cloudfactory

The Outsourcers' Guide to Quality

Tech innovators depend on high-quality data to get to market fast, outpace competition, and power disruptive products. In the race to usable data, no problem looms larger than the significant time investment required to clean and prepare data.

According to a 2019 study by Cognilytica, more than 80% of artificial intelligence (AI) project time is spent on data preparation and engineering tasks.

Due to this considerable roadblock, the market for third-party data labeling solutions has grown significantly, estimated to reach over $1B by the end of 2023. Yet outsourcing this foundational data work is not without its own unique challenges.

Data quality is vital when creating reliable algorithms. For companies looking for a solution that equals the quality of their in-house team, it can seem as though outsourcing is an impossible option.

Luckily, there are a few key areas that, when analyzed, can give companies a good sense for the level of quality that a data labeling partner is able to provide.

WHAT IMPACTS QUALITY?

There are three dominant factors which stand out as being most important for predicting the level of quality you can expect from a data labeling provider.

PEOPLE

The selection, development, and management of workers for your data labeling project.

PROCESS

How a data labeling provider executes on your work, from onboarding, to task instructions, to quality control workflows.

TOOLS

The work and team management tools and technology put in place to maximize quality outputs and throughput.

But first, let's be clear what we mean when we talk about quality, especially when it comes to capturing useful Key Performance Indicators. Many often confuse accuracy with quality. When measuring quality, it's important to keep in mind that accuracy and quality are two different things.

- **ACCURACY** in data labeling measures for how close and consistent labeling is to real world conditions. This applies whether building computer vision or natural language processing models.
- **QUALITY** in data labeling is about accuracy across the overall dataset. In other words, does the work of all labelers look the same and is it consistently accurate across datasets? This is relevant whether you have 29, 89, or 999 data labelers working simultaneously.

ASSESSING A DATA LABELING PARTNER FOR QUALITY

Evaluating any provider for quality begins by thoroughly understanding the elements that directly impact the caliber of work.

We'll step through each of the key areas, providing insights on what to look for and what questions to ask when exploring your options. We'll also share best practices based on the high-quality work CloudFactory has delivered for clients over the past ten years.

PEOPLE

Quality starts with the people that do the work. The workers' level of experience coupled with the investment made and instruction provided to them significantly impacts the level of work they're able to deliver. This is the heart of human-in-the-loop and the first place you should look when considering a data labeling provider.

There are multiple approaches to developing and maintaining a workforce for data labeling, each with their own strengths and weaknesses. When it comes to the human impact on quality, there are two main aspects of an offering to examine:

VETTING & SELECTION

TRAINING

VETTING + SELECTION

Worker assessment and selection is the first opportunity for data labeling providers to impact data quality. By instituting different skill and personality assessments, providers can ensure that all of their clients are supported by workers with a minimum acceptable skill set. On the flip side, for providers with a totally open platform, there is a high risk for workers who will inherently underperform and provide sub-par data quality.

The CloudFactory Approach

CloudFactory follows a multi-step screening process to vet and select top candidates for our clients' project beginning with the CloudWorker, what we call our valued team members. Everyone is vetted for character, competency, and community. As many grow into team leads, these attributes become a critical example for the rest of the cloud team.

STEP 1: RAW SKILLS ASSESSMENT

This initial evaluation determines a candidate's English literacy and ability to read online documents, analyze images, conduct web research, and complete other basic tasks. These raw skills help us determine competency fit and assign tasks appropriately and according to their strengths, immediately setting them up for success.

STEP 2: CHARACTER INTERVIEWS

We then interview each candidate to ensure they have the desire to grow in character and the heart to take the lead in community service programs. Vetting CloudWorkers with the right expertise and character provides our clients with an invested workforce that helps them stay competitive and agile, while reinforcing CloudFactory's purpose and community focused culture.

STEP 3: CLIENT SPECIFIC TASK ASSESSMENT

Finally, before any CloudWorker is assigned to a client's project, they are tested on specific skills relevant to individual client work/tasks. This crucial step ensures that clients get the best cloud team possible for their unique work requirements.

Team Leads and Team Captains

Each client is assigned a dedicated team lead that works alongside their Client Success Manager to ensure their work and relationship with CloudFactory is successful. The team lead is responsible for training and managing the cloud team and applying best practices from their experience managing similar projects to maximize throughput and quality.

To support the team lead, CloudFactory nominates high-performing CloudWorkers to be team captains. The team captains' primary role is to ensure that communications from the team lead and client are cascaded to all of the CloudWorkers. In addition, they are experts in the task and conduct sampling and review of completed tasks as part of each shift.

TRAINING

Depending on the difficulty and complexity of a task, varying levels of customized training are required to ensure quality outputs and the continued skill development of the data worker.

For very simple yes or no tasks, minimal training may be enough to deliver sufficient quality levels. However, for tasks with a range of complexity, nuance, or subjectivity, higher level training programs ramp up workers quickly while still ensuring quality throughputs.

