Designing for Video Annotation

ALEGION

Introduction

What are the key success factors you need for a Machine Learning Project?

Take a few seconds to think of a list of three or four, but no jumping ahead until you're done!

Got your list? Great. Does it include "user experience"? If not, put it on there. At Alegion we've learned that user experience is foundational to any successful ML effort, and here's why: **A machine learning model is only as good as the data it is trained with, and training data is only as good as its labels.**

How do the majority of datasets get labeled? By people. These people work against tight deadlines and are asked to interpret information across a wide range of unfamiliar problem domains, usually outside the context of their culture or primary language.

If you don't seek to understand the conditions and challenges these people face as they attempt to do their work, it is very difficult to design processes and software that will enable them to create the high-quality annotations required to train models.

TABLE OF CONTENTS

The Challenges of Annotating Video

Alegion provides training data to many industries, and it was our work in the retail space that prompted us to better understand the factors that drive project completion time and cost. We focused our analysis on shopper behavior in self-checkout scenarios because those use cases are particularly lengthy and complex. Our analysis showed us that accuracy and efficiency play a big role in determining project success. Both of these measurements quantify the behavior of annotators as they perform their work, so understanding the experience of an annotator is key.

All data labelling tasks impact an annotator's cognitive load (hereafter referred to as CL). CL influences concentration, mental well-being, and physical energy levels, all of which affect accuracy and efficiency. Compared to image annotation, video annotation can increase impact on CL exponentially because it introduces the variable of time into the equation. An annotator must repeat UI interactions and walk a mental decision tree over and over in order to evaluate the content of each frame.

You can predict the degree of impact by thinking about the annotation task in terms of **intrinsic** and **extraneous** CL.

Intrinsic CL refers to the effort involved with the annotation problem itself. For example, asking an annotator to identify whether or not a pedestrian is present in a crosswalk for a given frame has a fairly low intrinsic CL. Conversely, our case study's use case requires annotators to track multiple people and objects, and their temporal relationships to each other, over long periods of time. This represents high intrinsic CL.

Extraneous CL refers to the effort associated with how the task is presented to the annotator, i.e - the video annotation tooling's user interface, the quality of the video, the device they use to perform the work, the nature of the training material, etc.

Intrinsic CL tends to be fixed because it's dictated by the requirements of the project and content of the video to be labelled. Extraneous CL tends to be malleable because practically every aspect of task presentation can be altered. User-focused research helps us understand which aspects of extraneous CL to address. These insights provide the decision-making information a business needs to mitigate risk and maximize its ROI

Solutions Begin with Research - 3 Guiding Principles

“You must understand the ‘as is’ before you can envision the ‘to be’” is a design aphorism that captures the idea that you can’t solve for a problem if you don’t understand what the problem is in the first place. Research is the cornerstone for well-designed solutions because it is the process by which you understand a problem.

The most effective tool you have at your disposal when it comes to conducting research is curiosity. You might have heard the phrase “Good design requires empathy.” While true, you don’t get empathy for free. It has to be built up over time. Research, by its very nature, deals with the unknown and the unfamiliar. It is hard-wired in our brains to be suspicious of the unknown, and that subconscious bias affects our research efforts. When you make the conscious decision to approach the project with curiosity you are short-circuiting that automatic response in a way that materially leads to you being more observant, asking better questions, and building true empathy for the people you are trying to understand.

The second most-effective way of understanding the “as is” is to become your own research subject . For our project, Product Designers, Engineers, members of the Customer Success team, and even the Sales team used our video annotation solution to complete highly complex labelling work. Nothing builds empathy faster than spending a few hours annotating a video, physically experiencing the eye and hand strain, and encountering an unexplained error, or worse, losing your data.

Third, build your body of research from a variety of sources and types. A diversity of source material leads to more nuanced and expressive synthesis, and more creative possibilities for solutions. The Design team gathered qualitative data by conducting interviews with folks who were directly performing the labelling work, as well as Customer Success and Production Operations team members who were responsible for setting up the task structures. We collected observational data by watching working sessions captured in Fullstory. Lastly, we used Periscope reports that tracked task time and action counts to generate quantitative data .

