Training data platform buying guide

BROUGHT TO YOU BY

V7

TRY IT NOW →

Personas Analysis

	Typical Concerns	Metrics that matter to them	Role in evaluation
Annotators	Is the tool easy to learn? Is the tool easy to use? Does the tool allow me to automate routine tasks? Does the tool allow me to dispense with non-core tasks?	Ramp time for annotators Speed of annotation at Month 1, Month 3, and Month 12	Annotators should be allowed to try the software to evaluate the UI and provide feedback
Annotator Management	Can I easily add new annotators or manage my existing workforce? Does the tool have different permissions/access roles and easy authorization Can I easily report on annotator performance? How can I minimize annotator downtime? Can I easily assign work and roles to annotators?	Annotator efficiency (how much data can they process in a given hour) Annotation quality and accuracy Speed of administration (effectively, end-to-end, how long does it take to manage a new project)	The Annotator Manager should be a core part of the evaluation process They should be a key stakeholder in testing the software and provide reports on annotator performance They should provide efficiency benchmarks for comparison
Data Science / CV Team	Does the tool have a robust SDK/API for easy use? Does the platform minimize my involvement in the annotation process? Does the platform allow me to run experiments and explain the link between training data and model performance? Can the platform scale with our growing ambitions? Is there good support and a performant platform?	Speed of administration (effectively, end-to-end, how long does it take to manage a new project) Accuracy of annotation Performance of Models	The CV Team typically lead the technical evaluation process. This should be not only an evaluation of ease of use of the platform itself, but also understanding how it would fit into the broader stack. They should spend time evaluating the dataset management and administrative functions of a training data platform to ensure it will minimize their time in the weeds of these processes, but allow for better AI
Executive Stakeholders	Will this platform save me money? Will this platform increase the performance of our models? Will this platform free up more of my top engineers' time? Is the platform and our data secure?	Return on Investment Model performance SLAs & Security	The Executive Stakeholders should review the business case of the product to understand it's ability to affect their north star metrics and KPIs. They should easily be able to estimate the ROI of the solution and understand how it might enable their team to hit their targets

TRY IT NOW →

V7

Features Analysis

Feature	Weighting	Vendor 1	Vendor 2	Vendor 3
Data and Annotation Types				
Privacy and Security				
Annotator Speed and Efficiency				
Quality Assurance				
External Annotators				
AI Models				
Dataset Management				
API & SDK				
Support				
Totals				

TRY IT NOW →

V7

Data Types and Annotation Types

Feature	Vendor 1	Vendor 2	Vendor 3
JPEG/PNG			
Video			
Audio			
DICOM			
TIFF			
TIF			
SVS			
NDPI			
Bounding Box			
Keypoint			
Keypoint Skeleton			
Polygon			
Ellipse			
Polyline			
Cuboid			
Classification Tags			
Total Score			

Privacy and Security

TRY IT NOW →

Feature	Vendor 1	Vendor 2	Vendor 3
Where is my data stored?			
Can my data be accessed by others?			
Will my data be used to improve services for others?			
Can I keep data in my own environment?			
GDPR Compliant?			
HIPAA Compliant?			
FDA Part 11 Compliant?			
SOC2/ISO27001?			
SSO?			
2 Factor Authentication?			
Cloud Storage Integrations?			
Total Score			

TRY IT NOW →

V7

Dataset Management

Feature	Vendor 1	Vendor 2	Vendor 3
Ease of import			
Provisioning and project management			
Class Balance and Overview			
Annotator Metrics			
Querying			
Complex Querying			
Dataset Versioning			
Experiment models			
Total Score			

TRY IT NOW →

V7

Quality Assurance

Feature	Vendor 1	Vendor 2	Vendor 3
QA Workflows			
Universal Batch Sampling			
Granular Sampling (e.g. person by person)			
Comment Tool			
Consensus Labelling			
Ground Truth / Gold stages			
Automated QA process			
Ease of re-queuing or restarting workflows			
Total Score			

TRY IT NOW →

API & SDK

Feature	Vendor 1	Vendor 2	Vendor 3
Uploading data			
Creating Datasets			
Assigning Work			
Setting up workflows			
Webhooks			
Managing Datasets			
Complex Analytics			
Creating Releases			
Moving data into other platforms			
Ease of integration			
Total Score			

TRY IT NOW →

V7

Annotator Speed and Efficiency

Feature	Status Quo	Vendor 1	Vendor 2	Vendor 3
Time to accurately annotate 100 Bounding Boxes				
Time to accurately annotate 100 tags				
Time to accurately annotate 100 polygons				
Time to accurately annotate 100 frames of video				
Time to accurately annotate 100 images				
Time to accurately annotate 1 DICOM Series				
Time to accurately annotate 1 Large Image (TIF/WSI)				
AI-Assisted Labelling?				
Models in the Loop?				
Totals				

TRY IT NOW →

V7

AI Models

Feature	Vendor 1	Vendor 2	Vendor 3
Upload tagged data?			
Upload pre-annotated data (polygons and bounding boxes)?			
Integrate our own models into the labelling loop?			
Combine the output of multiple models in a single labelling workflow (cascading)?			
Train Models on the platform?			
GPU Orchestration?			
Quality of models trained on the platform?			
Ease of integration of models into the loop?			
Speed of training and inference?			
Model API?			
Total Score			

TRY IT NOW →

V7

External Annotators

Feature	Vendor 1	Vendor 2	Vendor 3
Does the platform offer an annotation workforce?			
Ease of setting up a labelling job?			
Quality of communication			
Cost of annotators			
Quality of annotation			
Security of annotators (Cloud vs tenured employees)			
Ease of integration of additional humans into the loop?			
Specialist annotators?			
Scale?			
Turnaround time?			
Internal Project Management and review?			
Total Score			

TRY IT NOW →