



## Small Weight Loss Habits



overwriting the knowledge it gained during the initial training phase while still adapting to the specific task of generating summaries.

- Monitor the model's performance on a separate validation dataset, and stop fine-tuning when the performance plateaus or starts to degrade to avoid overfitting.

By combining these techniques—balancing the datasets and fine-tuning your model—you can create a ChatGPT-like model that effectively generates summaries of PDFs from your core dataset without being corrupted or overshadowed by the additional diverse data sources.



i think you misunderstood a part of my purpose of creating this model. Only the pdfs will be provided as core dataset, summaries will not be provided. I expect the model to create the summaries when requested by a user. Is there a problem with this approach? and why do you suggested this " Ideally, the summaries should be human-generated"



Apologies for the confusion, and thank you for the clarification. If you don't have the summaries available and your goal is to create a model that can generate summaries for the PDFs in your core dataset, you can still train the model using a different approach.

Since you don't have a specific dataset of summaries for fine-tuning, you can use a more generic summarization dataset or task to help the model learn the general concept of summarization. This can be achieved through the following steps:

1. Train the model on the combined dataset, including your core PDFs and additional diverse data sources, as previously described.
2. Find a suitable summarization dataset or create one that contains text passages and their corresponding summaries. This will be used to teach your model the general concept of summarization. A well-known summarization dataset is the CNN/Daily Mail dataset (<https://cs.nyu.edu/~kcho/DMQA/>). This dataset contains news articles and their

Send a message...



ChatGPT Mar 23 Version. ChatGPT may produce inaccurate information about people, places, or facts.