The Vanderbilt University
CENTER FOR
DIGITAL
humanities

GPT-2 Habla Español:

Training a Transformer to Write like Spanish Authors

Caroline Colquhoun and Sahai Couso Díaz | 29 April 2021

Project Description

Our project interrogates and responds to intersecting manifestations of Anglocentrism--in technological developments like artificial intelligence and in digital humanities methods and practices--by seeking to fine-tune Open-AI's GPT-2 natural language model in Spanish--and in the tone of authors of Spanish literature. The aim is that these models will serve as engines for literary and archival discovery, allowing students and researchers to explore well-known authors and their works in novel and creative ways.

Following a series of dead ends and failed attempts, we successfully built on the work of Josue Obregón and Berny Carrera and fine-tuned their (previously fine-tuned) "Small Spanish" model of GPT-2. We used datasets of texts by three different authors, Miguel de Cervantes, Emilia Pardo Bazán, and Sor Juana Inés de la Cruz to fine-tune three specific models that generate texts in the "authorial voices" of the writers whose works they were fine-tuned with.

SCAN ME

Scan to view our website where you will find detailed process descriptions, links to our datasets, and sample texts generated by our fine-tuned models. .

Central Problems & Guiding Theories

Programmed and algorithmic biases plague artificial intelligence and a wide range of technologies, a problem charted extensively by media and science scholars like Adam Hadhazy, Safiya Umoja Noble, and Sara Wachter-Boettcher.

In GPT-2's case, we observe that biases from the primarily Anglophone training dataset are embedded in the outputs and capabilities of the transformer: when prompted in Spanish, it does not generate coherent texts.

As Simon Mahony and Jin Gao, Roopika Risam, Thea Pitman and Claire Taylor, and others note, linguistic legacies of implicit and systemic biases—themselves rooted in the coloniality of knowledge—are pervasive in digital humanities tools and practices.

Tools & Processes

GPT-2: a large-scale, open-source, transformer-based language model programmed by OpenAI. Trained on a very large text corpus--8 million web pages--this language model is designed to predictively generate the next "token" (word) in a given sequence of text. GPT-2 stands for "Generative Pre-Trained Transformer, version 2."

The original GPT-2 model is trained primarily and most effectively to generate texts in English. However, we found a fined-tuned version for Spanish: "GTP-2—small-spanish," trained by the Datificate research group. They processed 3GB data of the Spanish language Wikipedia and applied transfer learning and fine-tuning techniques from the Hugging Face libraries to obtain the resulting model. According to them, the training took around 70 hours with four GPU NVIDIA GTX 1080-Ti with 11GB of DDR5.

We used their model as a baseline to train our own. To fine-tune each author model, we prepared datasets ranging from 0.5 to 6GB containing a corpus of the texts of each author. We fine-tuned then three separate GTP-2 models using the Hugging Face transformer library. Our software platform was a Google Colab notebook which executes code using a cloud-based GPU environment.

Results: Generated Texts

El acto de la escritura es el elemento clave en toda explicación de la existencia humana. // *The act of writing is the essential element of every explanation of human existence*.

-GPTmilia Pardo Bazán

La escritura es cosa muy curiosa y muy rara, y así los unos en la lengua inglesa, la otra en el arábigo; pero . . . no hay artificio más extraordinario que el artificio que el español usa. // *Literature is a curious and odd thing, and so some [do it] in the English language, another in Arabic; but . . . there is no craft more extraordinary than that which Spanish employs*.

El acto de escribir es muy dificultoso. // *Writing is hard*.

-Cide GPTmete Benengeli (Miguel de Cervantes GPT Saavedra

Mi escritura es del alma–por ser ella la deidad–sin haber sido sólo de los sentidos. // *My writing is from my soul–as she is the deity–without having been only of the senses.*

El acto de escribir la vida fue un delito bárbaro en Egipto. // *The act of writing one's life was a barbaric crime in Egypt.*

-Sor Juana GPTnés de la Cruz

Lingering Issues & Future Plans

While the ostensible aim of our project is to respond to--and help to begin to remedy--linguistic biases in AI technology and digital humanities, certain limitations of the tools at our disposal inherently--and inevitably--reproduce implicit biases related not only to language but also to race and gender.

Wikipedia training dataset: Because the "Small Spanish" model of GPT-2 is fine-tuned with a Wikipedia training set, our fine-tuned model likely contains hidden biases based on the source data's issues of selection and representation (Wikipedia has been scrutinized both for its exclusion of certain perspectives and for its questionable factuality)

Public domain, digitized training dataset: The texts we used in our fine-tuning datasets needed to be available in the public domain--published in 1925 or before--and digitized (with character recognition). Given the historical exclusivity (in terms of gender and race) of the literary canon, we were limited in our ability to include diverse voices. We attempted to reconcile this to some degree by selecting two women authors.

Our initial aims were largely experimental and exploratory--to test the limits of the open-access AI and to test our own technical capabilities. We hope to continue learning and developing this project, ideally through the creation of an interactive web application that will allow users to test out our text-generating author models.

Selected References & Works Cited

"Datificate/gpt2-small-spanish." *Hugging Face.* huggingface.co/datificate/gpt2-small-spanish. Accessed 2 Mar. 2021.

GUILLOU, Pierre. "Faster than training from scratch — Fine-tuning the English GPT-2 in any language with Hugging Face and fastai v2 (practical case with Portuguese)." *Medium*, July 20, 2020 medium.com/@pierre_guillou/faster-than-training-from-scratch-fine-tuning-the-english-gpt-2-in-any-language-with-hugging-f2ec05c98787. Accessed 5 Sept. 2020.

PITMAN, Thea, and Claire Taylor. "Where's the ML in DH? And Where's the DH in ML? The Relationship between Modern Languages and Digital Humanities, and an Argument for a Critical DHML." *DHQ: Digital Humanities Quarterly* 11.1, 2017.

RISAM, Roopika. *New digital worlds: Postcolonial digital humanities in theory, praxis, and pedagogy*. Northwestern University Press, 2018.

SCHMID, Philipp. "Fine-tune a Non-english Gpt-2 Model with Huggingface." September 06, 2020, www.philschmid.de/fine-tune-a-non-english-gpt-2-model-with-huggingface. Accessed 2 Oct., 2020.

For more resources, visit our project website: https://rb.gy/kkuih0

Acknowledgements

This work was funded by an Andrew W. Mellon Fellowship for the Digital Humanities. We are grateful for the collegiality, comments, and constructive criticism of our fellow 2021 Mellon DH Fellows. ¡Gracias a todes!

This presentation poster was designed by FPPT.