1 **Curiosity**

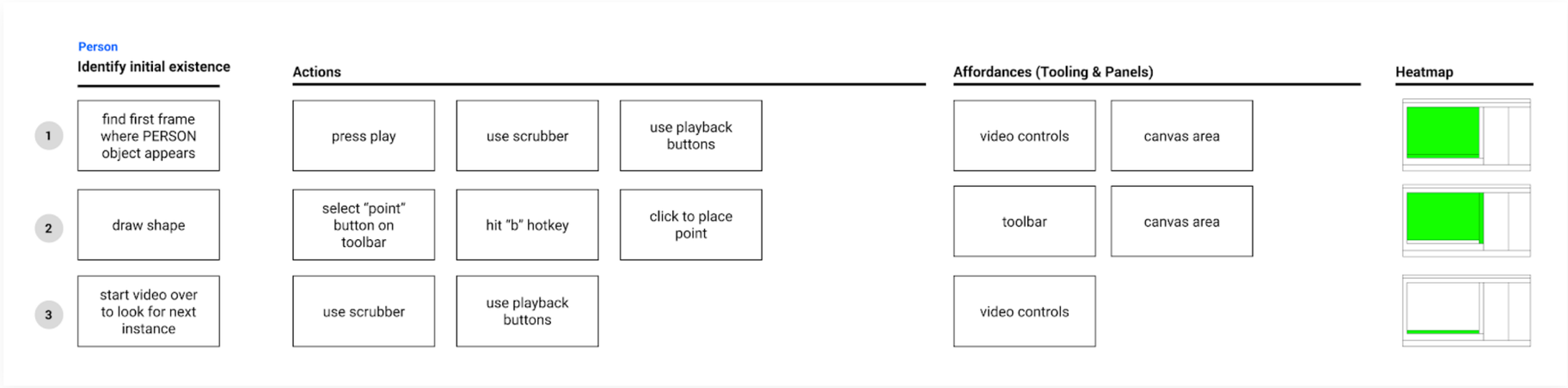
2 **Become Your Own Subject**

3 **Build from Diversity**

Interpreting User Feedback

There are lots of ways to synthesize research findings. In general, synthesis is used to detect patterns through the process of grouping and connecting data points via affinity or theme. Synthesis results in the creation of one or more design artifacts that tell stories and uncover insights. For our project we chose to create a variation of a user journey map . We say “variation” because, regardless of what process and artifacts you choose, it is critical to allow the data to tell its own story rather than forcing the data to fit a specific embodiment.

