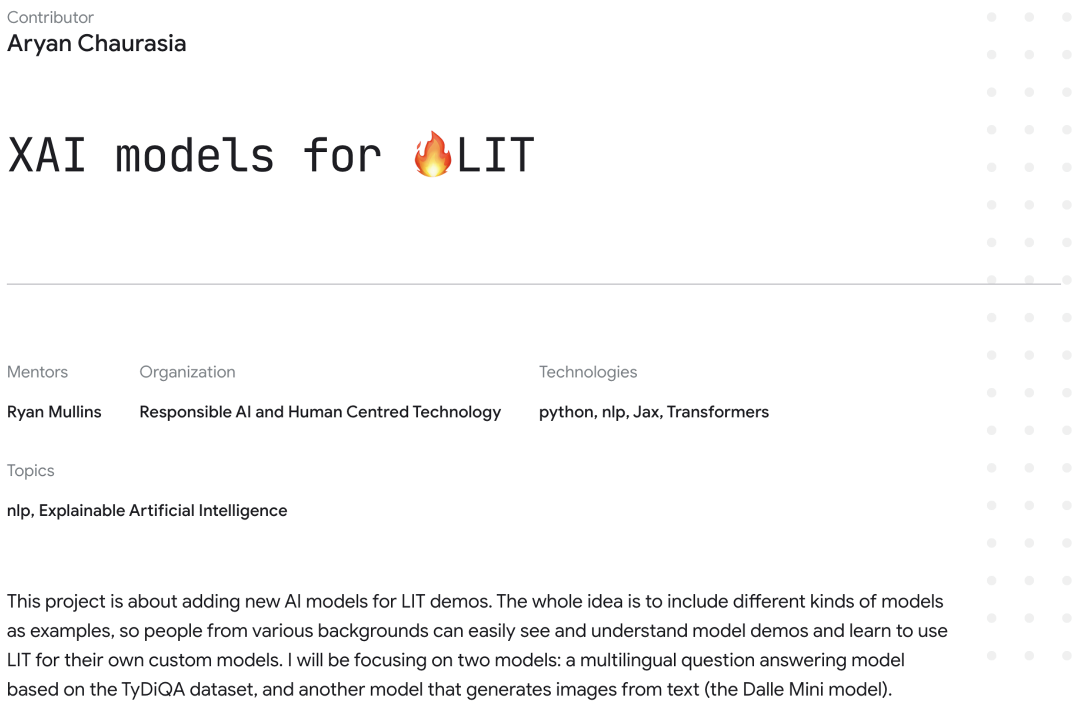
XAI Demos for 🔥LIT



Contributor

Aryan Chaurasia

XAI models for 🔥LIT

Mentors	Organization	Technologies
Ryan Mullins	Responsible AI and Human Centred Technology	python, nlp, Jax, Transformers

Topics

nlp, Explainable Artificial Intelligence

This project is about adding new AI models for LIT demos. The whole idea is to include different kinds of models as examples, so people from various backgrounds can easily see and understand model demos and learn to use LIT for their own custom models. I will be focusing on two models: a multilingual question answering model based on the TyDiQA dataset, and another model that generates images from text (the Dalle Mini model).

Brief

For GSoC 2020, my project was about adding new models to LIT (Language Interpretability Tool). LIT is a tool created by the Google PAIR team specifically to "interactively analyze NLP models for model understanding". I was specifically interested in this project because of my interest in NLP and also getting into software development.

From my viewpoint, working on adding new models wouldn't have required higher experience in software development. At the same time, it would allow me to learn a lot about the development practices and make open source contributions. It gave me an opportunity to make contributions to production-level applications and also get to work and learn under a very experienced and talented mentor.

Token of appreciation

My personal goal, other than completing the project and making open-sourced contributions, was to be able to work with someone who has tons of experience in the field and also receive mentorship which I believe was very valuable.

