

# Sama platform Process and Feature Overview

The Sama platform is our proprietary, full-service task distribution platform. This web-based task management system helps facilitate large data projects through the customization of task workflows, task distribution, multi-tiered quality control, and project-based training. We have a full-time product and engineering team dedicated to continuously improving our platform to serve our clients' dynamic industries.

The Sama platform has four main functions: task creation and delivery, task distribution, data structuring work (image annotation, data categorizing, other writing/research) and quality management.

- The Sama platform's external APIs enable programmatic querying of task status, posting new tasks to be done, and automatically receiving results once the task has been completed
- The Sama platform's work distribution system allows a large team of data annotators to work simultaneously on one project's queue of tasks.
- The Sama platform's data structuring work allows data annotators to perform custom work, such as annotating images with specific labels.
- The Sama platform's quality management system, Sampling, allows quality managers to check the quality of work and provide feedback to annotators to ensure teams are working efficiently and effectively.

## Table of contents

### 1. Overall Sama platform project workflow

- A. Project setup
- B. Task production
- C. Quality management and assurance
- D. Delivery

### 2. Computer vision offerings

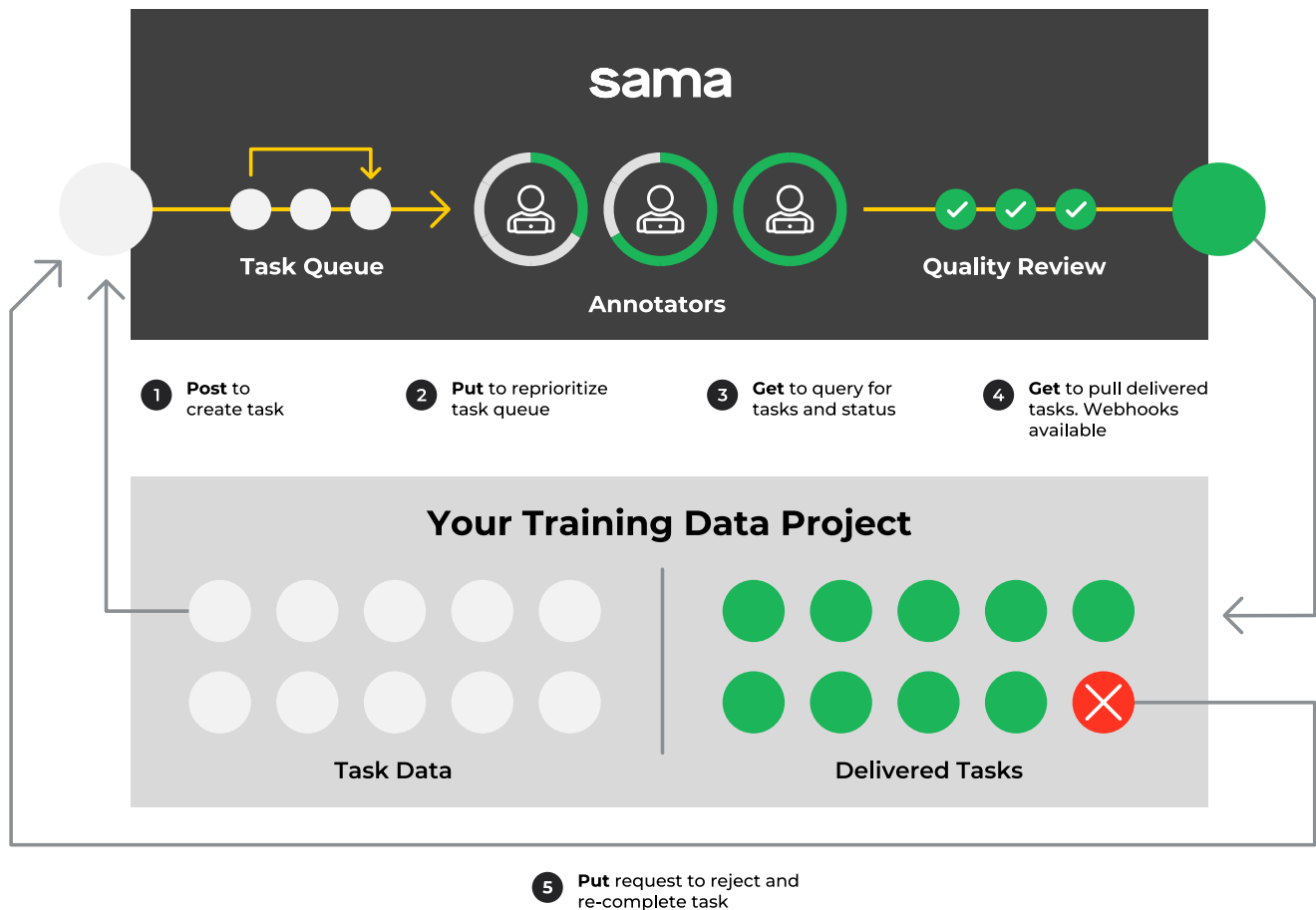
- A. Vector annotation for 2D images
- B. Semantic segmentation for 2D images
- C. Vector annotation for 2D videos
- D. Cuboid annotation for 3D LiDAR point cloud data

