



sama

REPORT

2023 ML Pulse Report

The latest trends and challenges in machine learning—as reported by ML professionals

Machine learning continues to explode—but plenty of details, practices, and evaluations need to become more sophisticated to be truly useful. Translating business needs into successful models is hardly a linear path, and growing pains are still evident.

We wanted to understand the biggest concerns, challenges, and motivations for ML professionals. And, of course, what they think about the biggest tech development in these times—Generative AI (GenAI).

Will GenAI change the world? What will the ML industry be thinking about over the next 12 months? We surveyed ML pros to get a read on these questions, and more.

Key takeaways

01

Generative AI is hyped and still maturing—but **yes, it is a big deal**. From routine tasks to its inevitable impact on the economy as a whole, GenAI will be the tech game-changer.

02

Model success metrics are varied and not always linked to business outcomes; this will be an area that will develop as the ML field matures and businesses become more cohesive.

03

There's an overall lack of model confidence—many we surveyed said they have concerns that their models won't survive production and scale correctly—and the root cause is lack of resources. No high-quality quality assurance, no confidence.

04

With some imagination and experimentation, the uses of GenAI will proliferate. But it is, at the end of the day, still a tool. Approach it as one.

1 | GenAI Insights

Hyped But Excited

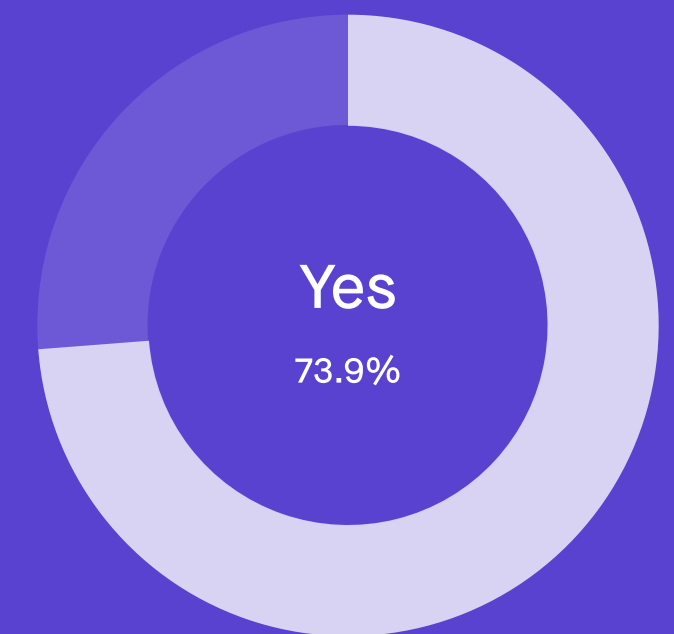
The claims are big: Generative AI will change the world. Is it true or just a bunch of hype?

We think it's true—but it's going to take longer than the hype suggests.

