INTELLIGENT CROWDSOURCING:
What the Next Generation of Crowdsourcing Looks Like
• Introduction
• Traditional Crowdsourcing’s Shortcomings
  • It’s Work
  • It’s Anonymous
  • It’s Unsophisticated
  • It’s A Hassle
• Intelligent Crowdsourcing: Overview
  • Productivity Instead of Work
  • Community Over Crowd
  • Advanced Targeting
  • Machine Learning for Optimization & Quality
  • Engaging Interface & Task Design
  • Mobile-First
  • Relevant Use Cases
• Intelligent Crowdsourcing: Details
  • Use Cases & Task Types
    • Data Enrichment (aka Data Classification, Data Completion, Data Attribution, etc.)
    • Data Clean-Up (aka Data Cleansing, Data Verification, Data Validation, Data Authentication, etc.)
    • Data Labeling
  • Real-World Examples of Intelligent Crowdsourcing in Action
    • PivotDesk Saved $100K+ & Eliminated 100% Of Inappropriate Images By Automating With Spare5
    • Avvo Forecasts 60% Increase In Onboarding Based On Results With Spare5
    • Moz Saves Time, Costs, & Effort Using Spare5 To Collect Machine Learning Training Data
  • The Business Value of Intelligent Crowdsourcing
• Summary
• About Spare5
INTRODUCTION

The advent of crowdsourcing was a game changer for the business world. It enabled companies to break up overwhelming data projects into thousands or even millions of tiny, individual tasks for human “workers” to complete (usually for pennies). It empowered businesses with a remote workforce to which to outsource data projects — projects that quickly became mind-numbing to any one human to work on for too long, and that computers could not readily handle (at the time). It was a breakthrough invention — but that was over a decade ago. And as has becomingly increasingly apparent in today’s world of artificial intelligence (AI), machine learning, personalization, and optimization, crowdsourcing in its original form is no longer adequate.

Traditional crowdsourcing was not designed, for example, to provide training data for machine learning models. It was never designed to automate critical business processes. It was never designed to handle complex tasks that require specialization and thereby enable employees to focus on their highest value-add (example: engineers offloading the annotation and attribution of products so they can focus on writing code). Crowdsourcing has the potential to encompass all of these goals — but only in an evolved form of its original design.

TRADITIONAL CROWDSOURCING’S SHORTCOMINGS

1 IT’S WORK

This is obvious, of course, but the key here is that it is thought of and set up as work. Crowdsourcing “workers” can potentially sit down to hours-long sessions of monotonous, nearly mindless tasks — and are very poorly compensated, from a conventional work perspective. People don’t particularly enjoy completing rote tasks for pennies for hours on end, yet crowdsource platforms and their customers make little effort to keep workers alert and engaged. Crowdsource workers have complained endlessly about benevolent addiction, “serfdom,” disappearing work, low payouts, and even post-traumatic stress disorder. Imagine what this means for the quality of their efforts.

2 IT’S ANONYMOUS

Crowdsource workers are by definition members of a crowd. Members of a crowd are largely anonymous. The problem with anonymity? It breeds cheating and poor performance. It’s far easier to deliver sub-par results or scam the system when one’s name or reputation isn’t attached to the action. Crowdsourcing customers constantly complain of fraud, bots, and high noise-to-signal ratios — all undoubtedly a result of, or exacerbated by, anonymity.
Academics have published terrific best practices on how to apply algorithms, workflows, and game theory to improve results, but the problem remains: lack of clear, personal identity means lack of accountability.

3 IT’S UNSOPHISTICATED

Crowdsourcing companies were created to solve simple problems with massive scale, but today, computers are beginning to emulate humans’ ability to perform rote tasks — what those “simple problems” break up into — with better scale and speed. Thirteen years ago, computers couldn’t reliably discern the subtle differences between, for example, two pages selling the same product, or when an image contains fruit versus faces — humans had to do it. That’s changing. Fast. Al, machine learning, and the broad availability of massive computing power are advancing with breathtaking velocity, displacing a lot of traditional crowdsourcing work.

Humans are still very much needed, of course. Machine learning models are only as good as their training data — and that comes from humans. Traditional crowdsourcing didn’t account for this use case in its inception, though, nor the ability to handle complex tasks. Crowdsourcing as a concept is still a smart solution to data problems, but once again, an updated model is required.