## Overall Sama Platform Project Workflow

### Project setup

Project workspaces are customizable by our project manager using the Sama platform's markup language, H3ML.

- Our platform supports a variety of data inputs such as images, text and URLs. Project managers upload tasks via CSV or tasks can be sent by the client via API.
- Project managers add questions to the workspace to elicit the correct output data from annotators. Question types: multiple choice questions, free text and image annotation or segmentation workspaces.
- Sama hosts assets such as images on a secure Amazon Web Services S3 server. For security reasons, each project has its own folder in S3 that is only accessible by Sama admins and the client.



## Task production

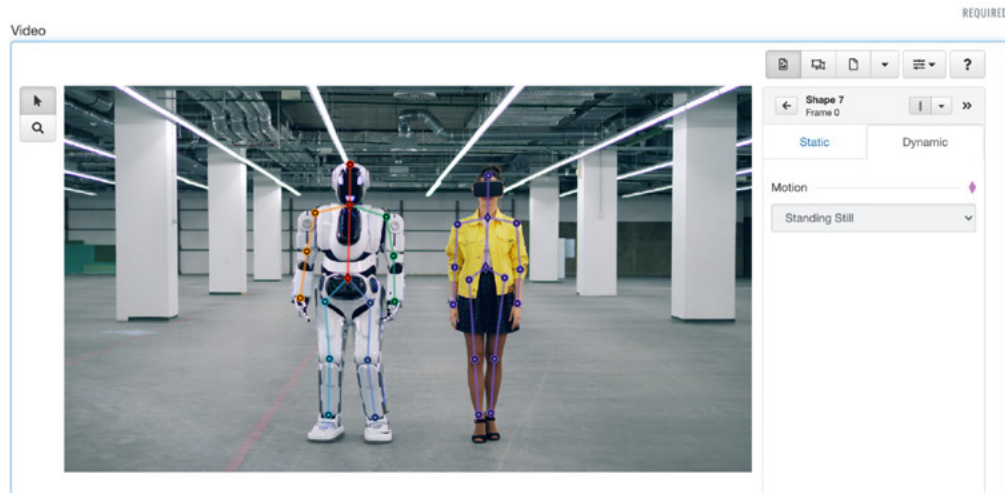


Figure 1: Task workspace is fully customizable. Sama project managers are able to add or modify a variety of output labels and options.

The customized Sama platform workspace can include guiding data such as links out to the internet, embedded images, or images to be annotated. Questions can be required and/or dependent on other questions to ensure data completeness.

- Allow annotators to pull the next available task from the general pool of tasks, prioritizing throughput. Tasks can be assigned a relative priority to move them forward or backwards in the queue.
- Serve a group of related tasks to a single annotator in an ordered sequence if the tasks are related in some way and should be worked on by the same person.

ID	CREATED	FILTERS	SIZE	QUALITY	REJECTED
	Jan 22	DEC 1 - JAN 22 COMPLETED	5		
969109	Jan 22	10%	1 of 5	100%	
969108	Jan 22	56%	3 of 5		
	Jan 22	DEC 1 - JAN 22 COMPLETED	...		
	Jan 22	JAN 15 - JAN 22 COMPLETED	...		
	Jan 12	JAN 5 - JAN 13 COMPLETED	...		
	Jan 12	JAN 5 - JAN 13 COMPLETED	...		

