Towards Visualisation Specifications from Multilingual Natural Language Queries using Large Language Models

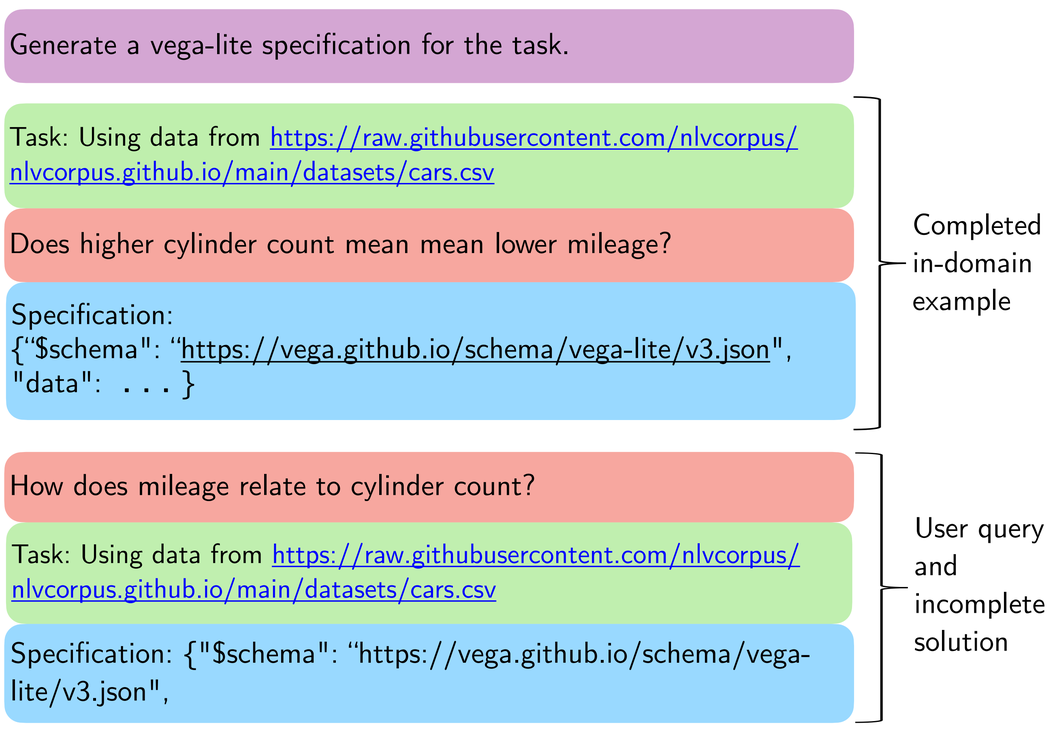
Maeve Hutchinson*, Aidan Slingsby, Radu Jianu, Pranava Madhyastha*
Department of Computer Science, City, University of London
* Corresponding Authors

Recent research into pre-trained Large Language Models (LLMs) suggests that generating data visualisations from natural language is an increasingly feasible technique, but current systems predominantly focus on English language. We demonstrate that a LLM approach can generate visualisations from multilingual queries, a first step towards the development of accessible multilingual interfaces for data visualisation.

Experimental Setup

We use the **dataset** by Srinivasan, et al.[1] of 814 natural language queries, each relating to a specific Vega-Lite visualisation.

We use a **prompt-based learning** approach. This involves engineering a natural language prompt input that contains a task description and a set of valid in-context examples. We pair each query with another example from the dataset and its completed specification.

