How We Generated Millions of Content Annotations
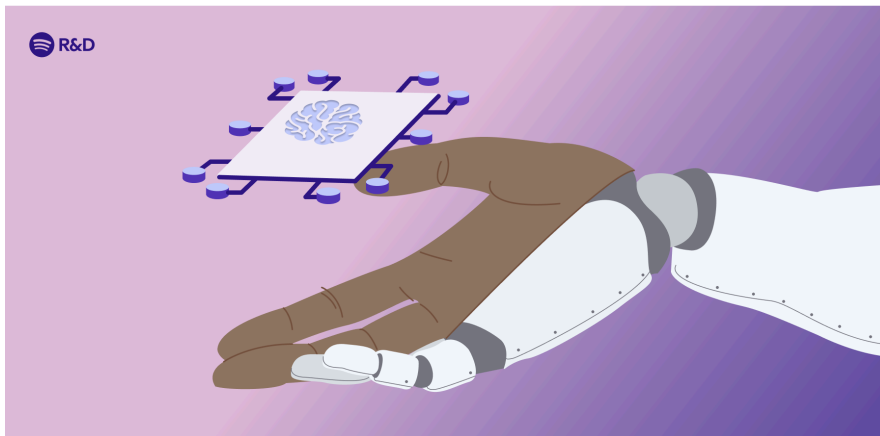
October 21, 2024
Published by Dana Puleo (Annotation Platform Ops Manager), Meghana Seetharam (Staff Engineer), and Katarzyna Drzyzga (Staff Engineer)

With the fields of machine learning (ML) and generative AI (GenAI) continuing to rapidly evolve and expand, it has become increasingly important for innovators in this field to anchor their model development on high-quality data.

As one of the foundational teams at Spotify focused on understanding and enriching the core content in our catalogs, we leverage ML in many of our products. For example, we use ML to detect content relations so a new track or album will be automatically placed on the right Artist Page. We also use it to analyze podcast audio, video, and metadata to identify platform policy violations. To

CATEGORIES

Data

Machine Learning

Related articles

Are You a Dalia? How We Created Data Science Personas for Spotify's Analytics Platform

hundreds of millions of tracks and podcast episodes. To implement ML at this scale, we needed a strategy to collect high-quality annotations to train and evaluate our models. We wanted to improve the data collection process to be more efficient and connected and to include the right context for engineers and domain experts to operate more effectively.
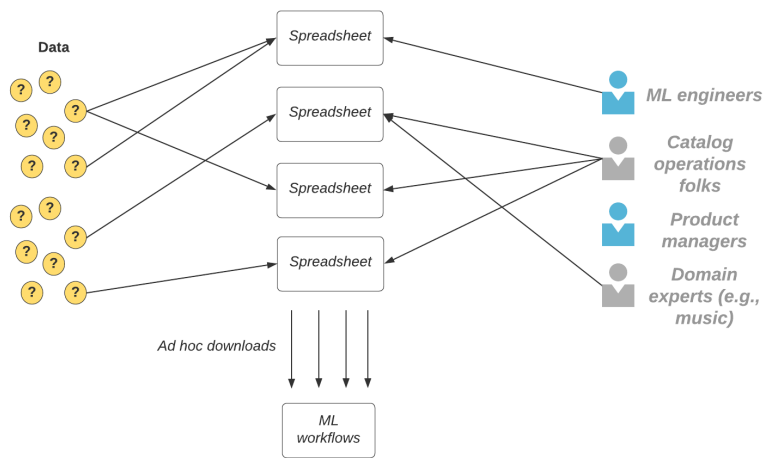

Figure 1: Ad hoc data collection processes.

To address this, we had to evaluate the end-to-end workflow. We took a straightforward ML classification project, identified the manual steps to generate annotations, and aimed to automate them. We developed scripts to sample predictions, served data for operator review, and integrated the results with model training and evaluation workflows. We increased the corpus of annotations by 10 times and did so with three times the improvement in annotator productivity.

Taking that as a promising sign, we further experimented with this workflow for other ML tasks. Once we confirmed the benefits of our approach, we
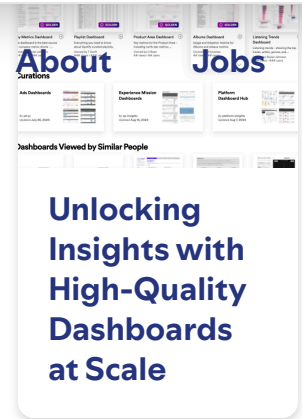
Sign up for updates

example@ema

@SpotifyEng on Twitter

platform that would scale to millions of annotations.