Before going into the project in detail, I would like to thank my mentor Ryan Mullins who is an excellent mentor and provides great feedback and also code reviews that were both helpful and constructive. He has a wealth of experience and knowledge to share and was always willing to help me learn and grow. Also, he was very patient and supportive even though I think he definitely had tons of other things to do. And he still created a positive and supportive learning environment. Thank you for being an excellent mentor.

Main Tasks

- Multilingual Flax model based TyDi QA dataset: This was the first task for LIT which involved adding a question-answering model based on the TyDi QA dataset.

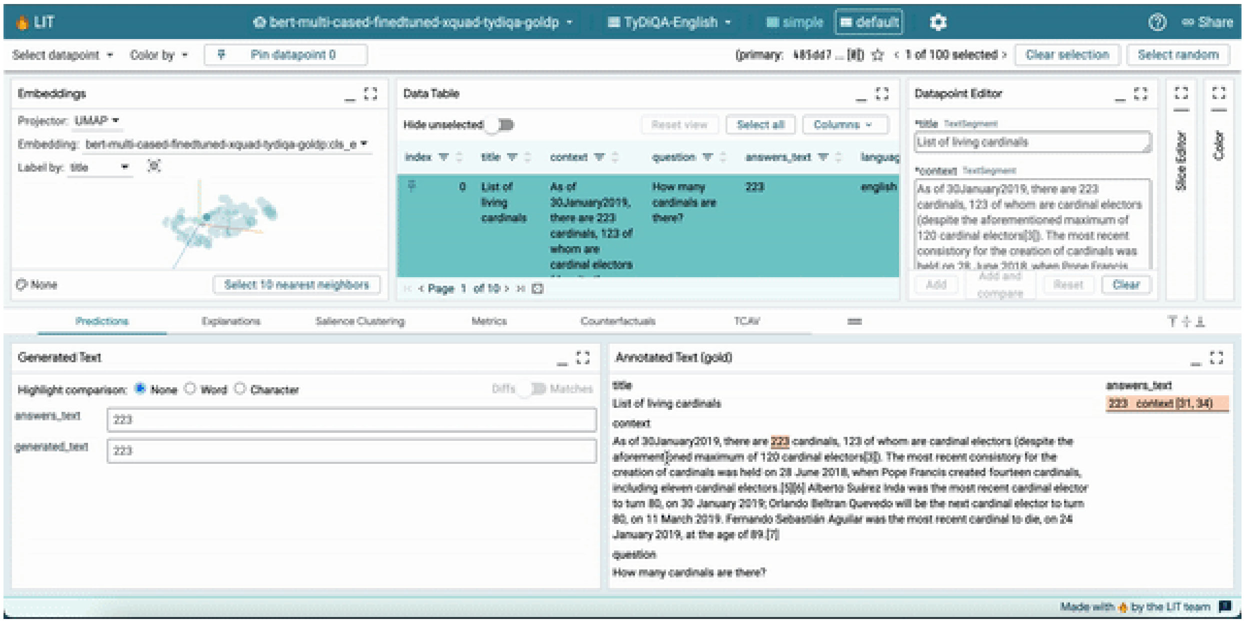
- Text-to-Image Generation Demo: The second task was about adding text to image generation model, and we decided to go with Dalle-mini, which was just released around the same time and had gained huge popularity.

Pull Requests

- Multilingual Demo for the TyDi QA task
- Text-to-Image Generation Demo with Dall-E Mini

Project Highlights

- We had very well-organized weekly meetings, with an agenda and notes for the week being noted. Not only did it help to track progress, but also I was able to discuss any other issues which I faced during the week.
- I started with a question-answering model based on the Tydi QA dataset. After a few initial rounds of code reviews and feedback, the demo was looking good.
- A few issues faced while working on the Tydi QA model were mostly on Exact Match metrics which had to be implemented for the model, and proper rendering of the new data type that this model was using (Multi-Segment Annotations) on the front end. Again, all thanks to the mentor with the help of his expertise, this issue was solved.