The CloudFactory Approach

CloudFactory understands proper training is a key part of delivering high-quality data labeling work. We kick off each client project with our targeted onboarding, an opportunity to develop, validate, and execute a custom training program for our clients' hand-picked cloud team.

DEVELOP

A dedicated team lead geographically located near the cloud team collaborates with the client to understand their task through instruction and daily sprints. This train-the-trainer approach removes the burden from the client and ensures consistency in education across the cloud team.

VALIDATE

The team lead tests the process, validates task instructions, and establishes bespoke QA process/controls before onboarding a single CloudWorker. This step ensures the client and team lead are in sync before any workers are added, minimizing errors and accelerating ramp up time.

EXECUTE

Team leads deliver comprehensive training on task instructions, cultural context, and impact. Training is done through a combination of in-person and online programs led by team leads who have now learned by doing. Using this applied knowledge and client-provided materials, team leads tailor to individual learning styles and bring CloudWorkers up to speed as quickly and effectively as possible.

Bigger Picture / Greater Impact

We train team members not only on what to do, but why quality work is critical. We explain what tasks contribute to and the impact of their work on the overall project.

For example, bounding boxes for autonomous vehicles need to be tightly drawn for pedestrian safety. Accurate chatbot training can mean the difference between someone getting direct and appropriate customer support from an insurance provider versus wasting time and potentially getting misinformation when trying to find critical medical help.

These examples underscore the importance of each task by showing the potential human impact.

PROCESS

A solid process ensures a better outcome, plain and simple. Relying on anything less from a data labeling provider can increase the risk of failure in both accuracy and quality.

But good process doesn't have to mean so much rigidity that improvements and flexibility are unattainable. Successful process allows for a scalable approach with tight quality controls and task precision. It also integrates open communication and collaboration for a comprehensive methodology that accommodates use case agility and the ability to quickly pivot to meet evolving business goals.

If you've already been doing the work in-house, you want to make sure that a potential data labeling partner is willing and able to take your pre-existing processes and modify them to work for their workforce. However, if you're starting a new process or want to update the one you are currently using, you want to find a partner that will help you craft one from scratch. Ideally, you'd find a partner that can do both.

The CloudFactory Approach

Our agile processes scale to the needs of any client and apply across all types of use cases. We put robust controls in place to tightly baseline, track, and improve quality for maximum task precision. There are levers in place to manage and improve quality throughout the client lifecycle, from onboarding to production to use case changes.

ONBOARDING

Our onboarding process integrates quality control measures into each step starting with the project kickoff, then through to iteration and optimization, but doesn't end there. This focus on quality drives our ongoing efforts for continuous improvement in process as client projects proceed and new use cases begin.

- **KICKOFF**
  A comprehensive kickoff session ensures every CloudWorker is fully educated on project details and that they have an opportunity to engage with one another, ask questions, remove ambiguity and confusion, and ultimately reduce any risk or complications before the project even begins.

- **ITERATE**
  Our iterative approach has become a vital best practice for CloudFactory's highest standards. The key to this approach is the daily sprints which allow us to test quality factors immediately. Along with these tightly managed daily sprints, we've incorporated a quality scorecard with a simple grading system for clients to provide quick and clear feedback for rapid improvement.

- **OPTIMIZE**
  To elevate and optimize quality outputs, our processes incorporate some core fundamentals; basic essentials that distinguish a quality-focused, managed team from the anonymous workforce. These fundamentals include quality over speed, pair learning for new team members, and CloudWorker incentives like leadership opportunities and metric-driven bonuses.

PRODUCTION

As CloudWorkers ramp up, get used to the task, and start meeting throughput requirements the need for client input also decreases. We have developed a recommended review cycle which our clients can taper as they go, eliminating massive and long-term time constraints. Below is a recommended review schedule for most CloudFactory clients.

TIMELINE *Approximate*	**RECORDS REVIEWED BY CLIENT** *Recommended as best practice*
Week 1	**100%**
Week 2	**50-100%**
Week 3	**20-50%**
Week 4+	**5-20%**

Additional optimization efforts are incorporated as part of the overall performance improvement process. These efforts include:

- Visual insights from quantitative and qualitative feedback and actions
- Reassignment or offboarding of team members when necessary
- Best-practices and shared insights from other WorkStreams
- Agility and flexibility to accommodate scope changes

All of these best practices and project fundamentals ensure a cohesive process for the highest quality standards and delivery outputs. The strength of this methodology promises greater task precision, proficiency, and accuracy on the most complex data labeling projects.

USE CASE CHANGES

As clients' needs and goals change over time, we work with them to design new processes that enable their cloud team to transition quickly from one task to another. Client task review and CloudFactory's own internal QA processes increase for a short period of time to catch edge cases and ensure task understanding. However, due to the cloud team's understanding of the client's business and existing processes, time to full throughput and quality is typically much shorter.