4 IT’S A HASSLE

Traditional crowdsourcing requires a lot of hard work for its customers. It requires experts, often including a PM, data scientist or engineer, and someone with the design and research chops to create or customize an interface and its questions. It’s a lot of effort to sort out the “signal from the noise” of largely anonymous workers completing rote tasks in a model that treats humans as an on-demand, commoditized API. High investment and questionable return make for a weary group of users. It also means that teams are not focusing on their core competencies, goals, and products.

Over the years, some traditional crowdsourcing customers have engaged middleware providers, but many others have given up and sought alternatives because it’s just too much hassle.

INTELLIGENT CROWDSOURCING

OVERVIEW

Intelligent crowdsourcing is the new and improved version of crowdsourcing. It’s a model that recognizes and solves for today’s data problems. It’s all the good of traditional crowdsourcing (human insights, scalability, speed) with new features and functions to address the bad. Those features and functions: productivity instead of work, community over crowd, advanced targeting, machine learning for optimization and quality, engaging interface and task design, mobile-first, and relevant use cases.

A bit more detail on each:
Intelligent crowdsourcing is not a work model. It does not aspire to be a proxy for a job; it does not aspire to involve “workers.” Instead, intelligent crowdsourcing is a way for people to spend their free time productively — the spare moments they have waiting in line for coffee or commuting to work by bus or train. Intelligent crowdsourcing’s taskers — not “workers” — do not endure disengaging work sessions for a measly compensation. They participate voluntarily in intelligent crowdsourcing as a way to contribute their attention and time to something productive (which would likely otherwise be spent on checking social media networks for the nth time that day, or playing a game that might even cost them money), and to make a bit of spare change as they go — as a bonus. As a result of this fundamentally different mindset, intelligent crowdsourcing sources the best insights from individuals’ optimized performance. When taskers seek short-term stimulation, they engage. When they’re done, they sign off.

Imagine, for example, that you need 20 million insights from tasks that require roughly 30 seconds each. In theory, you might be able to hire 100 full-time people to complete them in a year (at roughly 100,000 productive minutes per employee-year). In practice, you’re much more likely to get quality data if you ask 100,000 people to complete them in an average of 100 minutes per person — and at least an order of magnitude faster. This scaled model doesn’t function for all types of work, but when you need confident human insights at scale, it changes the game in terms of quality, speed, and use cases.

This next wave of crowdsourcing rejects the crowd mentality, and instead intentionally seeks community. The difference: members of a community have identities, communicate with each other, feel connected, feel valued, and feel accountable. With accountability comes quality output (or “insights”). And with identity comes profiles. All people have unique demographics, specialities, skills, interests, and performance reputations. For a crowdsourcing solution to handle complex tasks, it must know, track, and empower the use of its community members’ individual qualifications. Intelligent crowdsourcing accomplishes this. An intelligent crowdsourcing platform encourages its members to communicate, connect, help, and challenge each other.

With a known community of identified members, crowdsourcing users can get the specialized insights they need for complex problems — via targeting. In the new, intelligent model of crowdsourcing, businesses can target taskers who match specific profiles: a fashionista, a medical expert, a pet owner, a pro writer, a new mom, etc. It’s no longer acceptable to “spray and pray” for data projects, as is the case with old-school crowdsourcing. Targeting is non-negotiable.
MACHINE LEARNING FOR OPTIMIZATION & QUALITY

Machine learning has affected crowdsourcing in other, perhaps more positive, ways than just the aforementioned displacement of work. Crowdsourcing in an intelligent form uses machine learning algorithms to a) automatically identify the best potential taskers for a given task based on past performance indicators, and b) filter results for an exceptionally high level of quality. At this point, a crowdsourcing solution that is not making significant use of machine learning is undoubtedly inadequate (and headed for extinction). The best of breed continuously optimize themselves, increasing the efficiency with which it assigns tasks, establishes ground truth, and completes batches.

ENGAGING INTERFACE & TASK DESIGN

The best platforms have community members who play their tasks as much as they complete them. Intelligent crowdsourcing keeps taskers alert and fulfilled with game-inspired user interfaces and task designs. High-quality insights are not likely to come from tired, bored, unfocused taskers — great, purposeful design counteracts that. Incorporating learnings from psychology, gaming, and more, intelligent crowdsourcing prioritizes fun, rewarding user experiences to ensure happy taskers and quality results.

MOBILE-FIRST

To reach people in their free moments, and enable them to spend those moments productively (a goal of intelligent crowdsourcing), a crowdsourcing solution must be available for use by taskers anywhere at any time. Everywhere we go, each of us sees one of the world’s three billion smartphone users looking for something productive to do on the computer in their pocket. Additionally, desktop-based platforms alone often cannot produce the volume and velocity today’s businesses require for their data projects.