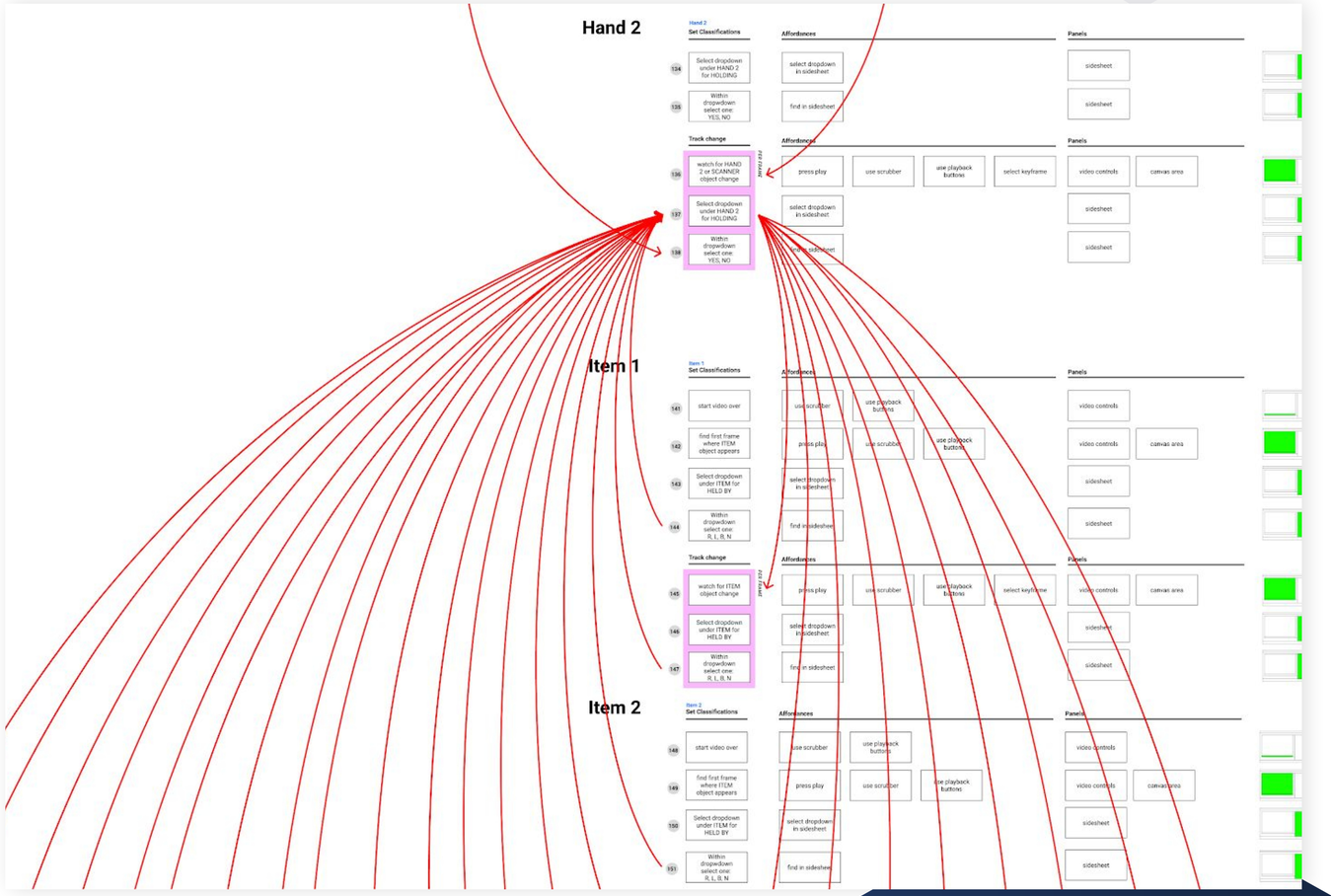
Our quantitative data told such a clear and compelling story that it practically dictated the structure of our synthesis document. We visualized the discrete steps that an annotator takes to complete a task in the form of an algorithm-like flowchart. The screenshot below shows how we grouped an annotator’s physical and mental behaviors into a **Doing** section and identified the UI affordances and focal areas into a **Seeing** section. We augmented this information with a **Thinking/Feeling section** (not pictured).