Projector: UMAP

Embedding: bert-multi-cased-finedtuned-xquad-tydiqa-goldp:cls_e

Label by: title

Hide unselected

index	title	context	question	answers_text	languag
0	List of living cardinals	As of 30January2019, there are 223 cardinals, 123 of whom are cardinal electors	How many cardinals are there?	223	english

*title TextSegment: List of living cardinals

*context TextSegment: As of 30January2019, there are 223 cardinals, 123 of whom are cardinal electors (despite the aforementioned maximum of 120 cardinal electors[3]). The most recent consistory for the creation of cardinals was held on 28 June 2018, when Pope Francis

Highlight comparison: ● None ○ Word ○ Character

answers_text: 223

generated_text: 223

title

List of living cardinals

context

As of 30January2019, there are 223 cardinals, 123 of whom are cardinal electors (despite the aforementioned maximum of 120 cardinal electors[3]). The most recent consistory for the creation of cardinals was held on 28 June 2018, when Pope Francis created fourteen cardinals, including eleven cardinal electors.[5][6] Alberto Suárez Inda was the most recent cardinal elector to turn 80, on 30 January 2019; Orlando Beltran Quevedo will be the next cardinal elector to turn 80, on 11 March 2019. Fernando Sebastián Aguilar was the most recent cardinal to die, on 24 January 2019, at the age of 89.[7]

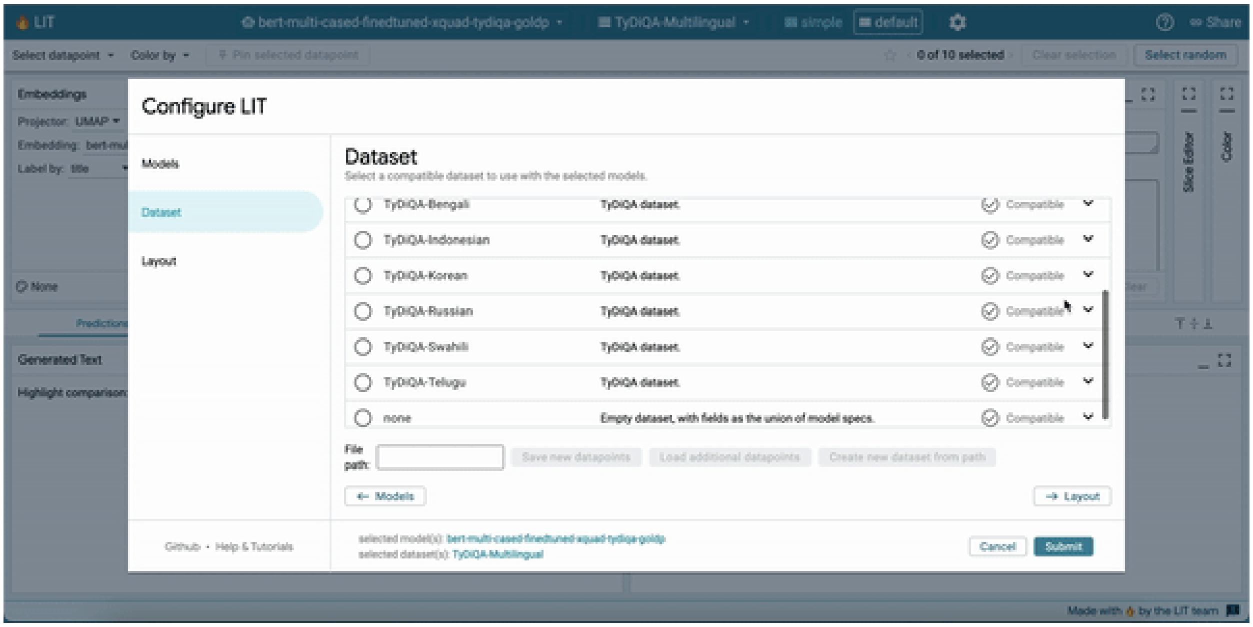
question

How many cardinals are there?

