larger graphical design goals. Extending image generation into graphic design would be a real accomplishment and a huge step forward in the space of image generation.

CHAPTER 2: LITERATURE (3-5 PAGES)

Literature Overview

The image generative space of machine learning was kicked off by Google's DeepMind research team in 2015 with their *DRAW* model. *DRAW, Deep Recurrent Attentive Writer,* [1] was revolutionary in the field of image generation as the results from this recurrent neural network (RNN) were the first of its kind in creating images that were indistinguishable from the human eye. The model was trained on the Street View House Number dataset which is a common image machine-learning dataset of street house numbers and from that dataset, they were able to produce fake house numbers. The key advancement in this paper that was used in later works and this project was two parts "an encoder network that compresses the real images presented during training, and a decoder that reconstitutes images after receiving codes." (Gregor et al., 2015)

The next major step in artificial image generation was a paper released by researchers from the University of Toronto in 2016 titled *Generating Images from Captions with Attention* [2]. This paper made several key advancements including caption-based image generation and that has been since built on greatly in later models like DALL-E [3]. The alignDRAW model splits the input caption into sequences of words and then references known image compositions to output a similar composition to what is known. The model iteratively repeats this process and then sharpens the image to make a somewhat recognizable but definitely identifiable artificially generated image. This process is the foundation for most image generation models including this project.

The most groundbreaking paper in the space of AI image generation was *Zero-Shot Text-to-Image Generator* [3] which was the release paper that came along with the release of DALL-E. The core question behind this paper was, “Could dataset size and model size be the limiting factor of current approaches?” (Ramesh et al., 2021) The current approaches in question were from GPT-3 model developed by OpenAI which uses “in-context learning” to determine the next word which was then extended into some basic images and sounds. This paper uses a transformer to compact image texts down to a single stream of data for the model to ingest. This is a massive improvement from the previous models as it drastically increases image quality and reduces input parameters by a substantial amount. From this advancement, this process of a single-stream inflow of image-text data has become standardized and drastically approved with newer releases like DALLE-2.

DALLE-2 is the gold standard for artificial image generation and with its release paper, *Hierarchical Text-Conditional Image Generation with CLIP Latents,* [4] came massive improvements to the space. With this release, the team at OpenAI was able to harness the image-capturing qualities of *CLIP* (which is an image-processing model trained on natural language supervision) and used that to create a *CLIP* embedding for the input text prompt and then decode that embedding to produce an image. The process from this paper is helpful in any image generation project as it establishes an easy and effective way to encode images and then decode them into variant images.

For the technical aspect of this project, we learned a lot from Kris Kashtanova of "AI Tutorials" [5] who runs an AI tutorial YouTube page. Her videos played a key role in understanding the more advanced sections of code and software used in image generation. The main takeaway from her tutorial series was learning how to train image generation software on your own photos locally. This is a major portion of the project is training on local images in order to teach the model logos.

Another inspiration for this work was a project by GitHub user *jn2clark* named *Combining stable diffusion with semantic search to tag and query images at scale*. [6] What this project did was use stable diffusion, a competitor to DALL-E 2, to generate a dataset of 100,000 hot dogs and then leverages Marqo, a tensor search model, to index the data and query it. This project used a lot of very interesting techniques like generating one's own image dataset and then using that dataset to query for desired results.

With the scope of the model being very small it is important that the computational requirement of training the model be small enough to be run on a single high-quality GPU or a small cluster of online cloud GPUs. The paper, *High-Resolution Image Synthesis with Latent Diffusion Models*[7] provides a new method to significantly reduce the computational cost associated with training a diffusion model(DM). "To enable DM training on limited computational resources while retaining their quality and flexibility, we apply them in the latent space of powerful pretrained autoencoders. In contrast to previous work, training diffusion models on such a representation allows for the first time to reach a near-optimal point between complexity reduction and detail preservation, greatly boosting visual fidelity." (Rombach, et al) By using learned latent space through pretrained autoencoders, they were able to successfully reduce how much spatial information the model learns. In simpler terms, they provide essentially "lossy" training on the image dataset compared to the "lossless" training of some of the larger models like DALL-E.