Figure 2: Sampling landing page for quality management in the Sama Platform.

## Quality management and assurance

### AutoQA

- AutoQA improves client delivery quality and decreases agent rework and manual QA time by programmatically evaluating 100% of agent work for specific logical errors in the workflow requirements.
- Annotators receive clear feedback on the errors to correct prior to submission.

### Sampling home

- Project managers, team leads and customers may use the Sampling function to inspect completed tasks on a project to confirm they meet quality standards.
- Reviewers filter tasks by created or last submitted date, specific annotator, answers given on a task, or a batch ID group name (such as a folder).
- Reviewers inspect a subset of the filtered tasks (example: 10%) in detail and then make a quality judgement on all of the tasks based on the quality of the subset.
- The Sama platform offers several inspection strategies: percent of tasks, fixed number of tasks per annotator and view all tasks in order. "View all" is useful for projects in which the order of tasks is important for data consistency.

### Task inspection and scoring

- Reviewers at the delivery center use Sampling to evaluate annotators' work and give it a quality score. Each task receives a score and the overall quality score is the average score of all tasks. Annotators receive daily quality scores to track their work.
- Sampling has other tools for reviewers to give annotators feedback to help them correct tasks and improve their work in general.
- Individual questions (outputs) can be scored to help annotators easily identify which portion of the task needs improvement.
- Reviewers may add free text comments or choose a specific "reason for error" from a list. Reasons for error can be used to track error trends for an annotator or for the project team. Trends can identify an annotator that needs more training, or a topic that needs more training for the entire team.
- Any work that does not meet the average quality score specified in the SLA is rejected to be revised.
- Approved work that has completed QA and meets quality is queued for delivery.

### Task feedback

- At the beginning of each shift, the team meets to discuss common points of feedback.
- Data annotators see any feedback left by reviewers on rejected tasks.

### Review step

- Projects that require a short turn-around time via Sama's API sometimes use a review step. The review step enables a second, more senior annotator to check the first annotator's work and make any changes before the task is delivered. The review step is an alternative quality management process to Sampling.

### Multiple submission steps

- Projects with subjective requirements sometimes will use multiple submission steps. Each step allows for a unique answer to the task. Multiple submission is an alternative quality management using the Sampling feature.

### Delivery

- When tasks have been approved for quality, clients can use an API to receive the delivered work. See Sama's API documentation for more detail. Sama has several APIs available for task delivery, acknowledgement and <https://docs.sama.com>
- Project Managers can also provide a delivery file (JSON, .XLS, or .CSV) on a specified cadence. This file can be posted to a secure shared file drive of the customer's choice.
- For more details, please request a sample delivery file or API test from Sama.

## Computer Vision Offerings

### Vector annotation for 2D images

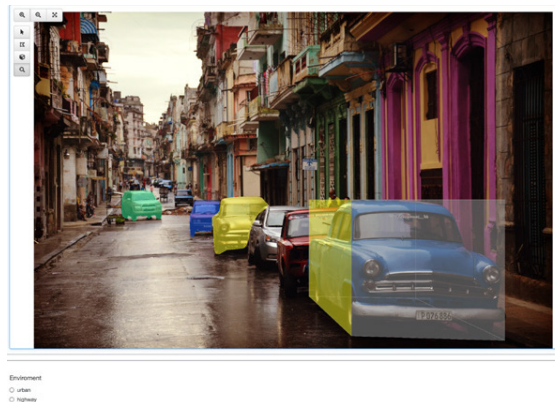


Figure 3: Task workspace view of a vector image tagging task in the Sama Platform. An additional metadata question appears below the image tagging canvas.

