

Former President William Howard Taft is a special interest of mine and so for my final project I thought it would be fun to try and create a William Howard Taft chat bot that I will call Willy bot. My project had 3 parts to it: building the dataset, training the model, and then making the chat wrapper. I feel that the result was mostly successful. There is definitely a lot of room for improvement but given the hardware I have available and the timeframe, my final product is passable. To use the chat bot all that needs to be done is to run chat.py.

The first step of this project was to compile the data into a usable format. The brunt of the data I used was collected from Project Gutenberg and then I also found a few primary examples from Faft just from a simple google search. After obtaining the raw text, I did some preliminary scrubbing of the data. This basically amounted to removing the Project Gutenberg headers and also the random white space from when I copied over the text from the websites. I then wrote buildDataset.py which boils down to chunking the data up, examining the chunks for key words that would indicate a topic, and then filling in a random question template. So the resulting json file would contain an instruction with the identified topic, an empty input, and then an output value that was pulled from the specific chunk. Later on there is an apparent issue that is presented by the limited amount of question templates I used. The data was then dumped into a json file which I then fed into a supplementary program to split it into the training set and eval sets.

The next part of the project is where most of the headache comes into play. The first step was finding an appropriate model that my desktop could actually train without taking days. After some research I discovered that the Qwen2.5-3B-Instruct model would hopefully be better than the tiny Llama one we used in class but also be possible for me to train on. Using this llm did not require me to download an external code base, I just needed to use pip to install a few libraries.

One important note here is that I had to install an older version of python because some of the libraries are not fully updated to be used on python 3.13, so the environment I set up needed to use python 3.10. After setting up the environment I got into writing the training program. This is where I ran into a lot of issues with getting everything to work because a lot of the hugging face blogs and git hub repositories had snippets of code, but never really showed how to set everything up. I found the documentation for transformers.Trainer and the linked blog for bitsAndBytesConfig the most useful. Also I noted in the code itself but I used chat GPT to help me fill in the gaps where things weren't working.

The final part of this project was the chat interface itself. This was pretty straightforward to code, I basically just initialized the trained model and then had a while loop so that the user could ask questions indefinitely. From here I asked it a variety of questions including some that were policy based and some other random questions. One major pitfall I made was mentioned briefly above, but the limited number of question templates leads the Willy bot to answer every question with the same few templates. When asking questions, I found the more direct policy ones gave me a fairly decent answer, but when I asked more open ended or just simple questions, the Willy bot failed. I provided a couple question and answer responses that worked well, but I also asked it a couple simple things like "spell strawberry" and "count to 10" to which it just gave me a big blurb of nonsense. I also tried asking some pitfall questions like "how many r's are in strawberry" to which I also received a big blurb of nonsense.

I included a couple of my results in the form of some Q&As. It is clear that my project worked to a degree. These responses show that when asked a question, Willy Bot pulls historical context and uses it in its answer. However there is definitely a lot of room for improvement. I think one area that could be significantly improved is the training question templates. I

mentioned this previously but I think a broader spectrum of templates would make the answers feel a little less “railroaded.” Another improvement would be to use a larger language model.

Q&A with Willy Bot:

Q: What are your thoughts on demolishing large trusts?

A: My views on demolishing large trusts were grounded in constitutional principle. I believe that the present condition of things in regard to the banking and trust companies is one of great danger to the safety of the country, because there are a few banks and trust companies which have become so powerful that they can influence legislation and public opinion at will, and this is true not only in New York and Boston but also in other cities of the country. Such power ought not to exist in any free government. I believed restraint and respect for law were essential. I believed restraint and respect for law were essential. Opposition to my bill was based upon the assumption that it would give the Government a right to interfere with private business in order to protect its own interests. This belief guided my conduct as President. In foreign affairs, my administration promoted peace. It was impossible to expect that a new country should be governed without some friction. Such judgment best served the nation. The belief guided my conduct as President. I believed restraint and respect

Q: What are your views on the judiciary?

A: In reflecting upon my public service, I have been a firm believer in the necessity of maintaining the purity of the military and naval personnel. My belief is that it is a matter of common interest to all Americans to see that the military and naval officers who are responsible

for the defense of the country should be selected with special care, and that they should be kept as pure as possible from all influence of politics or party. This belief guided my conduct as President. Such judgment best served the nation. This belief guided my action as President. I am pleased to say that in every State in the Union the law is now in force. This belief guided my conduct as President. I am pleased to say that in every State in the union the law is now in force. This belief guided my action as President. My course has always been conservative and cautious. This belief guided my conduct as President. I am pleased to say that in every state in the union the law is now in force. Such judgment best served

Q: Tell me about yourself

A: In reflecting upon my public service, A good judge of the quality of the evidence is more likely to be found in a judicial court than in any other place. This principle was recognized by the courts when they held that a person who violates a statute prohibiting the use of liquor is guilty of a misdemeanor and subject to punishment as such; but the courts have also held that a conviction for this offense is not an absolute bar to the prosecution of the same person for a subsequent offense under some other statute. This belief guided my conduct as President. I am proud of the progress which has been made in the training of our naval officers and men. Such judgment best served the nation. This belief guided my conduct as President. I am proud of the progress which has been made in the training of our naval officers and men. Such judgment best served the nation. You sought to establish the fact that there are no laws against the sale or consumption of liquor by women, nor any law making the possession of liquor by women

Q: How do you view executive authority?

A: To understand my position on executive authority, one must consider the constitutional framework. The present Congress, in the exercise of its legislative power, has enacted many wise and well-conceived laws, and has already made some important provisions for the prevention of the crime of murder, but it is now proposed to institute a new system of police and criminal procedure, under which the state shall have an inspector general with power to investigate all murders committed within the state and to lay the crime before Congress, and if Congress approves the action of the inspector general, he shall be empowered to prosecute the murderer. This approach best served the nation. I believe that the new system of criminal procedure will greatly strengthen the arm of the law against those who commit the most heinous crimes. Such judgment best served the nation. This context is essential to sound judgment. My actions were guided by duty rather than ambition. I believed that this government was entitled to the respect due a great nation. This context best served the nation. Executive authority best

bnb_config/loraConfig https://huggingface.co/docs/bitsandbytes/main/en/fsdp_qlora

Trainer arguments: https://huggingface.co/docs/transformers/main_classes/trainer