A huge inspiration for undertaking this project was in part from a YouTuber called Two Minute Papers who produces videos on release papers for major AI advancements. Almost every paper discussed above has been covered by him which is a really great resource to use when first trying to understand image diffusion. Videos like "OpenAI's Image GPT Completes your Images with Style!" [8] (Zsolnai-Fehér, Károly) In this video he goes over some applications of image outpainting where the model is trained to "complete the picture" on the inputted image. This is very similar to what would be required with logo generation in that text or colors would have variations depending on the type of logo. For example, image outpainting in logos could look like a college's mascot holding different objects to resemble different clubs and organizations.

The most recent release in the ongoing competition for the best image generator was Stable Diffusion 2.0. Along with upgrades to performance and quality, Stable Diffusion 2.0 introduces Depth-to-Image Diffusion Model which functions very similarly to the desired effect that the model used in this project would have using sample logos to generate alternative logos.

"Depth-to-Image can offer all sorts of new creative applications, delivering transformations that look radically different from the original but which still preserve the coherence and depth of that image." [9] (Cusick, 2022) Basically depth-to-image would be useful in logo generation as it is able to learn the structure of the image while radically changing what the image is of. This is very useful for logos as a general logo shape can be iteratively used to generate alternative logos for different icons/mascots.

One of the most groundbreaking advancements in the deep learning field has been the transformer which has revolutionized how large models are trained. In pretty much every image generator model, a transformer is used. "In Transformers, the input tokens are converted into vectors, and then we add some positional information (positional encoding) to take the order of the tokens into account during the concurrent processing of the model." [10] (Hashesh, 2022) Where recurrent neural networks and convolutional neural networks are lacking in effectively learning positional information from the text, transformers exceed. Transformers have since then been extended into learning image and text relationships which is a very important concept in generating logos. Designing a logo using text would have to make extensive use of a transformer so that the model can understand what the input text is asking and then output an image that it thinks closely relates to the request.

Software

The current era of generative images using machine learning has produced many fantastic projects and libraries in an open-source format. Some major research projects with corresponding libraries are stable diffusion by Stability AI and CLIP/DALLE2 by OpenAI. These projects provide massive documentation and improvements to the image generative process. Stable diffusion, in particular, has the most open-source code including training weights which is very crucial in cutting down training time. One project that has been a massive inspiration for this project *Combining Stable Diffusion with Semantic Search to Tag and Query Images at Scale*, which is a project exploring stable diffusion's mass generative abilities and then querying the resulting dataset. This process could be used in graphical design to mass-produce logos and then query them for specific criteria.

CHAPTER 3: METHODS

Design

The approach that we took in making this project was to use the power of stable diffusion to generate logos. Using automated captioning tools on sample images to use as a training set, we then used stable diffusion in order to train a logo checkpoint. From that checkpoint, a massive list of sample prompts gets passed into a text-to-image stable diffusion model in order to produce a bunch of sample variation logos to store in an output image dataset. The resulting image dataset is domain-based meaning that the dataset created is only specific to the company/organization that is using it. From this dataset, the desired logos can be queried.