answers_text

223 context [31, 34)

Tydi QA Model on LIT



Dataset

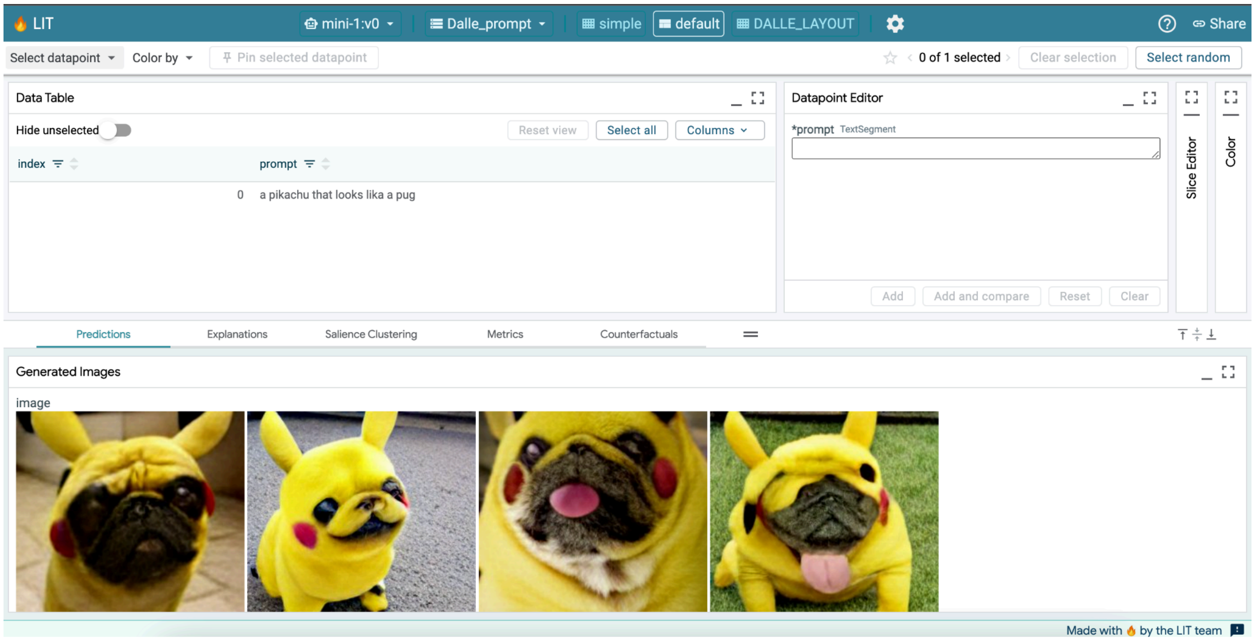
Select a compatible dataset to use with the selected models.

○	TyDiQA-Bengali	TyDiQA dataset.	Compatible
○	TyDiQA-Indonesian	TyDiQA dataset.	Compatible
○	TyDiQA-Korean	TyDiQA dataset.	Compatible
○	TyDiQA-Russian	TyDiQA dataset.	Compatible
○	TyDiQA-Swahili	TyDiQA dataset.	Compatible
○	TyDiQA-Telugu	TyDiQA dataset.	Compatible
○	none	Empty dataset, with fields as the union of model specs.	Compatible

selected model(s): bert-multi-cased-finedtuned-xquad-tydiqa-goldp
selected dataset(s): TyDiQA-Multilingual

Example of other languages loaded into the LIT Module from the dataset.