Building and scaling our annotation platform

We centered our strategy around three main pillars:

1. Scaling human expertise.
2. Implementing annotation tooling capabilities.
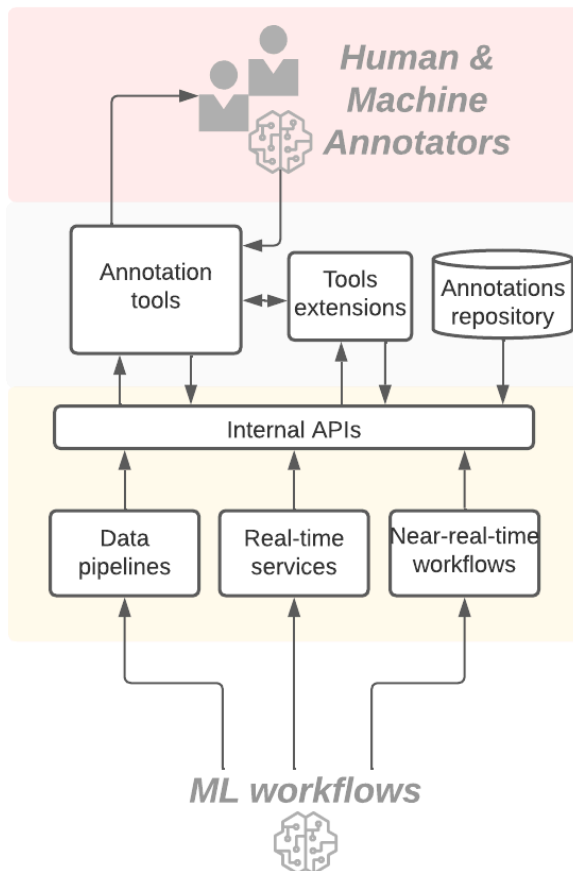3. Establishing foundational infrastructure and integration.

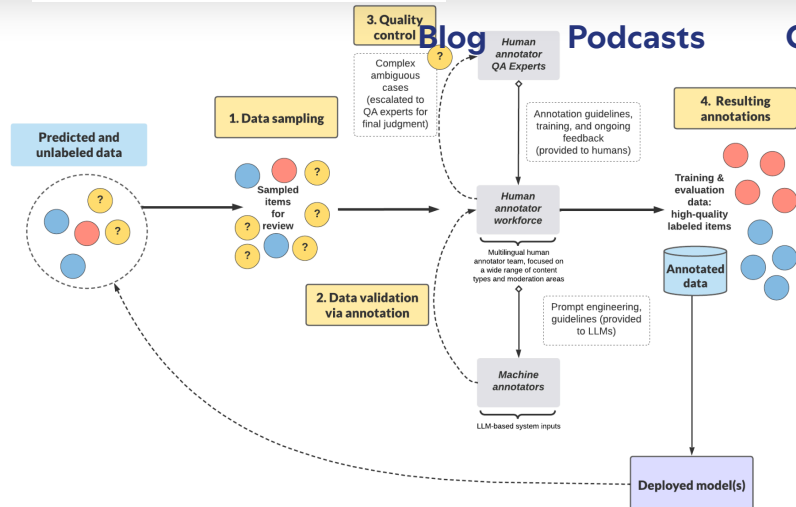*Figure 2: Pillars of the annotation platform.*

Figure 3: Annotation workflow diagram.

In order to scale operations, it was imperative that we defined processes to centralize and organize our annotation resources.

We established large-scale expert human workforces in several domains to address our growing use cases, with multiple levels of experts, including the following:

- **Core annotator workforces:** These workforces are domain experts, who provide first-pass review of all annotation cases.
- **Quality analysts:** Quality analysts are top-level domain experts, who act as the escalation point for all ambiguous or complex cases identified by the core annotator workforce.
- **Project managers:** This includes individuals who connect engineering and product teams to the workforce, establish and maintain training materials, and organize feedback on data collection strategies.

to the human experts. It has allowed us to significantly grow our corpus of high-quality annotation data with low effort and cost.