RELEVANT USE CASES

An intelligent crowdsourcing solution accounts for use cases in artificial intelligence and involving complex tasks. Data scientists are in desperate need of clean, labeled training data for machine learning models. Businesses need humans — often specific humans — to rank, label, edit, classify, categorize, attribute, associate, etc. their unstructured data and content that’s growing at a dizzying rate. While traditional crowdsourcing platforms can be used for such specialized tasks, it’s not what they were built for. Intelligent crowdsourcing had these things in mind from the beginning.
So what precisely do these new use cases and task types look like in the intelligent model of crowdsourcing? What are some real-world examples of companies using it? How can the human insights gained via intelligent crowdsourcing be used to boost business? Below, the details:

**USE CASES & TASK TYPES**

**Data Enrichment**
(aka Data Classification, Data Completion, Data Attribution, etc.)

It’s well-established that to thrive today, businesses need high-quality, reliable data. The challenge is that managing data like images, documents, social media comments, videos, and audio is often chaotic. By having all that “unstructured data” enriched — categorized, classified, tagged, edited, etc. — companies are better able to use it, and take action on it.

For example, do you have all the right, specific, subjective product attributes in place so that your search engine delivers the right results to the right customers? Does Google drive traffic to all your pages, or your competitors’? The right human insights into your unstructured content addresses these scenarios (and make these data undeniably more valuable).

How does intelligent crowdsourcing help with this? Those right human insights. Micro-tasks for data enrichment via human insights include content classification; rating; adding / editing keywords, tags, titles & descriptions for media; video scrubbing; image editing; association (e.g., outfit creation and “more like this”); surveys; and more. As taskers complete the tasks, the associated human insights are gathered, validated, and organized into a structured format. **With these high-quality data, companies can train and feed tools and algorithms that improve conversion rates, the search & browse experience, SEO, and machine learning capabilities.**

**Data Clean-Up**
(aka Data Cleansing, Data Verification, Data Validation, Data Authentication, etc.)

Inaccurate and out-of-date data is simply inexcusable these days — Google has effectively reset our collective expectations for online (or internal) information. When a company’s info is invalid or incomplete, it loses. It’s that simple (and frustrating). But no doubt, it’s a huge undertaking to keep all product details, directory listings, prospect contact info, and other information clean and accurate. That’s where intelligent crowdsourcing helps — a community can verify and update a business’s data, helping organizations realize the benefit of high-quality data, such as increased customer loyalty, reduced support costs, accurate directory listings, improved marketing campaigns, and increased sales effectiveness.

Which data needs cleansing? Generally directory listings, product details, and prospect contact info is what gets stale and annoying to update for most companies. These data can reach astronomical numbers, and can be extremely difficult to stay on top of. Intelligent crowdsourcing is the solution.
In the data clean-up use case of intelligent crowdsourcing, community members tackle a variety of micro-tasks, like making phone calls to verify business info, scouring the web to retrieve key data, taking surveys, transcribing media, and more to help companies clean up and maintain their data. By creating and dispersing data cleansing tasks to proven members of the community, businesses can keep their data clean in a scalable, cost-effective way, while improving their web content, and marketing and sales efforts.

Data Labeling

The world’s best data scientist can build the world’s best machine learning algorithm or artificial intelligence model, but without a supply of high-quality labeled training data, the technology will never be able to reach its full potential. In fact, depending on the quality of the training data, it may even be quite limited.

Many data scientists have historically used traditional crowdsourcing methods to generate large amounts of training data, but the quality was low while the effort was high. Intelligent crowdsourcing solves for that. Rather than needing to lean heavily on traditional crowdsourcing (or panels, outsourcing, interns, and in-house editors) intelligent crowdsourcing customers are finding that this new model of crowdsourcing provides the transformative, trustworthy training data they need — quickly and easily.

Customers with machine learning projects have run a variety of data labeling tasks through intelligent crowdsourcing like tagging, rating, classifying, titling, associating, comparing, annotating, and survey-taking. For clean, accurate, training data — and a scalable, cost-efficient way to get it — intelligent crowdsourcing wins every time.