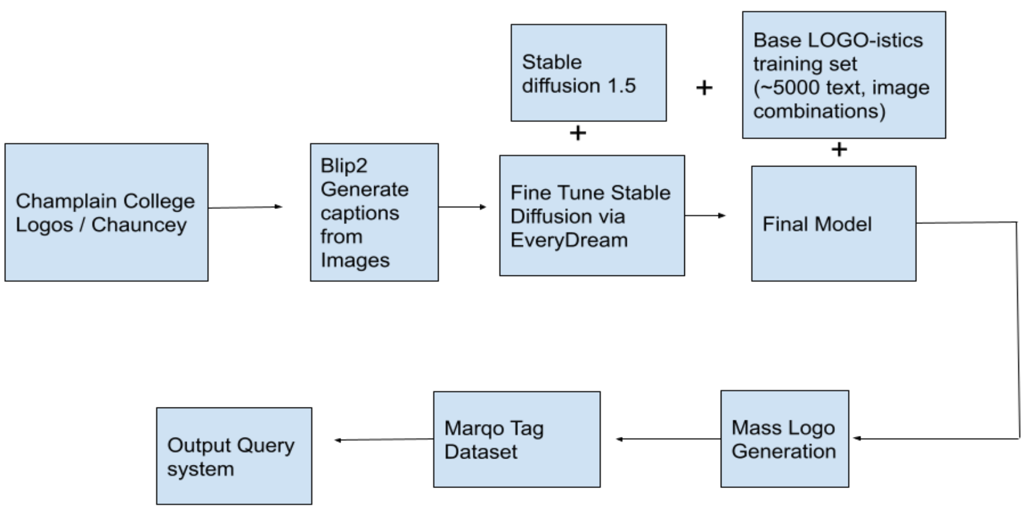

Fig 3.1 LOGO-istics Flow Chart

Frameworks

The main import for this project is stable diffusion which offers several different pretrained models for which the logo-generating model can be built off of. Using a pre-trained model drastically cuts down on the computational, economic, and environmental constraints of this project. The text-to-image diffusion model uses an inputted text to generate variants which is exactly the desired input to the output of logos for this model. Python is the most common coding language used in any data science/machine learning project so there really was no other framework to work under for this project.

Algorithms

The main algorithm used in this project is the generative image function which takes sample images and then generates variants of the logo and stores them in a database. This algorithm will leverage stable diffusion to skip the lengthy and expensive pre-training process. The model does not fully skip the training process, however. The model is then trained on logos to be fine-tuned to make logos. Given the desired number of outputs, the model will store that many outputs in the domain database. The other main algorithm queries the database for images based on text prompts. This will be done by indexing the images using Marqo and then querying the database based on user input. The result of this query will be the output logo or logos.

Analytical Methods

The analytical methods of this project revolve around model performance and comparisons between predicted outcomes and requested outcomes. At a high level when generating an image from text or even from another image, there is a predicted output vector that is generated based on what the model thinks the user is requesting, and then there is the actual output vector that is the closest image generated to what the predicted output vector was. The next part of the project that heavily uses analytics is tagging images using Marqo. What Marqo does is effectively convert images to tensors from which to be searched. The input query to search the database will also be converted to a predicted tensor for which to search the dataset for the closest match.

Features

The main feature of LOGO-istics is a BLIP feature that inputs images and outputs a text caption to go along with the input image in training. BLIP leverages zero-shot learning which is a type of machine learning where the model is trained to detect information that can be inferred from what it already knows. This method of learning is much more advanced than supervised and unsupervised learning as it's trying to teach the model how to understand things that it will be introduced to that it hasn't seen before. BLIP also uses a Large Language Model or LLM in conjunction with an image encoder and query transformer to translate an image into text. For this project, BLIP was used in both its caption generation and visual questioning and answering aspects. The texts derived from BLIP are used in training with the images to automate the process of mainly captioning the images for learning image-text associations.

Another feature of our project is a text-based query system that searches through the dataset for relevant features based on the input. The dataset from which the model will be queried will be added as more sample images are introduced and then tagged using Marqo thus increasing the robustness of the project. Using stable diffusion's text-to-image to generate alternative logos of the sample logo based on the positions and prompted text is the core component of the main feature of the project. The querying system will be used after the user trains the model on their logos. The user can input a query like "college mascot logo with a Halloween mask" and the model with that input will use the Marqo encoder to generate a predictive tensor for which to search the tagged dataset. The model will then output the closest match to the predicted tensor of the input request.