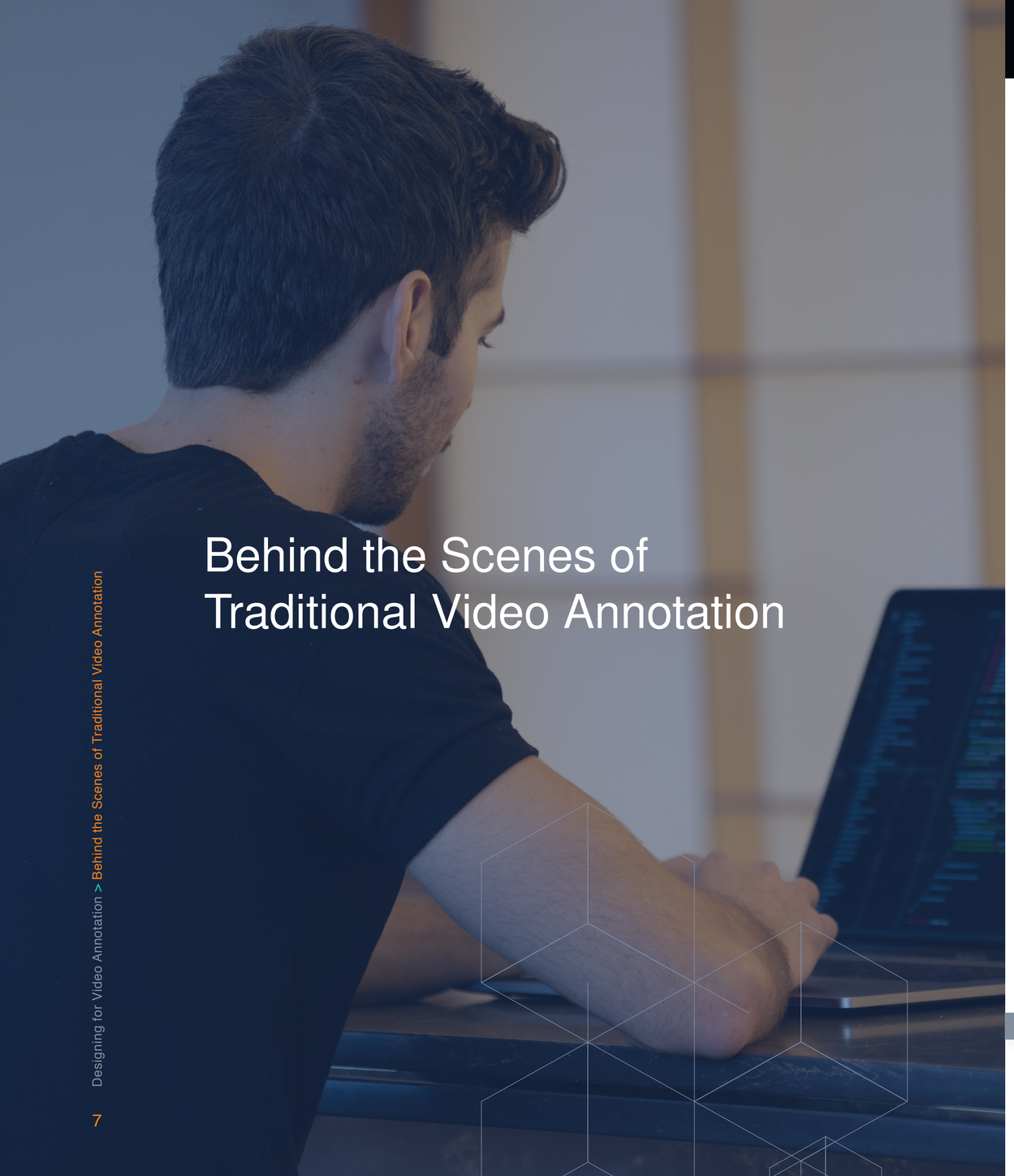


Interpreting User Feedback

In the image to the right, notice how certain blocks of behaviors are grouped in pink. These indicate behaviors that happen on every frame . You'll also see an explosion of red arrows cascading out of the screenshot. Each one of these arrows identifies the tracking of a hand-to-product relationship. For every frame in the video there potentially exists an exponential number of mental calculations that an annotator has to perform in order to capture all of the information permutations

Many videos had upwards of twenty items to be tracked. The complexity of this visual clearly communicates the power of research and the importance of providing properly designed software to annotators





Behind the Scenes of Traditional Video Annotation

One 5 Minute Video Requires:

7,200
FRAMES TO BE ANNOTATED

108,920
ACTIONS TAKEN

48
HOURS OF ANNOTATION

It’s Got To Ship To Count

Our research revealed several ways to improve the video annotation experience.

With so many opportunities available, how do you decide where to focus time and energy?

Design solutions can’t exist in a vacuum. They must consider factors beyond the screen such as Engineering constraints, roadmap deadlines, and business requirements. If a solution only focuses on how something looks or how a workflow behaves but fails to incorporate itself into the larger context of the business, there is a high chance the product will never see the light of day; or if it does ship, its quality will be compromised. Either outcome ends up failing to achieve Product Design’s ultimate goal of meeting the user’s needs.

We struck a balance between the goals of improving the annotator’s experience and decreasing project costs by defining the solution in terms of three strategic focus areas.

This holistic way of looking at the design phase of the project emphasized the need for several organizations within the company to collaborate in order to reach a successful outcome.

1 **IA & UX**

Manage and track entities and their relationships over time.

Control what data gets displayed based on context and task-at-hand.

2 **Feedback & Training**

Easily access guidelines, task instructions, and project changes.

Receive clear feedback on annotation actions and work results.

3 **Experience Engineering**

Optimize task design to reduce training and labelling complexity.

Provide a responsive and stable application environment.