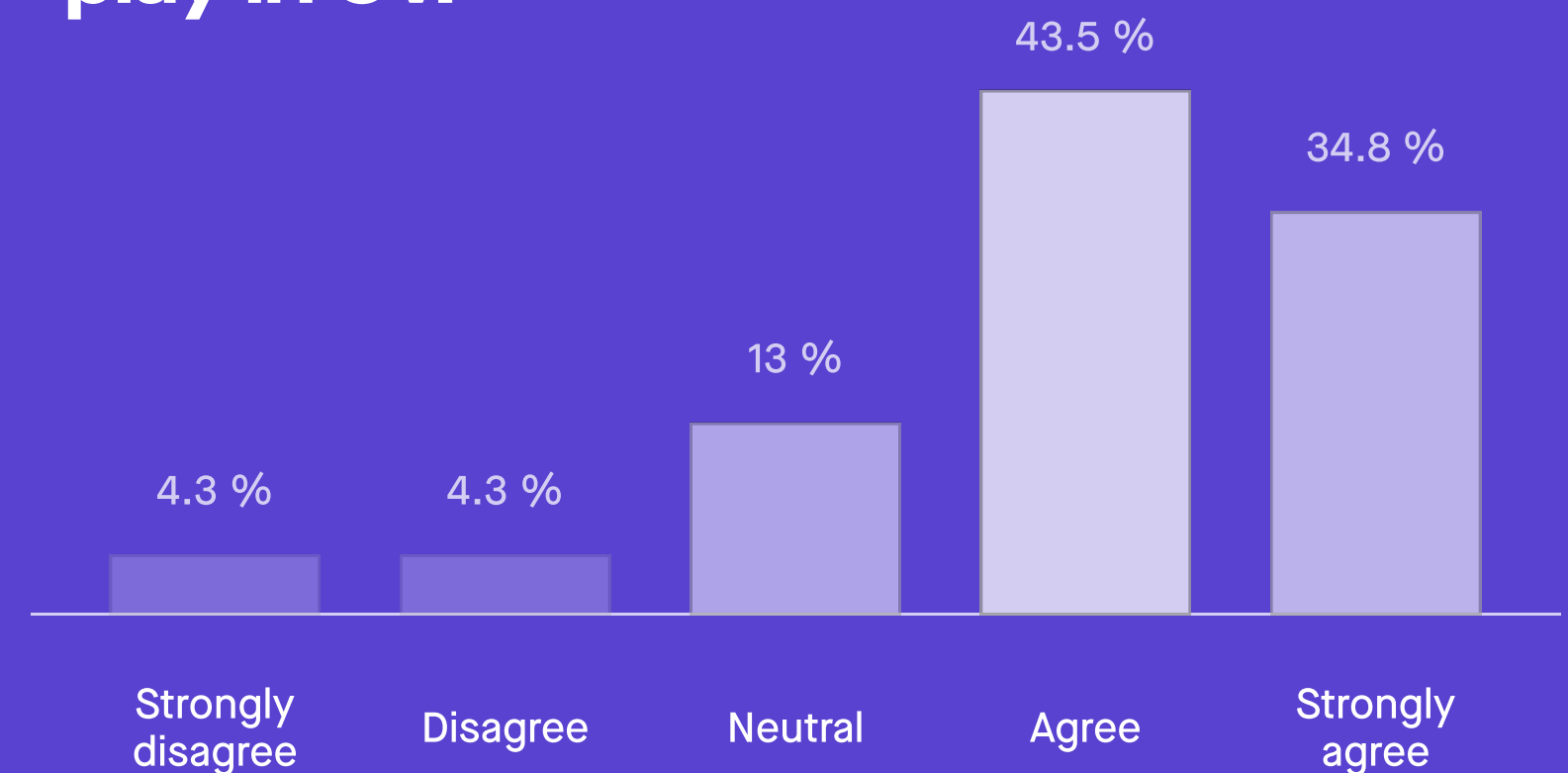
GenAI is perhaps the biggest technological advancement since the mobile phone. As with many advanced technologies, the initial enthusiasm can overshadow the reality. But the reality in this case is still pretty good. ML engineers simply don't quite know yet exactly what that impact will be.

Nearly 74% of respondents said that GenAI for computer vision lives up to the hype, yet most of those—61%—also think we're not there yet. As the technology matures, which could take years, its promises and uses will be better realized.

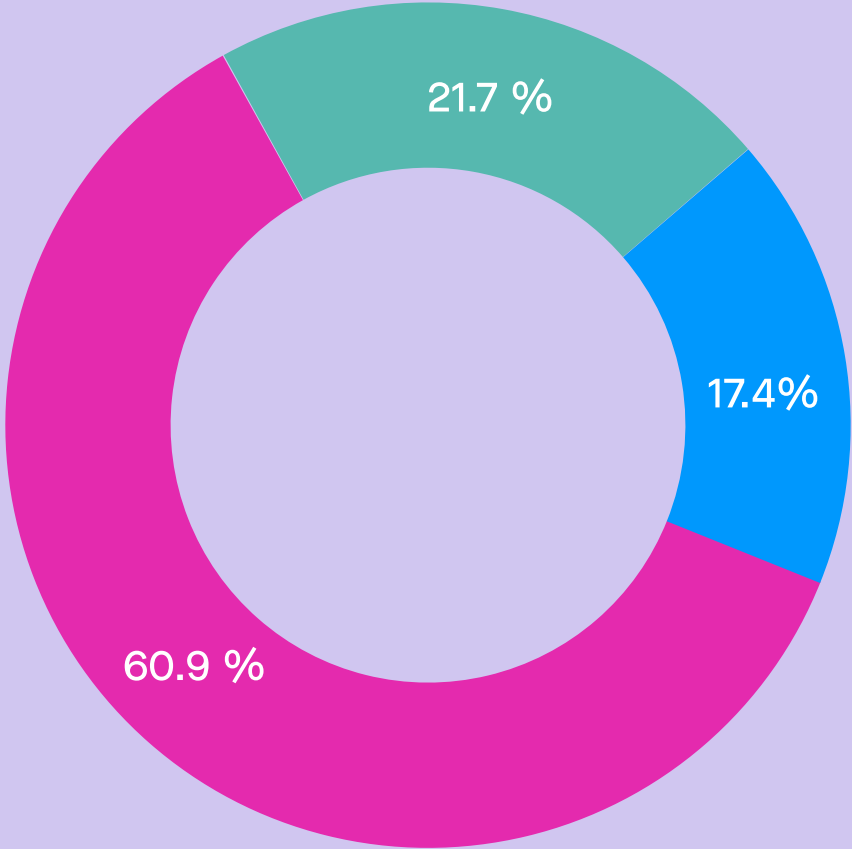
Do you think
GenAI is worth
the hype?