TOOLS

Like any project or task, without the proper tools, you simply can't do a good job. If you're stuck with outdated or incompatible technologies, productivity decreases and quality suffers. Additionally, the financial implications can dramatically slow down and even halt many well-intentioned and innovative projects. On the flipside, implementing effective tools can improve outcomes, increase speed, and reduce project costs.

For example, one of our clients has seen seen a 3x throughput increase just by embedding a Google Maps view into the same screen rather than opening Google Maps in a new tab. A simple tool enhancement that had a major impact.

When it comes to quality, you can put tools in three distinctive buckets:

ANNOTATION | **COMMUNICATION & COLLABORATION** | **PERFORMANCE MONITORING**

ANNOTATION

If a provider requires you to be locked-in to their annotation tools, the technology may not be up to achieving your unique task or use case. For example, if the provider has been tasked with a complex image annotation project but doesn't offer a robust enough tool to allow the broadest set of annotations according to client needs, the quality falls significantly short, as does the ability to leverage the data work in the most advanced applications.

The CloudFactory Approach

At CloudFactory we believe that a tooling agnostic approach is the only way for us to provide the best quality for our clients. Instead of forcing clients to work in a tool we created, we partner with best of breed tooling providers while also leaving the door open to take on work using our clients' own tools. In addition to maximizing quality outputs, this allows us to invest our focus and resources in the area where we excel - the managed workforce.

COMMUNICATION AND COLLABORATION

It almost goes without saying, but communication and collaboration is key to good quality outputs for tasks with any level of complexity. Data labeling providers should have a comprehensive, unified platform for clients to quickly

convey feedback and task adjustments before the data work goes too far, minimizing rework, lost time, and higher resource investments. And communication doesn't go just one way, workers should be able to ask questions and share suggestions to improve processes, tools, and outcomes.

The CloudFactory Approach

Communication, collaboration, and transparency are vital to CloudFactory for achieving our highest in quality standards. Within the WorkStream application, messaging channels enable a seamless quality feedback loop between the client and Team Lead. This functionality provides increased visibility and transparency so that everyone involved can stay up-to-date on progress, activities, and future use cases.

PERFORMANCE MONITORING

As Peter Drucker said, "If you can't measure it, you can't improve it." Performance reporting is one of the most important inputs that helps both client and data labeling provider manage quality. But, you have to measure the right things.

A good data labeling partner will work with you to understand what your quality measures are and tailor their reporting tools to ensure they are tracking the things you want to optimize. One-size-fits-all solutions can get you part of the way there but you'll be missing out on maximizing those metrics that really matter most to you.

The CloudFactory Approach

To provide proper reporting to the client and team lead, each of our Workstreams is enabled by a suite of workforce management tools that allow us to track worker performance and overall Workstream health. Clients can log into the WorkStream app to see key metrics on demand.

SECURE BROWSER

Our exclusive browser technology is our unified view into all client and project data and the tools leveraged. It collects engagement data like keyboard strokes and mouse clicks which can be analyzed by team leads. This engagement data, in conjunction with throughput stats from the client, help maintain an efficient team. The browser guides CloudWorkers through Pomodoro bursts, our choice for time management.

Pomodoro

CloudFactory workers follow the "Pomodoro" technique of working, guided by our custom WorkStream Browser. This technique involves dividing work into timed intervals of 25 minutes which are spaced out by short breaks of 5 minutes, completing a 4-hour shift, twice a day.

The length of the Pomodoro cycle is adjusted based on the use case and worker performance insights. This results in higher productivity, accuracy, engagement, and retention rates. Short sprints ensure workers are focused and consistently productive. Regular breaks boost motivation and creativity.

All of this promises a more consistent team that builds on task knowledge and functions as an extension of your team.

THROUGHPUT MONITORING

If integrated in the clients work tools, our throughput monitoring tool can fire task signals from their environment to visualize real-time throughput metrics and provide further predictive analytics. Measures such as tasks per hour are sent to our WorkStream app providing visibility into worker performance. We can adjust the Pomodoro cycle based on those insights from our dashboard.

KEY QUESTIONS:
ASSESSING A DATA LABELING PARTNER FOR QUALITY

Now that we've walked through the three main elements that impact quality, you've got all the context you need to make an informed decision about what partner can provide the data quality you need. To help you approach this conversation, here are a few starter questions:

- Do you offer dedicated **success managers and project managers**? How will our team communicate with your data labeling team?
- How do you screen and select your **workforce**? Will we work with the same data labelers over time? If workers change, who trains new team members? Describe how you transfer context and domain expertise as team members transition on/off the data labeling team.
- Is your data labeling process flexible? How will you manage **changes or iterations** from our team that affect data features for labeling?
- How do you manage **quality assurance**? How do you share quality metrics with our team? What happens when quality measures aren't met? How involved in QA will my team need to be?

Are you ready to talk outsourcing?

UK • USA • NEPAL • KENYA
contact@cloudfactory.com
CloudFactory.com

cloudfactory