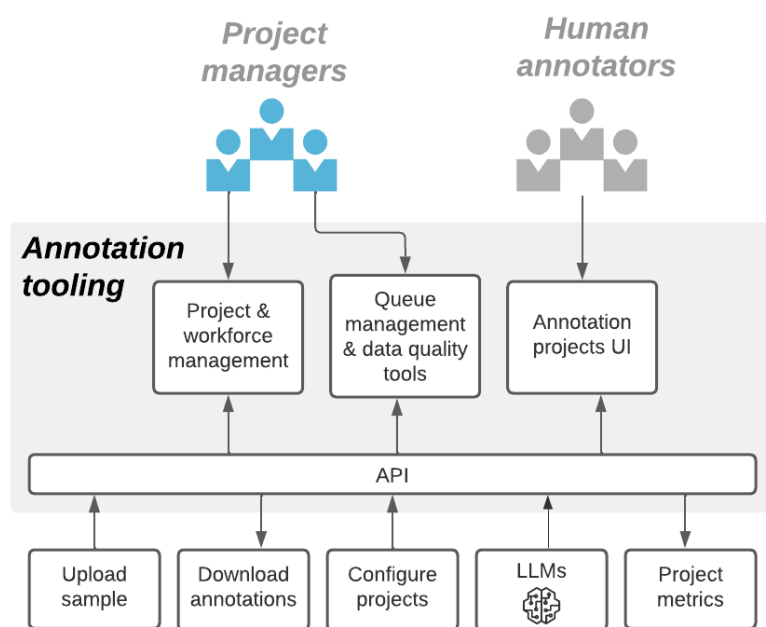## 2. Implementing annotation tooling capabilities.

Figure 4: Annotation tooling capabilities.

Although we started with a simple classification annotation project (the annotation task being answering a question), we soon realized that we had more complex use cases — such as annotating audio/video segments, natural language processing, etc. — which led to the development of custom interfaces, so we could easily spin up new projects.

In addition, we invested in tools to manage backend work, such as project management, access control, and distribution of annotations across multiple

experts remained productive across multiple projects.

Another focus area was project metrics — such as project completion rate, data volumes, annotations per annotator, etc. These metrics helped project managers and ML teams track their projects. We also examined the annotation data itself. For some of our use cases, there were nuances in the annotation task — for example, detecting music that was overlaid in a podcast episode audio snippet. In these cases, different experts may have different answers and opinions, so we started to compute an overall "agreement" metric. Any data points without a clear resolution were automatically escalated to our quality analysts. This ensures that our models receive the highest confidence annotation for training and evaluation.

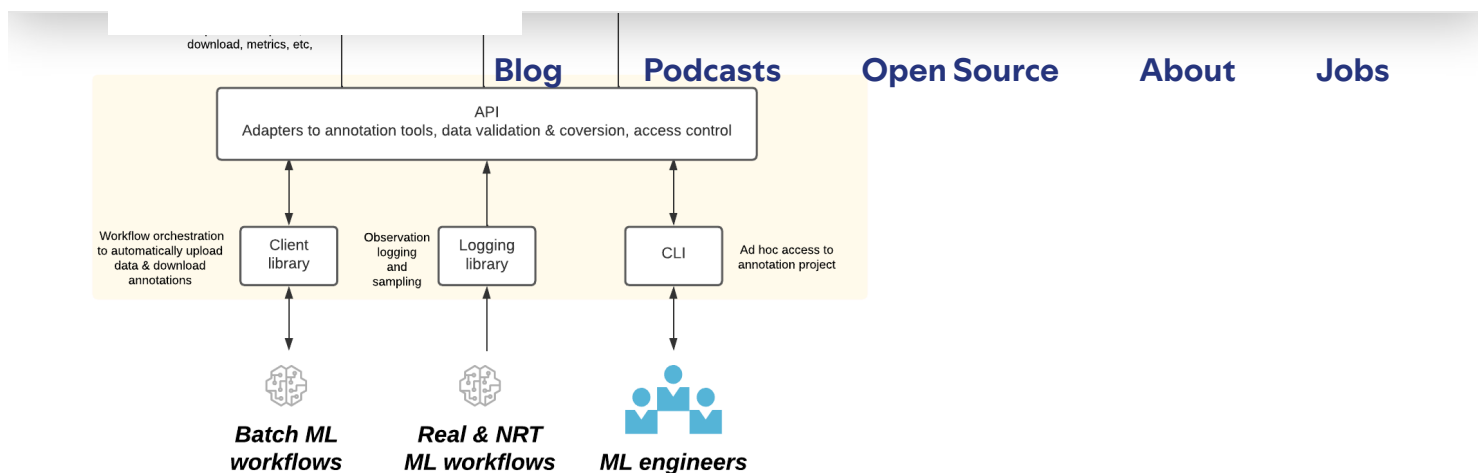# 3. Establishing foundational infrastructure and integration.