“I’m excited about
the role GenAI will
play in CV.”

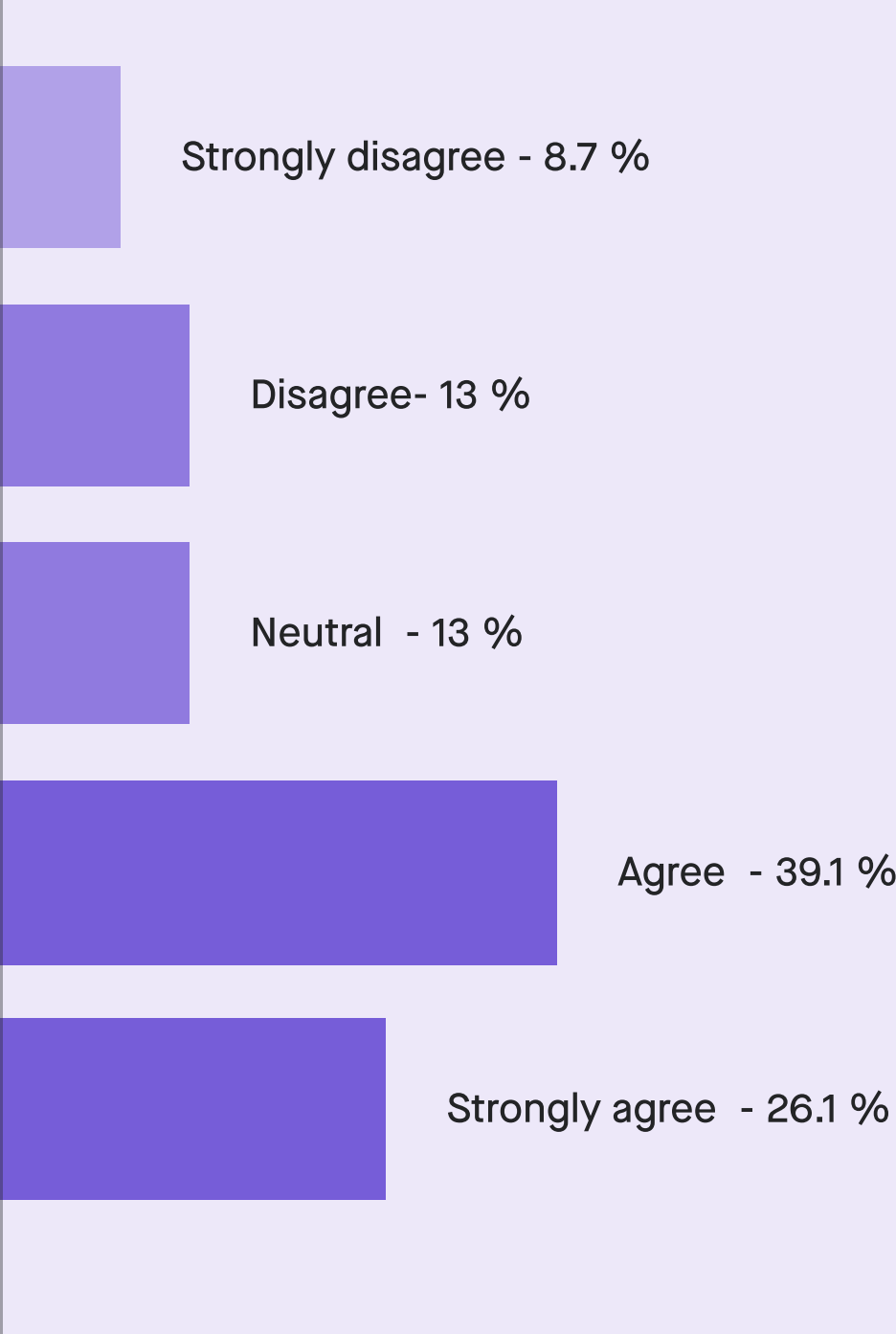


Do you think GenAI will have an impact on your CV work?



- Yes - As tech advances / Gradually
- Yes - Drastically / Quickly
- Not sure - I see some practical uses, but not at s
- No (0% of respondents chose No)

“I think GenAI will rapidly accelerate CV ML model development.”



Generative AI: Hype or not?

