Enable AI Systems to See the World like Humans do.

Supercharge your models' reasoning and factual accuracy with custom instruction-tuning datasets, grounded in real-world videos. Ramblr's advanced video annotations provide multimodal LLMs with precise contextual, spatial, and temporal understanding, boosting robustness, enabling processing of image and image-sequence inputs, and domain-specific fine-tuning.

Ramblr's Annotation Engine

Efficient and scalable cloud-based video annotation solution, purpose-built to provide MLLMs with a structured understanding of the real world and a deeper understanding of the relationships between different modalities

- AI-powered annotation pipeline for video and sensor data – with up to 97% automation
- Cloud-based visualization tool to manage and curate large multimodal datasets
- Cloud-agnostic infrastructure built with scalability, security, flexibility, and GDPR compliance in mind
- Leverage human-in-the-loop feedback to achieve even higher quality and accuracy

Ramblr's Advanced Annotations

High-dimensional data annotations to ground automatically generated context-specific QnA pairs in real world scenarios

Video Instance Segmentation
Spatio-temporal consistency, accurate masks, and category annotations for all context-relevant objects. Precise object coordinates provide an accurate training input and enhance spatial understanding.

Video Description
High-level descriptions of the scenario, intent, activity, and environment to provide a holistic understanding of the scene and context.

Action Annotation
Generate time-dependent action descriptions and correlate activities with sensor data (IMU, gaze, etc.).

Panoptic Video Scene Graph
Structured representation of event sequences consisting of spatial and semantic object-to-object, Ego-to-person, and Ego-to-object relations derived from HOI and gaze signal (t_0: Ego cuts tomatoes with a sharp knife t_1: Ego adds tomatoes to the pan)

Object Attributes
Provide additional object specific attributes to distinguish objects and enable instance awareness of the model (e.g., Where is my house key with the silver Eiffel Tower?)

Ramblr's Instruction-Tuning Engine

Fine-tune MLLMs to specific domains with Ramblr generated instruction-tuning datasets. Grounded in video context, Ramblr automatically generates QnA pairs leveraging contextual annotations

- Combinatorial explosion of context-specific QnA pairs: automatically generating positive and negative instruction samples
- Curated instruction tuning datasets: Enable models to reason and answer complex questions across various domains
- Domain specific adoption: Provide concise, relevant, and informative answers, tailoring their responses to specific questions
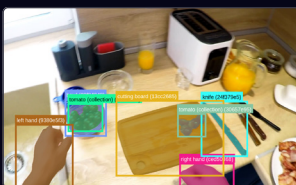- Enable LLMs to ground their reasoning in image, or video, context and alleviate hallucinations
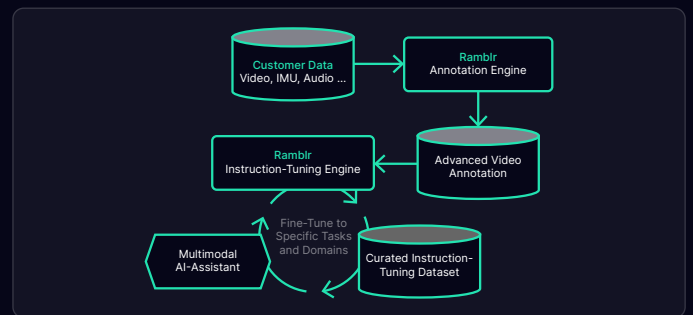
Ramblr Solves the "One Screw Left" Problem
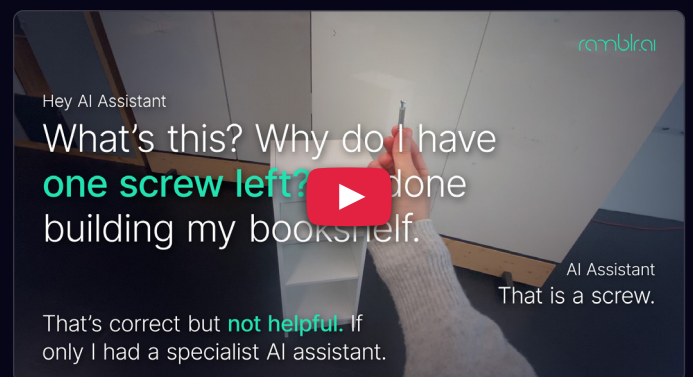
https://www.youtube.com/watch?v=hpI26UxAF_4

Use future and past annotations to formulate QnA pairs automatically

Instruction Samples
Q: How can the and the be combined?
A: The knife can be used to cut the tomatoes.
Q: How can be used to prepare brunch?
A: The tomatoes can be cut and fried in a pan together with other ingredients like mushrooms.

Multi-frame input for the model to learn consistent object representations, to bring current objects in context with other frames, and to summarize sequences
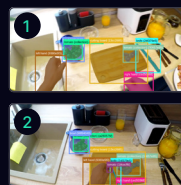
Instruction Samples
Q: Are any object annotations missing in the two frames?
A: No. The knife used to cut tomatoes in frame 2 is annotated in both frames. So are the hands and the cutting board.
Q: Is in frame 1 the same object as in frame 2?
A: Both regions contain hands, but in frame 2 is a right hand. in frame 2 would be the left hand.