

The Problem

Enterprises have millions of rows of unstructured text data (DOCX, PPTX, PDF, TXT files). Currently, they have people manually going through documents, by hand, to classify text, extract entities, answer questions, or find information, which has the following downsides:

- 1) **Time Consuming** - Slow, even with an unlimited labor force, it can take months to years
- 2) **Costly** - Expensive, takes up a significant portion of a team's budget
- 3) **Tedious** - Laborious, mundane and boring work that people shouldn't have to do
- 4) **Requires Expertise** - Hard, requires highly trained subject matter expertise (doctors, legal analysts, financial analysts) with domain-specific experience to find information
- 5) **Iterative** - Repetitive, when new data or new categories are added, or business requirements change, humans need to manually go through the data again

Our Solution

Anote's solution consists of 3 fully functional products, and has 3 main steps:

- 1) [Label Text Data](#) - Classify Text, Extract Entities and Answer Questions on Documents
- 2) [Fine Tune Model](#) - Run your Fine Tuned LLM locally with our model inference API
- 3) [Private Chatbot](#) - Chat with your documents while keeping your data private and secure.

The core technology is a fine tuning library that leverages state of the art few shot learning to make high quality predictions with a few labeled samples. There are 3 ways to do fine tuning:

- 1) **Unsupervised Fine Tuning** - Fine tune your LLM from your raw unstructured documents
- 2) **Supervised Fine Tuning** - Fine tune your LLM on your labeled data
- 3) **RLHF / RLAIF Fine Tuning** - Actively improve your models with human / AI feedback

For improved model performance, you can do supervised or RLHF / RLAIF fine tuning via our data annotation interface / API, which you can use with 4 main steps:

- 1) **Upload** - Create a new text based dataset
- 2) **Customize** - Add the categories, entities or questions you care about
- 3) **Annotate** - As you annotate a few edge cases, the LLM actively learns to predict the rest
- 4) **Download** - Download the resulting CSV / export the fine tuned LLM as an API endpoint

You can take your exported fine tuned model, and input it into our Private Chatbot, your accurate / private enterprise AI assistant. Private Chatbot has both a UI for enterprise users and a software development kit for developers. To use the Private Chatbot, there are 3 main steps.

- 1) **Upload** - Upload your documents
- 2) **Chat** - Ask questions on your documents with LLMs like GPT, Claude, Llama2 and Mistral
- 3) **Evaluate** - Get citations for answers, and mitigate the effect of hallucinations

Differentiation

There is a subset of problems where the categories, entities or answers to questions are too complex to be done purely by AI. For this subset, human in the loop tooling is needed to ensure accuracy requirements are met. Our technology enables customers to get information they need from their unstructured text data faster, cheaper, and more accurately, providing insights that otherwise were not possible to obtain, all while adhering to customer's requirements:

- 1) **Accurate Predictions** - Fine tuning and enhanced RAG for more accurate and tailored responses. Our AI models actively learn and rapidly improve from SMEs.
- 2) **Accurate Citations** - Accurate sources (page number, chunk of text, important features) to explain the models predictions and mitigate hallucinations.
- 3) **Comprehensive Capabilities** - Supervised, Unsupervised and RLHF / RLAIF fine tuning for classifying text, extracting entities, answering questions, and chatting with documents
- 4) **Easy to Use** - Accessible UI similar to ChatGPT, and simple SDK for developers where you can input a fine tuned model for improved results.
- 5) **Private** - On premise enterprise-grade solution (desktop app) using Llama2 and Mistral to leverage LLMs on your unstructured documents while keeping your data local/secure.
- 6) **Evaluation Framework** - Robust evaluation framework with metrics to show fine tuned model performance improvements vs. zero shot models.

Value Add

What we are building at Anote matters, and here is why:

Use Case	Before	After / With Anote
Data Labeling Approach	Manually Labeling their data themselves in a spreadsheet	State of the art few shot learning to make high quality predictions with a few labeled samples
Data Labeling Impact	Tedious, Time Consuming, Costly, Manual Iterative Relabeling	Less time, less expensive, higher accuracy, Rapid flexibility for changing business requirements
Document Processing Approach	Given raw unstructured documents, such as a 10-k or earnings call transcript, you can't get answers to the questions right if trying to extract info, where accuracy really matters.	After a few interventions, we go from 10 questions right, to 15 questions right, to 20 questions right, to enable insights that were otherwise impossible to obtain.
Document Processing Impact	Not accurate and largely manual extraction, Sub-optimal analytics for critical business decisions	Higher accuracy for raw unstructured documents, New insights that otherwise were not obtainable

Thank You

Anote is a startup in New York City helping make AI more accessible. We believe there is a massive gap between the tremendous power of AI models, and the everyday tasks that people care about. To learn more, visit our [website](#), [youtube](#), [linkedin](#), or email us at nvidra@anote.ai.