- A data annotator is presented with 2D images and asked to outline objects using vector shapes.
- Objects may be annotated with a bounding box (rectangle), polyline, closed polygon, directional polyline (arrow), 3D cuboid, point (vertex) or custom keypoint shape. After drawing the shapes, labels may be added, either from a menu of options or as free text.
- Annotated objects have automatically added unique IDs
- Annotators may label a single object with multiple pieces of metadata, i.e. vehicle make, model, color, license plate and direction. The label selected is reflected by the color of the shape in the workspace.
- Additional features: pixel-level zoom, copy and paste drawn shapes in an image or from a previous task, add a group label to related shapes to indicate that there is a relationship between them, view shapes one at a time to check accuracy, measure the pixel size of objects to determine if they should be labeled according to customer requirements.
- More metadata about the image as a whole can be provided through additional questions added to the workspace below the image.

The output for vector annotation projects is JSON containing the (x,y) coordinates of all shape vertices and their associated labels. This file can be delivered as a .CSV or raw JSON file.

- Annotation information around object of interest will abide by below format and syntax:
- Coordinate convention for each image: top left of the image will be marked (0,0). The x-axis is horizontal, and the y-axis is vertical.
- The format of each bounding box is  $[[x1,y1],[x2,y1],[x1,y2],[x2,y2]]$  where (x,y) are the coordinates of each corner of the bounding box.
- The format of each polygon is  $[[x1,y1],[x2,y2],[x3,y3],\dots,[xn,yn]]$  where (x,y) are the coordinates of the n vertices of the polygon.
- For more details, please request a sample delivery file from Sama.

## Semantic segmentation for 2D images

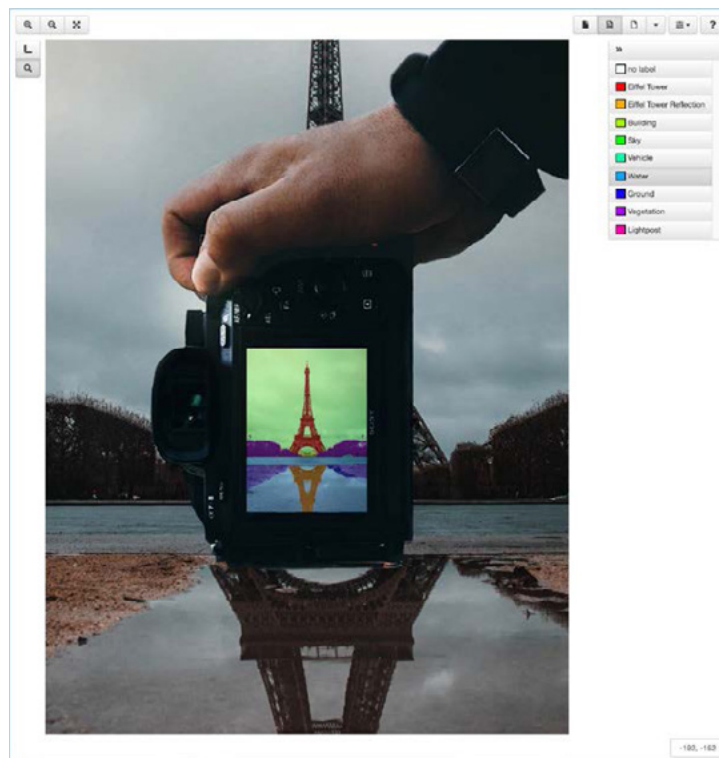


Figure 4: Task workspace view of a semantic segmentation task in the Sama Platform.

The Sama platform's pixel-level semantic segmentation tool allows teams to fully categorize and label every pixel in a 2D image using three tools: paintbrush, flood fill, and superpixels. Data annotators use the tools to draw a mask (area) and add a label from a fixed list to all objects with that mask. In the workspace, annotators view each label as a distinct color for usability. The delivery file is a grayscale PNG and contains numeric values for a computer to ingest.

Tools available for data annotators to label pixels efficiently:

- **Paintbrush:** Use the paintbrush tool for fine annotations. This tool is best used for fine tuning annotations or in small areas where color and contrast are similar but the human eye can identify the pixels belonging to different objects. The brush size can be adjusted.
- **Flood fill/magic wand:** Using color and contrast, our fill tool colors all pixels up to the boundaries of an object. This tool is best used to annotate large objects such as the sky or road.
- **Superpixel:** Superpixels break the image up into large tiles of similarly colored pixels to be annotated together. There is the option of applying 3 types of superpixel: SLIC, SLICO, and PFF algorithms to create the tiles.
- Segmentation contains many of the same usability tools as vector annotation: pixel level zoom and zoom to a specific region, view all masks with a specific label.
- For more details, please request a sample delivery file from Sama.