A potential large feature that could be added in the future would be adding a graphical user interface where the user goes through some design choices like shape or text before the logo generation occurs. Another feature could be implementing inpainting which is the combination of both image-to-image and text-to-image diffusion models. By selecting a region of an image to edit, one can input a text request to regenerate that area of the image according to the surrounding area and composition of the image. Inpainting in logo generation could have some very interesting applications for finer tuning one's logo and using a base logo from which to generate inpainting variations.

Test Plan

Testing will be split up into two parts one being a comparison threshold a generated logo has to meet in order to be a valid result. The other portion is the eye test where a generated logo has to be sufficiently pleasing to the eye. For the first testing part, the sample input will have a stable diffusion embedded value with which to be compared to the output's embedded value. A testing threshold will be set that the output value must reach before making it to the next half. If this testing step fails it would be due to the query resulting in an image that does not resemble anything near the sample image. The other half of the test is the more difficult test as making a logo that looks good is hard enough as is by hand. Passing the eye test of being indistinguishable from human-made is really the ultimate goal for any image generation project so being able to produce logos that are of graphic design quality would be an outstanding success.

Criteria and Constraints

As of writing, the current legal precedent for training models on copywritten material falls under the doctrine of "fair use." This was established in the ruling of *Author's Guild v. Google* by the United States 2nd District Court in 2015 and the Supreme Court declined to hear the arguments making this ruling the largest legal precedent set in favor of machine learning. The ability for researchers to use copyrighted material is so crucial in building high-performing models with general applications. However, in the ruling, there was one key line that puts projects like LOGO-istics on fairly shaky legal ground. “Chin stated that “Google Books enhances the sales of books to the benefit of copyright holders”, meaning that since there is no negative influence on the copyright holder it does not violate fair use.” [11] (Stewart, 2019) The reason why this line is so important is that training on copyrighted material falls under “fair use” if it benefits the copyright holders. Projects like LOGO-istics generally serve a negative interest for copyright holders of the data on which the model is trained. This may present grounds in the future for which a lawsuit by the copyright holders as their work is being automated upon.

The environmental impacts of deep learning projects are very overlooked in the machine learning space. Most discussion around the energy usage constraint in machine learning is about how many high-powered graphic cards one needs to train a model on. Many overlook just how long GPU computing hours are the training time of some of the larger deep learning models. As researchers from the University of Massachusetts found, “The sum GPU time required for the project totaled 9998 days (27 years). This averages to about 60 GPUs running constantly throughout the 6-month duration of the project.” (Strubell Et al, 2019) [12] The amount of energy consumed by training these models is vast and the results from these lengthy experiments could only produce marginally better results which begs the question of necessity in training models. By leveraging stable diffusion it removes the necessity of training a massive beefy model to learn text and image associations. Not only will this drastically reduce the environmental impact it will also reduce the economic constraint of the project which is the cost of cloud GPUs needed to train the model.

CHAPTER 4: RESULTS

Final User Interface

The final user interface is a guided GitHub and a labeled google drive folder where the images are labeled by BLIP caption generated from the *BLIP Mass Output* notebook. In order to query this directory one needs to use the *Search for Images* notebook to find desired images. The GitHub repository features directions and the code necessary to BLIP caption images, mass produce stable diffusion images, re-BLIP the output images, and then search the image directory. The output captions contain information about what the image is, the art style, color, action, and place of the image.