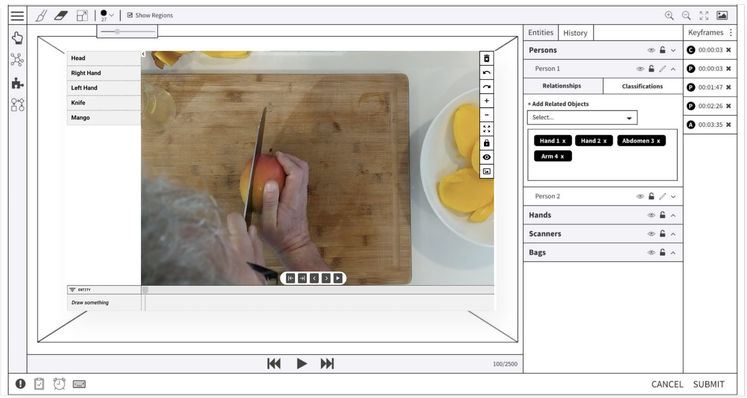
Creating Design Principles

The Product Design team condensed these focus areas down into three design principles. Similar to “Hills” in IBM’s Enterprise Design Thinking framework, these principles serve as landmarks for teams to refer back to whenever there is a question about a particular direction to take or if it becomes unclear what piece of functionality needs to be explored. The key to creating useful design principles is to describe them at just the right level of detail: too vague and you will be unable to meaningfully apply them to a given scenario, too specific and they will only be applicable to a limited number of problems.

1. **Make it easy to track change overtime.**
2. **“Confirm that I’m doing things correctly; help me understand when I’m not.”**
3. **Ensure data output is available and accurate**

It’s worth noting that although the third principal was largely out of the hands of Product Design, it was a valuable reference point that reminded us to consider how our interaction design choices could be used to minimize the emergence of confusing or “unhappy” paths.

With design principles in hand, we set about doing all the designery things one does when designing, such as brainstorming (6-8-5 is particularly effective for generating lots of UI ideas) and creating low-fidelity wireframes to talk through interaction ideas.)



Conducting Usability Studies

We conducted several rounds of usability studies using basic wireframe screens linked together as click-through prototypes. To generate as much feedback as possible we opened these sessions up to the entire company. We also spent in-depth time with the annotators who would be using the tool on a daily basis. Wireframes are a good starting point for exploring rough ideas, but they can only take you so far with a use case as complex as video annotation.

Computer vision-based labelling interactions are analogous with those commonly found in visual design, animation, computer graphics, and video editing software. We catalogued best practices and patterns from those industries and explored ways to adapt them to the data labelling problem domain. For video annotation, we hypothesized that building a keyframe-based timeline similar to those found in animation or video editing tools would go a long way to fulfilling our first design principle.

We had two primary concerns with committing to a timeline solution:

1. What features of a traditional timeline design make sense for video annotation?
2. What timeline behaviors will facilitate information parsing?

The complexity of testing this hypothesis required a more robust exploration tool than wireframes, or even a Figma prototype. We decided to implement a reference UI using HTML, CSS, and Javascript to explore not only timeline interactions and animations, but also ways to improve our localization UX, UI resizing behaviors, and panel collapse/expand interactions.

Conducting Usability Studies

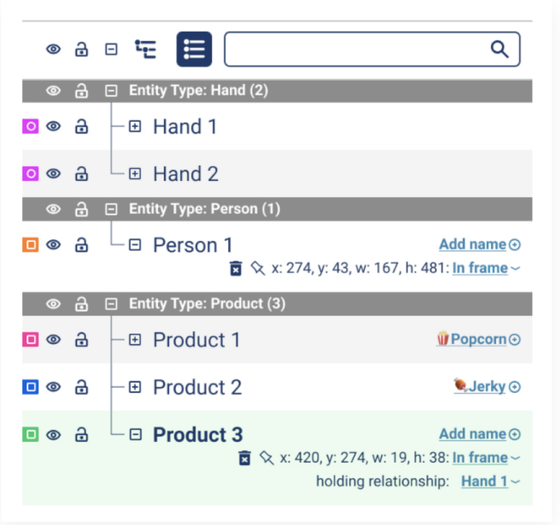
Being able to prototype at this level of fidelity allowed us to quickly explore complex interactions and behaviors, and to discuss ideas and options in real-time with users, Engineers, and other stakeholders. We were able to create an informed and validated design opinion for a potentially risky and complex feature before asking Engineering to dedicate time and resources to build it.