The output for segmentation projects is either .CSV or JSON formats and includes the image name, original image location in S3, and the PNG masked output file location in S3. Output masks are stored as PNG files in a secure folder on our S3 server. Each output mask is a grayscale PNG containing the layer index value for each pixel in the image. As the mask files can be large, it is easiest to download them directly from S3 onto a computer (as opposed to emailing them). Masks can be downloaded from S3 using a Python script provided by Sama. Sama can also provide direction to clients who prefer to write their own Python script.

## Vector annotation for 2D videos

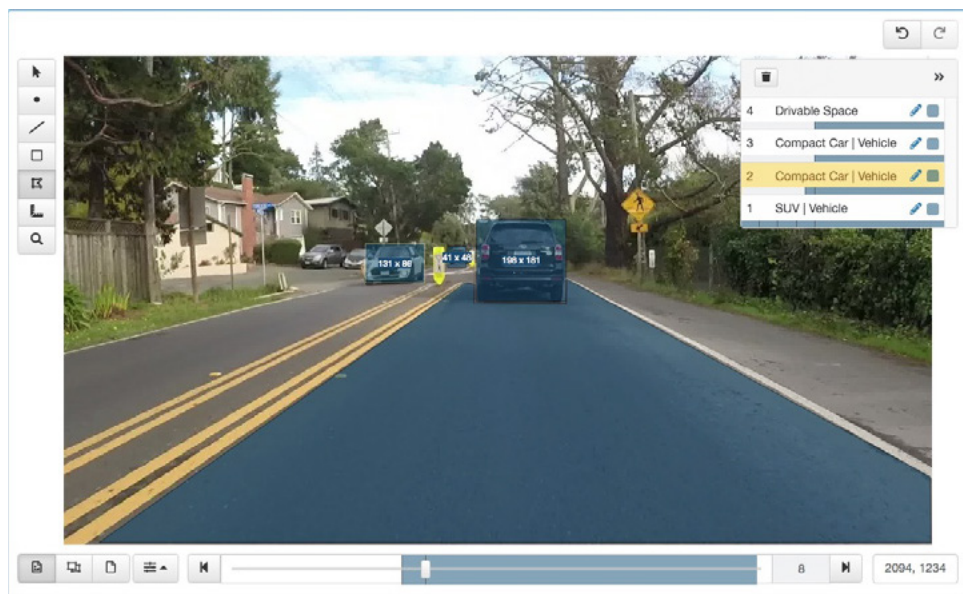


Figure 5: Task workspace view of a video vector annotation task in the Sama Platform.

A data annotator is presented with a 2D video, split up into images, and asked to outline objects using vector shapes. Annotators may scrub between frames to view the sequence.

- Objects may be annotated with a bounding box (rectangle), polyline, closed polygon, directional polyline (arrow), point (vertex) or custom keypoint shape. After drawing the shapes, labels may be added, either from a menu of options or as free text.
- Annotation is assisted by linear interpolation between key frames. Linear interpolation improves efficiency by automatically interpolating shape size and position between the key frames, the frames that were manually annotated.
- Annotated objects have automatically added unique IDs that carry through the entire video.
- The visibility of an object may be changed throughout the video while maintaining the same unique object ID. For example, if an object moves out of the frame and returns, it will have the same unique ID. Visibility settings are visible, occluded, not visible.
- A single object may be labeled with multiple pieces of metadata, i.e. vehicle make, model, color, and license plate. Object metadata can change from frame to frame (traveling direction, current lane, etc). The primary label selected is reflected by the color of the shape in the workspace.
- Additional features: pixel-level zoom to a region, add a group label to related shapes to indicate that there is a relationship between them, view shapes one at a time to check accuracy, measure the pixel size of objects to determine if they should be labeled according to customer requirements.
- More metadata about the image may be provided as a whole through additional questions added to the workspace below the image.

The output for vector annotation projects is JSON containing the (x,y) coordinates of all shape vertices and their associated labels. This file can be delivered as a .CSV or raw JSON file. The output organizes information by key frames and interpolated frames.