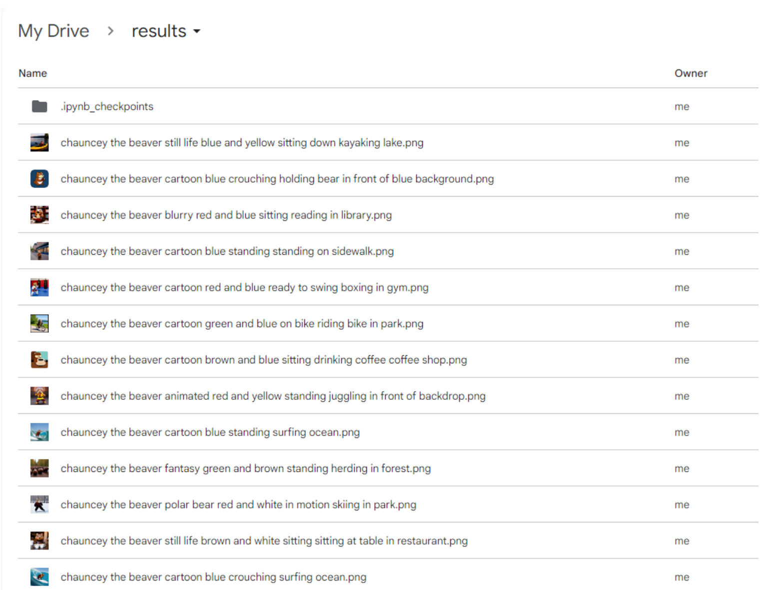

Fig 4.1 Example Google Drive output directory

Analytical Results

The only real quantitative analysis that appears in LOGO-istics is in the search functionality of the project. This quantitative analysis essentially calculates how closely the input query matches the output results. This is done by using Scikit-Learn's cosine similarity function which calculates the similarity between the query and the captions in the output directory. The score is heavily influenced by the "rare" words in the query and in the directory. For example, the words "Chauncey the Beaver" are weighted much less in calculating similarity than a word that shows up fewer like "biking." This method of calculating best search results based on rare words does leaps and bounds better than the prior frequency search method that was used.

```
{'chauncey', 'the', 'beaver', 'baseball'}
{'at', 'computer', 'looking', 'of', 'something', 's', 'screen', 'in', 'cartoon', 'the', 'beaver', 'he', 'chauncey', 'blue', 'like', 'front', 'holding'}
{'reading', 'in', 'cartoon', 'the', 'beaver', 'chauncey', 'park', 'green'}
{'motion', 'orange', 'in', 'and', 'cartoon', 'the', 'beaver', 'chauncey', 'playing', 'blue', 'baseball', 'on', 'field'}
chauncey the beaver,cartoon,green,reading,reading,in park.png : 0.4166666666666667
chauncey the beaver,cartoon,orange and blue,in motion,playing baseball,on field.png : 0.38461538461538464
chauncey the beaver,cartoon,blue,like he's holding something,looking at computer screen,in front of computer screen.png : 0.20833333333333334
```

Fig 4.2 Previous search method results for the query, "Chauncey the Beaver baseball"

```
for similarity, file_name in results[:1]:
  #prints the top match to query
    print(f"{file_name}: {similarity:.4f}")

chauncey the beaver,cartoon,orange and blue,in motion,playing baseball,on field.png: 0.4600
```

Fig 4.3 Current search method results

Testing Results

For any project where the deliverable is creative work the ultimate test of the quality of the product is the eye test, whether or not the work looks good. LOGO-istics mass produces images that naturally present their own advantages and disadvantages to this testing method. The advantage of having a ton of images is that there is bound to be at least one image per prompt that could pass the eye test. The disadvantage is that the majority of images that are produced would not pass this test.

To gauge the eye test of LOGO-istics to a non-bias viewer, LOGO-istics presented some results from the mass production of Champlain College promotional material to an audience of students and faculty. The overall reception was very positive and receptive to the idea of artificially generated logos and promotional material. However, there were some key recurrent criticisms that should be noted. The main criticism is that the generated Chauncey the Beaver did not resemble a beaver enough. This was due to the Chauncey concept being trained as "Chauncey the Beaver" and the input images were not traditional-looking beavers with buck teeth and black tails. A better training concept name would have just to use Chauncey given the training images.

Creative Results

The main results of LOGO-istics are the creative results from the mass generation script. The logo outputs come in varying shapes and sizes that differ from the base shield shape logo that the model was trained on. For the roughly 80 different logo prompts that went into the mass output, each one has a handful of usable results that could pass the eye test. The main limiting factor for a lot of the results not being able to pass the eye test is due to messed up lettering or random shapes being created in the image which could be removed in post-editing. Overall the results that do pass the eye test are very visually appealing and showcase the potential of a tool like LOGO-istics.