One complex problem we uncovered during prototype was how to render timeline at various zoom levels. For example, if we allowed a user to display the entire timeline for a long video all at once there would be no way to display all the information because there are more frames in the video than there are pixels on the screen. We knew that we wanted to solve this by creating a progressive data aggregation system similar to what you experience when zooming in and out of a map application, but the path ahead was not well-defined. We weighed the risk of sinking more time and resources into the solution against project timelines and business value, and decided to ship with a more basic feature. This buys us time to better understand how annotators will interact with the timeline before adding additional complexity.

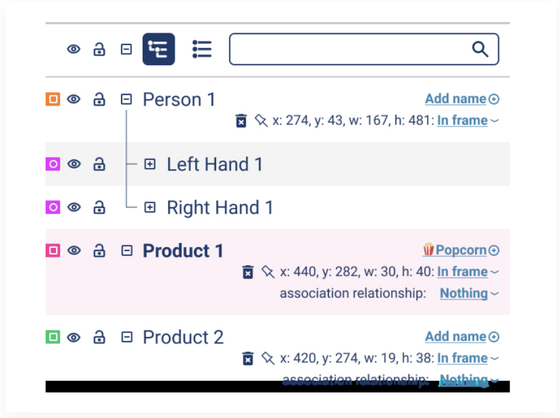
Conducting Usability Study

We adopted a layout pattern commonly found in the layers panel of an image editor for displaying lists of entities and their associated classifications. This structure provides a consistent information architecture and supports high information density.

Instances in the entity list share the same color treatment and selection highlights as their associated localization shapes, which are rendered on the video surface. Color coordination can make it easier for annotators to identify and track related pieces of data that appear in different parts of the interface.



Data views are another important UI affordance. An annotator's mental model of the information they see on-screen changes based on the context of the task at-hand. The previous screenshot displays entities grouped by their type, which is helpful during review to check coverage and consistency. The screenshot to the right shows entities grouped by their hierarchical relationship. This structure makes it easy for an annotator to understand, establish, and manage parent-child structures.



Conducting Usability Study

Displaying a list of tens or hundreds of entities all at once is of questionable value. People are not good at parsing long lists, especially when the list content frequently changes. Our research showed us that annotators are typically focused on tracking a subset of the total information at any given time, so we introduced the ability to use compound searches to quickly and non-destructively filter out data points extraneous to the current context.

It's also difficult to track multiple instances of the same entity type on both the video surface and in the entity list. For example, a shopper in a self-checkout video might have a basket of ten different products that must each be localized. We added the ability for an annotator to give each entity a nickname to make it easier to differentiate between categorically similar pieces of information. It's easier to remember "Lettuce" than it is to remember "Product." Additionally, if the data labeling flow is broken into multiple stages, these names make it easier for future annotators and reviewers to orient themselves to the previously labeled data.

Measuring Success

We shipped our new video annotation product on time and without major issues. That's an internal win for the many folks who gave the project their time and energy, but that achievement doesn't necessarily equate to success. Much like the labelled data of a video annotation project, success must be tracked across time, not examined for a single frame and then forgotten about. Design isn't finished until people stop using your product, and neither is measuring success.

Rely on evidence-based outcomes to keep assumptions in check, minimize data overfitting, validate hypotheses, and to understand what users value. At Alegion, we look to a few pieces of evidence to understand how we are doing.

Version adoption is one way that we gauge how well our products are doing. Because Alegion offers a full-stack data labelling service, we also have our own Global Workforce and Production Operations unit. This effectively creates an internal marketplace where new versions of our tooling become available over time. Both new and long-running projects have the option of adopting our latest releases. If work is not being transitioned over to the "latest and greatest" that is a clear indicator that something is amiss, and we can investigate by interviewing the Customer Success and Production Operations managers.