Annotation information around object of interest will abide by below format and syntax:

- Key frames for a single shape and its tags (labels) are listed first, in frame order with frame numbers, followed by all frames.
- For more details, please request a sample delivery file from Sama.

## Cuboid annotation for 3D LiDAR point cloud data

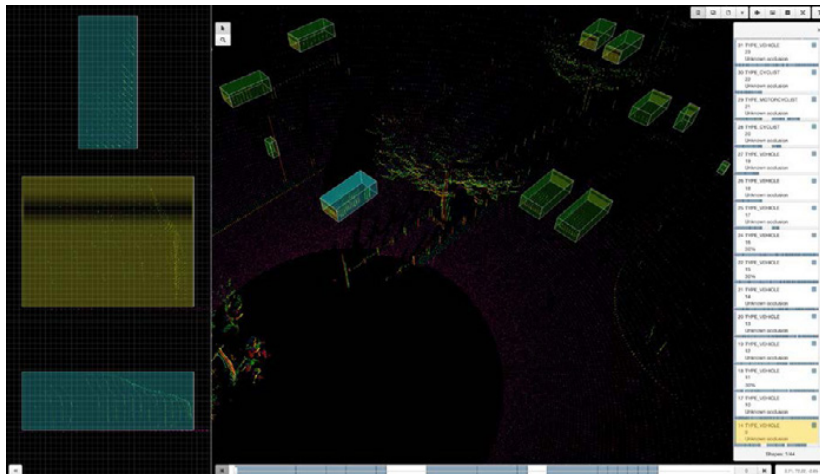


Figure 6: Task workspace view of a 3D point cloud annotation task in the Sama Platform.



A data annotator is presented with a sequence of 3D point cloud data, split up into frames, and asked to outline objects using vector shapes. Annotators may scrub between frames to view the sequence. Multiple 2D reference vantage points can be provided.

- Point cloud assets can have the ground points automatically identified and point intensity values normalized during pre-processing. If the sensor is moving over time, sensor location metadata can be provided to translate the point clouds to a fixed world orientation for improved annotation quality and efficiency.
- Objects may be annotated with a cuboid shape. After drawing the shapes, labels may be added, either from a menu of options or as free text.
- 3D shares many features with Video Annotation:
  - Annotation is assisted by linear interpolation between key locations.
  - Annotated objects have automatically added unique IDs that carry through the entire sequence of frames.
  - The visibility of an object may be changed throughout the video while maintaining the same unique object ID. For example, if an object moves out of the frame and returns, it will have the same unique ID. Visibility settings are visible, occluded, not visible.
  - A single object may be labeled with multiple pieces of metadata, i.e. vehicle make, model, color, and license plate. Object metadata can change from frame to frame (traveling direction, current lane, etc).
- Additional features: Orthographic view, select a cuboid face of interest, adjust cuboid size and position in 2D panels for precision, zoom, pan rotate, add a group label to related shapes to indicate that there is a relationship between them, view shapes one at a time to check accuracy.
- More metadata can be provided about the image as a whole through additional questions added to the workspace below the image.

The output for point cloud annotation projects is JSON containing the (x,y,z) coordinates of all cuboid vertices and their associated labels. This file can be delivered as a .CSV or raw JSON file. The output organizes information by key locations and interpolated frames.

Annotation information around object of interest will abide by below format and syntax:

- Key locations for a single shape and its tags (labels) are listed first, in frame order with frame numbers, followed by all frames.
- For more details, please request a sample delivery file from Sama.

## About Sama

Sama is the final authority on high quality data when it comes to getting AI models into production. A pioneer in data annotation, Sama is the only computer vision solutions company with its own expert workforce using its own innovative enterprise-grade platform. Sama routinely delivers reliable and predictable data quality needed for today's cutting-edge AI investments. That's why Walmart, GM, Ford, Microsoft, NASA, Google, and many others rely on us to help them deliver the best AI that actually works.

Driven by a mission to expand opportunities for underserved individuals through the digital economy, Sama is a certified B-Corp (a designation reserved for companies that balance purpose and profit) and has helped more than 60,000 people lift themselves out of poverty. Its training and employment program were recently validated by an MIT-led Randomized Controlled Trial.