Fig 4.4 Champlain College logo in different artist styles

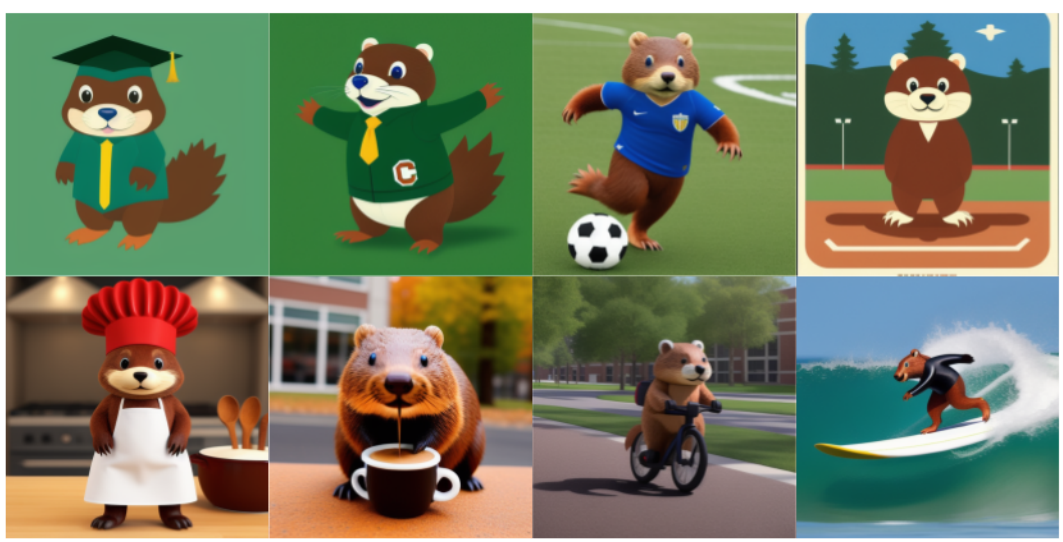

Fig 4.5 Chauncey the Beaver in various promotional images

CHAPTER 5: CONCLUSION

Challenges and Solutions

The main challenge presented throughout this project was obtaining the data used for training my base model as well as fine-tuning the model on Champlain's creative assets. Many logos for companies feature lots of lettering which makes it difficult to train a model on and introduces unnecessary information to the model. The main logo dataset that LOGO-istics is trained on was procured from a massive web-scraped logo dataset. A ton of time was spent on this process of only trying to introduce good information from the logos and not bog down the model with unwanted lettering. Ideally, the dataset to train LOGO-istics on would only use company logos that feature minimal words or letters that don't relate to the brand name.

Another major issue with the data collection for this project was that Champlain College does not have that many creative assets accessible and has no Chauncey the Beaver images in storage. To overcome this obstacle we created our own version of Chauncey using various different images of animated beavers, both real and generated. Ideally, LOGO-istics would be employed solely on a single domain's creative asset so there would be consistency in the design of the character.

One feature that couldn't be fully flushed out was the retagging of the mass outputs using Marqo. Setting up Marqo required using Docker which presented the massive problem of a Marqo image being created and running, however, unable to be used on Google Colab. After some struggles, it was decided that the retagging and search functionality could be done through lower-level means by retagging the outputs with BLIP and then using a Natural Language Processing search method. This search method uses Scikit-learn's TF-IDF Vectorizer which turns text into vectors and places heavy emphasis on words that are rarer rather than words that show up more frequently. This is an appropriate workaround solution to the search functionality that Marqo would have provided because unlike more traditional search problems we are looking for small differences in very similar text.

