[Fine-tuning model process](https://docs.google.com/document/d/1g691mZlQ1aNqh1y3H3bBo0Fq4a_lBpm2h2a2ZCEZRPI/edit?usp=sharing)

[Holy grail](https://platform.openai.com/docs/guides/fine-tuning)

Chosen model: GPT 3.5 Turbo

Dataset size: 10 (minimum)

1. **Data Collection and Preparation**
   1. **Collecting Content:** This content will serve as training data. It's important to have a diverse range of posts to cover various styles and topics the user might write about.
   2. **Formatting for Fine-Tuning:** Format this content into a JSONL file where each line is a separate JSONL object
      1. "**Title**:" and "**Post**:" in training data
         1. Consistency in Training Data: If training data consistently uses "Title:" and "Post:" as prefixes, the model will learn to expect and generate content in this format. This can be useful if we want the model to recognize and generate distinct sections of a post, such as distinguishing between a title and the body.
         2. Flexibility in Content Generation: We plan to generate content where the distinction between title and body is important, keeping these prefixes can help. The model will learn to treat content following "Title:" differently from content following "Post:", potentially leading to more structured and appropriately formatted outputs.
         3. **Training and Usage Alignment: If users will input prompts in a format that includes "Title:" and "Post:", then including them in your training data makes sense. Conversely, if users will input more general prompts, a training dataset without these prefixes might be more appropriate.**
      2. Example Format:
         1. {"messages": [{"role": "user", "content": "How to Create Strong Resume Bullet Points"}, {"role": "assistant", "content": "I cold-applied to 150 internships as a freshman…”}]}
   3. **Data Preparation**
      1. Convert to JSONL
      2. messages, Completion tags
      3. No “ s
      4. Remove new lines
2. Fine-Tuning the Model
   1. Upload the Prepared Data: Use OpenAI's CLI to upload your JSONL file.
      1. echo 'export OPENAI\_API\_KEY="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"' >> ~/.zshrc
   2. Fine-Tune a Model: Use OpenAI's fine-tuning command to train a new model based on your uploaded data. The fine-tuned model will learn the style and content patterns from dataset.