REAL-WORLD EXAMPLES OF INTELLIGENT CROWDSOURCING IN ACTION

At Spare5, we’re seeing the positive results of intelligent crowdsourcing in use every day with our customers. Here are a few notable examples:

PivotDesk Saved $100K+ & Eliminated 100% Of Inappropriate Images By Automating With Spare5

PivotDesk, an office-sharing online marketplace, was using a manual process to review, filter, organize, and give feedback on the many images their users upload to offer their excess office space to office seekers. The images are a critical part of the website’s experience and heavily influence conversion, and PivotDesk had four in-house employees — General Managers — handling the photo processing as part of their overall workloads. PivotDesk knew the employees’ time could be better spent, but also realized it was an important task that required focus, judgement, and skill — neither something a computer could handle nor just any unknown human.
Avvo is an online marketplace that connects consumers to attorneys. To remain a trusted source of legal info and lawyer profiles, Avvo must have a constant, steady supply of engaged attorneys available to interact with and serve customers. A major component of achieving this objective is onboarding new attorneys to the marketplace, and one big roadblock Avvo was running into in their onboarding efforts was a significant amount of gaps in key data.

Avvo was using some internal tools to address the data gaps, but the production volume was unpredictable. They found they couldn’t confidently rely on incoming info from the internal tools to supply the much-needed data in a timely manner, nor to make future plans for sales and marketing activities.

Avvo's data challenge was a perfect match for Spare5’s data clean-up expertise. We created custom web-based micro-tasks specifically designed to gather the info Avvo needed to fill in the holes in their data set, and we pushed them out to qualified, proven members of our community. Avvo was able to verify within a week that 90% of the intelligent-crowdsourced insights met their requirements, which was far superior than what they’d experienced with other solutions. The company is now projecting a 60% increase in onboarding over last year based on the data they’re able to collect with Spare5.

Moz is a marketing software and resources company. The company’s data science team had set out to build machine learning algorithms that extract structured information — specifically, author information — from web pages. Moz needed high-quality, labeled data to develop, test, and validate these algorithms.
Moz had sourced this kind of data with manual and/or semi-automatic internal processes in the past, but found that with this project, they needed an external solution. Spare5 created custom micro-tasks for our community to gather the data Moz needed to train their algorithms. When Moz received the first batch of data from Spare5, they compared the answers with what they’d done manually. Where there were discrepancies, the Moz team sided with the Spare5 answer versus what a Moz product manager had done most — “certainly more than half” — of the time. Based on this, Moz knew they could trust the Spare5-supplied data.

Moz found a lot of value in the ease of use of Spare5’s platform. They saved a significant amount of project management and QA time by using Spare5 over a traditional crowdsourcing solution or handling it internally, and they felt the cost was “little money considering the value.”

Read the full story here.

THE BUSINESS VALUE OF INTELLIGENT CROWDSOURCING

With many and varied use cases come many and varied benefits. Simply put, the human insights that result from intelligent crowdsourcing can be used to improve a website’s search and browse experience, a company’s critical online and offline data, and a machine learning algorithm’s “intelligence.” These benefits beget more and bigger benefits, including: improved SEO, increased conversion rates, increased customer loyalty, reduced support costs, accurate directory listings, improved marketing campaigns, increased sales effectiveness, and “smarter” machine learning models.

It’s clear that intelligent crowdsourcing is a valuable, much-needed solution for any company with unstructured data and/or complex data needs.

SUMMARY

Traditional crowdsourcing created a new industry more than ten years ago. However, more powerful compute power and software are solving more of the rote tasks crowdsourcing was originally designed to address with commoditized labor as an API. Today’s businesses and data scientists have new, intricate data needs, which the old model of crowdsourcing cannot readily address. A new model is necessary to be effective in our world of increasingly intelligent machines and relentless demand for personalization and optimization. Intelligent crowdsourcing is that updated version, and customers are already benefitting from this modern model.

Unlike traditional crowdsourcing, the death of which was inevitable from the beginning, intelligent crowdsourcing is built to adapt, grow, and survive.
ABOUT SPARE5

Spare5 is the intelligent crowdsourcing platform that delivers human insights for unstructured data. Companies work with Spare5 to clean up and improve their data and to gather accurate, labeled training data for machine learning models. We break down these large data needs into micro-tasks for humans to complete and provide their insights. We distribute tasks only to specific, qualified, proven members of our community. We then apply our proprietary QA processes and algorithms to ensure we provide only high-quality human insights. Forget old-school crowdsourcing; it’s time for an intelligent crowdsourcing solution.

Helpful links:
- How It Works
- Resources

Interested in learning more about how you could use intelligent crowdsourcing? Get in touch with us here.