



Maximize Impact, Minimize Effort with Active Learning

April 17, 2024

By Loris D'Acunto

Chief Data Scientist

Revolutionizing Banking Communications

In the highly regulated world of investment banking, the accuracy of machine learning models and how they are trained is a crucial aspect that cannot be underestimated. Large language models often have trouble when used in specific areas like finance, especially if they're not trained on the detailed types of data they need to work with. This problem gets bigger because many important datasets, including private financial communications, are not publicly available due to privacy and security concerns. This creates a significant hurdle in making models that can effectively understand and handle these data types.

To overcome this, banks need a team with a large set of different skills:

- Data scientists specialized in natural language processing, tasked with selecting examples for expert annotation and model training;
- Subject matter experts (SMEs) knowledgeable about the organization's internal processes, responsible for annotating data points (e.g., email intents, document terms);
- Machine learning engineers, who are charged with training the models, verifying their accuracy, and deploying them in production.

At DeepSee, we've created a system that learns as it goes, making these considerable tasks much more manageable and significantly cutting down the time and work needed to get accurate models in days, not months.

Active Learning: The New Frontier in Banking

Dynamic Selection

At the heart of our approach is the dynamic nature of our active learning system. Unlike traditional models that might randomly select data for training, our system intelligently adapts to the complexity and variability of language found in financial documents. By strategically choosing which emails to include in training, our system ensures that each selection maximizes the model's learning potential. This approach makes our system more efficient and more effective in dealing with the diverse types of communication that flow through investment banks.

Interactive Training

Our system is interactive, constantly refining its predictions through continuous training of intermediate models. This real-time learning process is accelerated by feedback loops that ask the users for confirmation of the model's predictions. Such interactions help the system rapidly improve its accuracy, adapt swiftly to new data, and ensure the learning process is as efficient as possible.

Proactive Quality Control

Moreover, our active learning system is proactive in maintaining the quality of the training dataset. It monitors for inconsistencies and human errors, automatically generating tasks for human-in-the-loop (HITL) review when needed. This ensures the highest possible quality of the training data, which is critical for achieving top-notch model accuracy.

Real-World Success: A Case Study

During a proof of concept (POC) that ultimately led to securing a new customer, we processed 90,000 emails with our system. Remarkably, one SME could annotate only 3,500 emails in 16 hours, but the model trained with this data achieved an exceptionally high accuracy of 98%. This success underscores the efficiency of our active learning system, which enables fine-tuning models in hours rather than weeks.

Furthermore, we extended our testing beyond average precision and recall scores on a small dataset. To demonstrate our model's efficiency and accuracy comprehensively, we conducted a separate session to categorize the whole set of emails. We then assessed the performance metrics against all labelled emails, not just a subset, providing our customer with undeniable evidence of our model's robustness across the entire dataset.

Transforming Investment Banking Communications

DeepSee's active learning system marks a significant leap forward in processing and understanding financial communications within investment banks. By leveraging dynamic selection, interactive training, and proactive quality control, we've enhanced model accuracy and dramatically reduced the human effort and time traditionally required for such tasks. This breakthrough represents not just an improvement in operational efficiency for our clients but a transformative approach to managing and automating financial communications.