Figure 5: Infrastructure to integrate with the tooling.

At Spotify's scale, no one tool or application will satisfy all our needs — optionality is key. When we designed integrations with annotation tools, we were intentional about building the right abstractions. They have to be flexible and adaptable to different tools so we can leverage the right tool for the right use case. Our data models, APIs, and interfaces are generic and can be used with multiple types of annotation tooling.

We built bindings for direct integration with ML workflows at various stages from inception to production. For early/new ML development, we built CLIs and UIs for ad hoc projects. For production workflows, we built integrations with internal batch orchestration and workflow infrastructure.
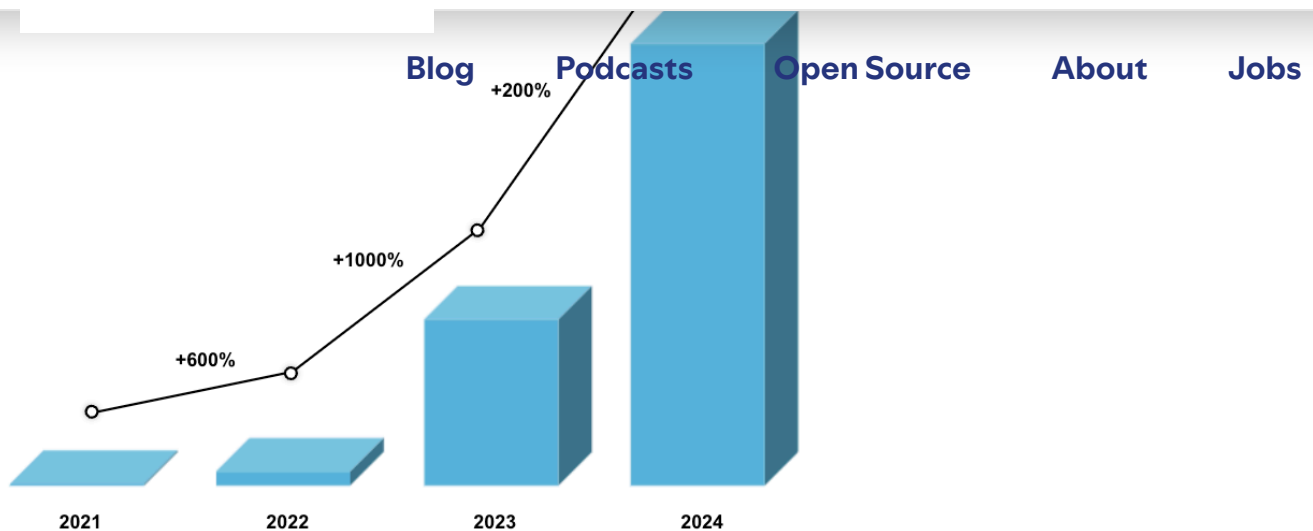
+200%

+1000%

+600%

2021   2022   2023   2024

Figure 6: Rate of annotations over time.

Conclusion

The annotation platform now allows for flexibility, agility, and speed within our annotation spaces. By democratizing high-quality annotations, we've been able to significantly reduce the time it takes to develop new ML models and iterate on existing systems.

Putting an emphasis from the onset on both scaling our human domain expertise and machine capabilities was key. Scaling humans without scaling technical capabilities to support them would have presented various challenges, and only focusing on scaling technically would have resulted in lost opportunities.

It was a major investment to move from ad hoc projects to a full-scale platform solution to support ML and GenAI use cases. We continue to iterate on and improve the platform offering, incorporating the latest advancements in the industry.

A special thanks to Linden Vongsathorn and Marqia Williams for their support in launching this initiative and to the many people at Spotify today who continue to contribute to this important mission.

Tags: machine learning

SHARE THIS:

Sign up for engineering updates

By clicking sign up you'll receive occasional emails from Spotify. You always have the choice to adjust your interest settings or unsubscribe.

Your Email

Sign up!

Spotify Jobs   Spotify R&D Research

Spotify.com   Spotify R&D Design

Newsroom