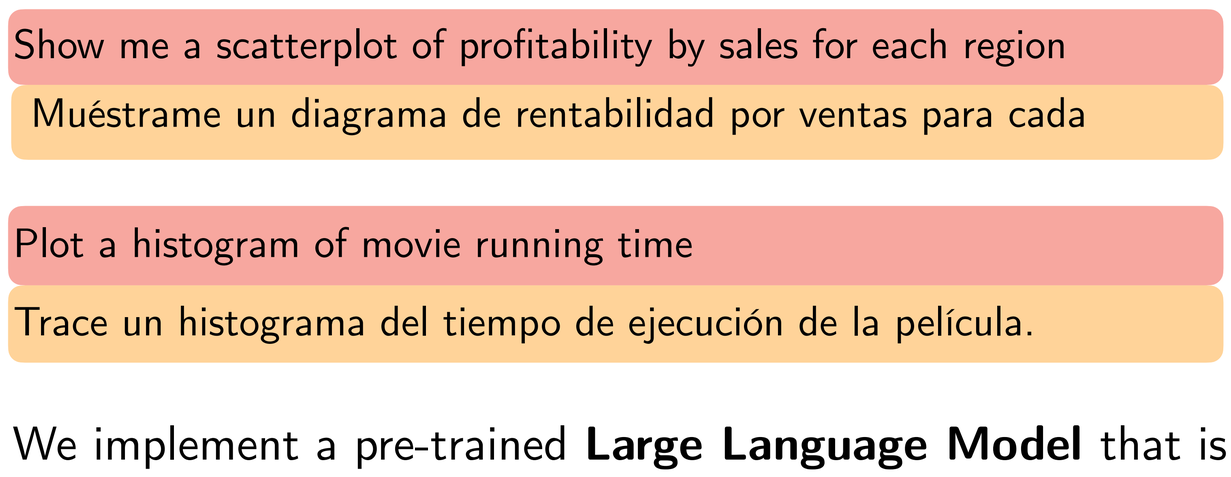


Examples are categorised as:

(a) *in-domain:* semantically similar to the user query, relating to the same dataset

(b) *out-of-domain*: not associated with the same dataset as the user query

For the **multilingual** step, we machine translate each query into Spanish, and pair with an in-domain example in English.

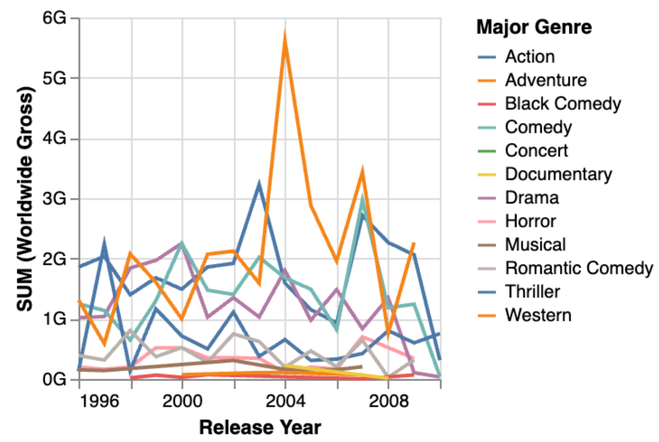
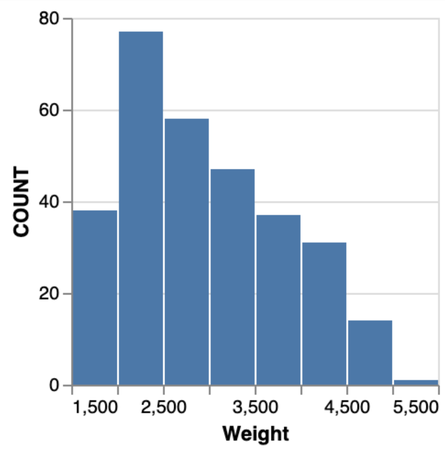


We implement a pre-trained **Large Language Model** that is able to generate text continuing from the input prompt. We use the openly available **BLOOM** model, which was built over data in multiple languages and large GitHub-based code repositories.

Initial Results

We use a series of unit tests to **evaluate** the generated specifications: specifications that are not valid python dictionaries are categorised as *syntax errors;* dictionaries that are do not compile to vega-lite are *validation errors;* and if the compiled chart does not match the target chart it is a *logic error.* The *Pass@1* metric represents the proportion of solutions that successfully match the target visualisation from the corpus.

Prompt Type	Pass@1	Syntax Error	Validation Error	Logic Error
English Out-of-domain	0.12%	4.4%	2.7%	92.8%
English In-domain	81.1%	3.4%	0%	15.5%
Spanish In-domain	78.1%	4.3%	0.12%	17.4%

User Query	Visualisation generated from in-domain prompt
Give me line charts of worldwide gross over year by major genres	
Histograma de automóviles por peso	

These results provide a proof of concept that a pre-trained LLM can produce a Vega-Lite specification from a user query in multiple languages using prompt engineering. Notably, the results for Spanish queries are very similar to English despite having no in-context examples in Spanish.

[1] SRINIVASAN, ARJUN, NYAPATHY, NIKHILA, LEE, BONGSHIN, et al. "Collecting and Characterizing Natural Language Utterances for Specifying Data Visualizations". Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, 1–10. ISBN: 978-1-4503-8096-6. DOI: 10.1145/3411764.3445400