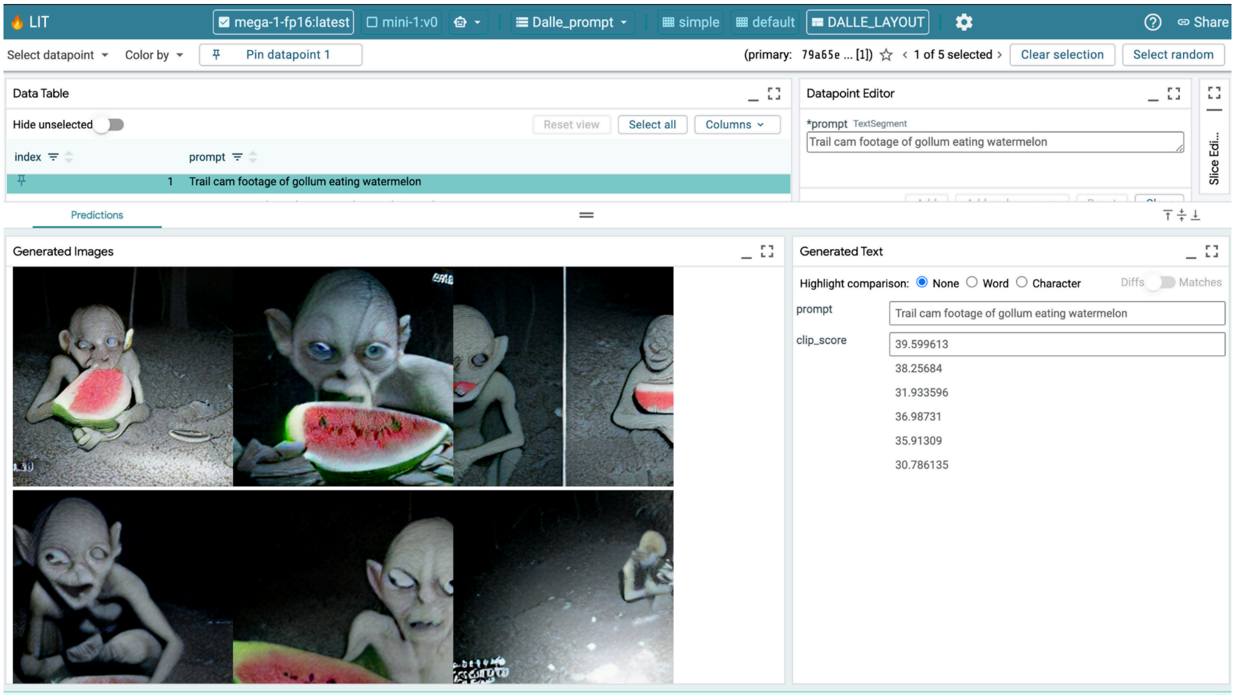
- On the other hand, the Dalle-mini model had a much simpler layout.
- Due to not having integrated graphics, it took me a lot of time to work with the Dalle-mini model since it was mainly using the CPU. To overcome this, I first tried implementing the pipeline on Google Colab using GPU runtime, which made the workflow a lot simpler.



index	prompt
0	a pikachu that looks lika a pug

Initial Example of Dalle-mini on LIT module

- The model looked very simple, so we thought of adding a saliency map for the text prompt. However, as of now, the Dalle-mini model doesn't output hidden states, due to this, including a saliency map in the LIT model is not possible.
- Soon I came across the [CLIP model]() by Open Ai which takes in an image and a prompt and tells how well they match by generating a score. CLIP was also used along with the Dalle-mini in a few examples. I thought it was important to include and something like this was also mentioned in a meeting that my mentor had arranged with other LIT team members.
- After incorporating CLIP into my Dall-mini model and a few rounds of code reviews, my second task was also completed.



index	prompt
1	Trail cam footage of gollum eating watermelon

*prompt TextSegment: Trail cam footage of gollum eating watermelon

Highlight comparison: ● None ○ Word ○ Character

prompt: Trail cam footage of gollum eating watermelon

clip_score: 39.599613
38.25684
31.933596
36.98731
35.91309
30.786135

Dalle-mini demo with CLIP score in LIT module

Conclusion

I believe overall it was one of the best decisions I made when I applied to GSoC. Not only did I make some new friends with other contributors, but I also got to know some very talented people who are way more experienced with domain-specific knowledge. Just getting even time to spend with them was so valuable.