Future Work

The future of LOGO-istics can be taken in many directions. The promotional material side of the project has massive value to any advertiser. Being able to mass produce promotional material as well as make variations of promotional material to target different demographics simultaneously would revolutionize advertising. The main issue that advertisers are facing these days is a lack of attention from consumers. With more personalized advertising, theoretically, content viewers would be more inclined to tune into the ad rather than out.

The current new AI tool that is in its infancy is text-to-video which takes input text and outputs a generative video. As of writing, text-to-video generates videos that closely resemble stop motion films or early 1900s movies with very little frame rates. In the next couple of years, text-to-video could drastically improve to the point where instead of using LOGO-istics for generated promotional marketing material we could use it to generate real advertisements.

Text-to-video will introduce a series of obstacles to researchers like computing power, data collection, and tagging outputs will be much more difficult.

Project Importance

LOGO-istics matters because it has demonstrated the more creative aspects of image generation outside of traditional AI art. Many people believe that the current state of artificial image generation is unable to produce images that could be used for graphic design work. LOGO-istics has shown that not only can you generate logos you can also generate promotional imaging. Further pushes in similar directions could lead to some really interesting results in work like architecture, interior design, or website design. But as always with models of these types they are only as good as the input data they are trained on which will be a massive obstacle for anybody moving forward in these areas.

Artificial image generation is a fantastic tool but there currently are not many ways to derive value from it outside of commission-based work. Through this project, LOGO-istics has shown that it has derived value for Champlain College. Theoretically, a group of students who want to start a club and need advertising material could approach LOGO-istics to generate promotional material for the club. Services like LOGO-istics could drastically change the advertising industry by making endless promotional material readily available to anyone at any time.

REFERENCES

CITATIONS

1. Mansimov, E., Parisotto, E., Ba, J. L., and Salakhutdinov, R. Generating images from captions with attention. arXiv preprint arXiv:1511.02793, 2015 (First paper on image generation)
2. Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., & Wierstra, D. (2015). DRAW: A Recurrent Neural Network For Image Generation. *arXiv*. https://doi.org/10.48550/arXiv.1502.04623
3. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. *arXiv*. https://doi.org/10.48550/arXiv.2102.12092
4. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv*. https://doi.org/10.48550/arXiv.2204.06125
5. Kashtanova, Kris. "Train Stable Diffusion Easier on Your Own Photos." *YouTube*, Ai Tutorials, 2 Dec. 2022, https://www.youtube.com/watch?v=A3iiBvoC3M8&t=0s.
6. Clark, Jesse. "Combining Stable Diffusion with Semantic Search to Tag and Query Images at Scale." *GitHub*, https://github.com/marqo-ai/marqo/blob/mainline/examples/StableDiffusion/hot-dog-100k.md.
7. Rombach, Robin, et al. "High-Resolution Image Synthesis with Latent Diffusion Models." *arXiv*, 2021, https://doi.org/10.48550/arXiv.2112.10752. Accessed 10 Dec. 2022.

8. Zsolnai-Fehér, Károly. "OpenAI's Image GPT Completes Your Images with Style!" *YouTube*, Two Minute Papers, 1 Sept. 2020, https://www.youtube.com/watch?v=-6Xn4nKm-Qw&list=PLujxSBD-JXgmB1AnewzycdtUtf5YVUyzU&index=10.
9. Cusick, Bill. "Stable Diffusion 2.0 Release." *Stability.Ai*, Stability.Ai, 24 Nov. 2022, https://stability.ai/blog/stable-diffusion-v2-release.
10. Hashesh, Ahmed. "Comprehensive Guide to Transformers." *Neptune.ai*, 14 Nov. 2022, https://neptune.ai/blog/comprehensive-guide-to-transformers.
11. Stewart, M. (2020). *The most important Supreme Court decision for Data Science and Machine Learning*. Medium. https://towardsdatascience.com/the-most-important-supreme-court-decision-for-data-science-and-machine-learning-44cfc1c1bcaf
12. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *arXiv*. https://doi.org/10.48550/arXiv.1906.02243