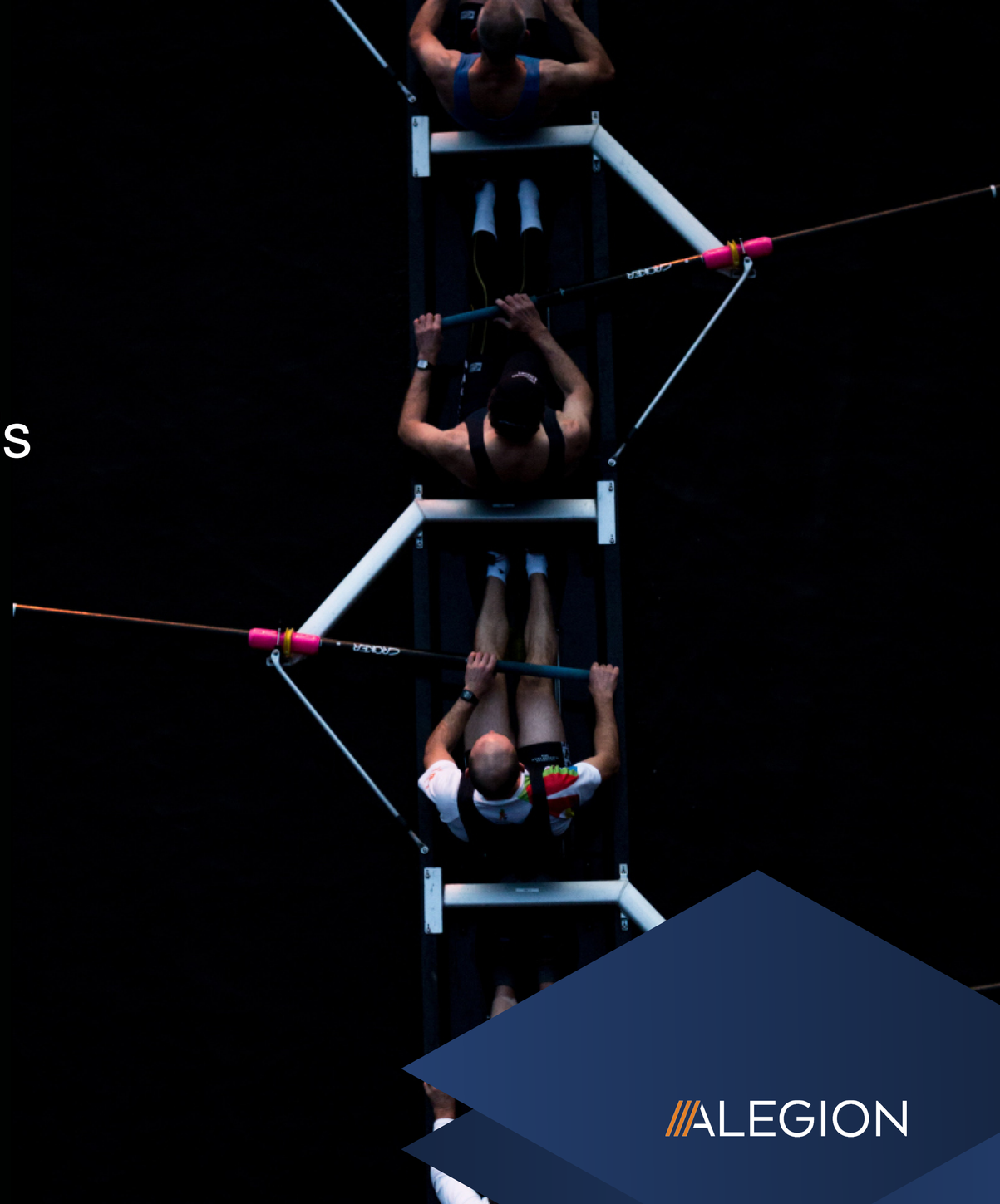
We also gain a better understanding of our product's success through direct experimentation. An example of direct experimentation is conducting time trial comparisons with other video labelling offerings. Repeating the same annotation task across products allows us to identify strengths, weaknesses, and gaps. This is a relatively quick and efficient way to uncover new learnings.

Finally, there is no substitute for user feedback. Our global workforce of annotators use our platform on a daily basis for several hours a day; they understand it in ways Product and Engineering never will. Alegion is a culturally diverse group of folks and for most, English is not their primary language. It's critical to factor this into the equation when eliciting and interpreting feedback. Schedule time to talk to your users, and if possible also ask others who are tangentially involved in the work to collect feedback on your behalf.

We think the first release of our video annotation tool has been successful, but that's a moving target. Our evidence collection has shown us that we need to make improvements around timeline and keyframe navigation as well as making it easier for reviewers to assess the work performed by annotators. We'll be incrementally introducing these features in subsequent releases

Want to learn more about Alegion’s Video Annotation Capabilities?

Reach out to:
Sales@alegion.com



ALEGION