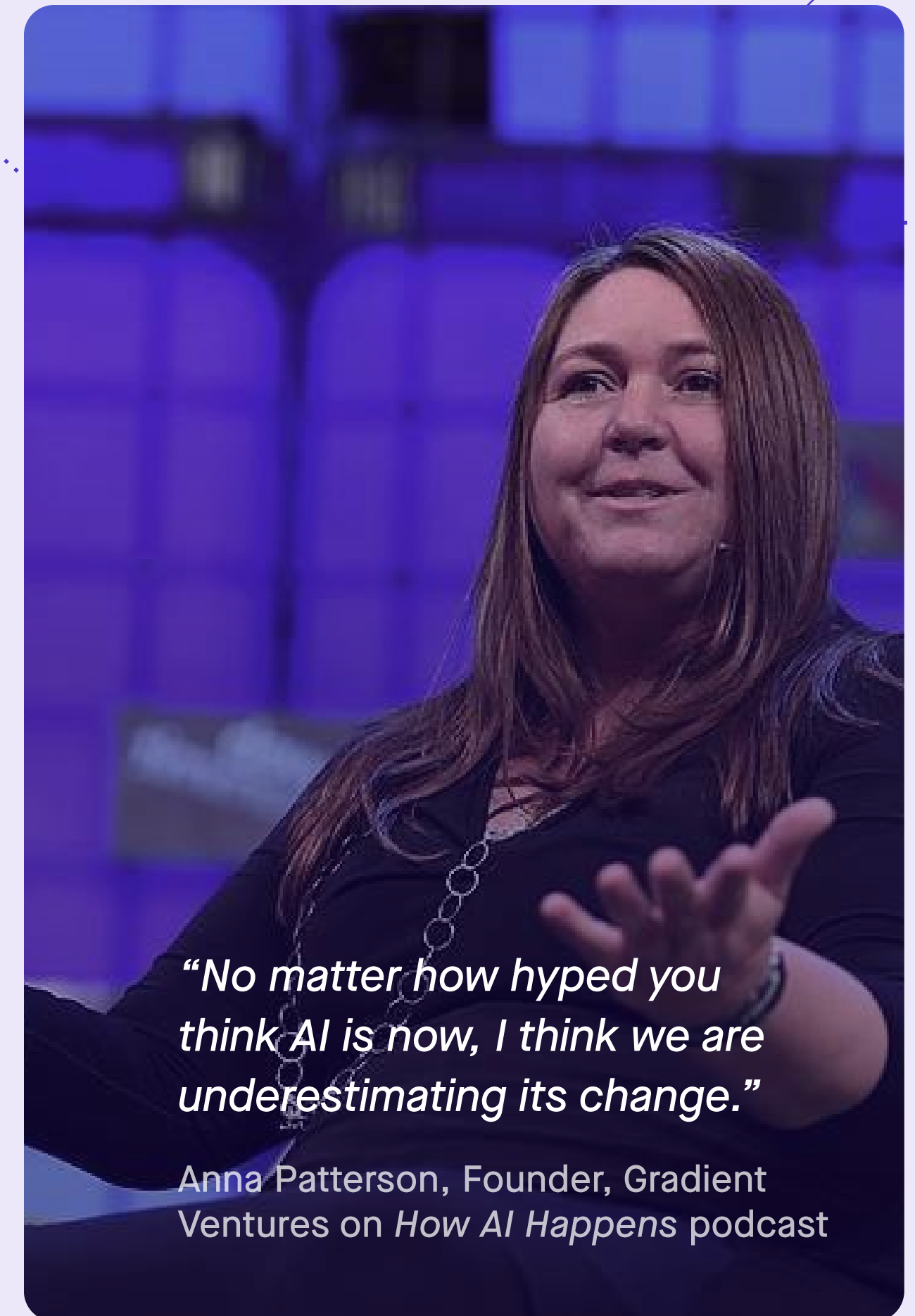
We can't predict the future, but we can say that right now GenAI is both under- and over-hyped.

The technology we use everyday, like our cell phones and the internet, have had an incredible impact on society. Yet the size of that impact was not predicted as those technologies were being developed. Clarity in the rearview mirror is always easier, of course, but there are similarities between the hype then and the hype now. In other words: don't get distracted by the hype peaks and troughs.

"Yes, [GenAI] is bleeding-edge technology, it is definitely going to radically change a lot of things in the medium to long-term. But it's not the be-all-and-end-all that it is being touted to be."

ML Lead, Services Industry, Master's Degree in Computer Science

Enthusiasm may heighten and it may wane, but the technologies continue to develop in the background. The speed at which large language models are improving, to look at one example, means they'll soon be useful much more broadly. They can already do many general tasks; now they need more training for specific tasks, and for increasingly complex tasks.



"No matter how hyped you think AI is now, I think we are underestimating its change."

Anna Patterson, Founder, Gradient Ventures on *How AI Happens* podcast

2 | Success Metrics

Confusion and Clarity

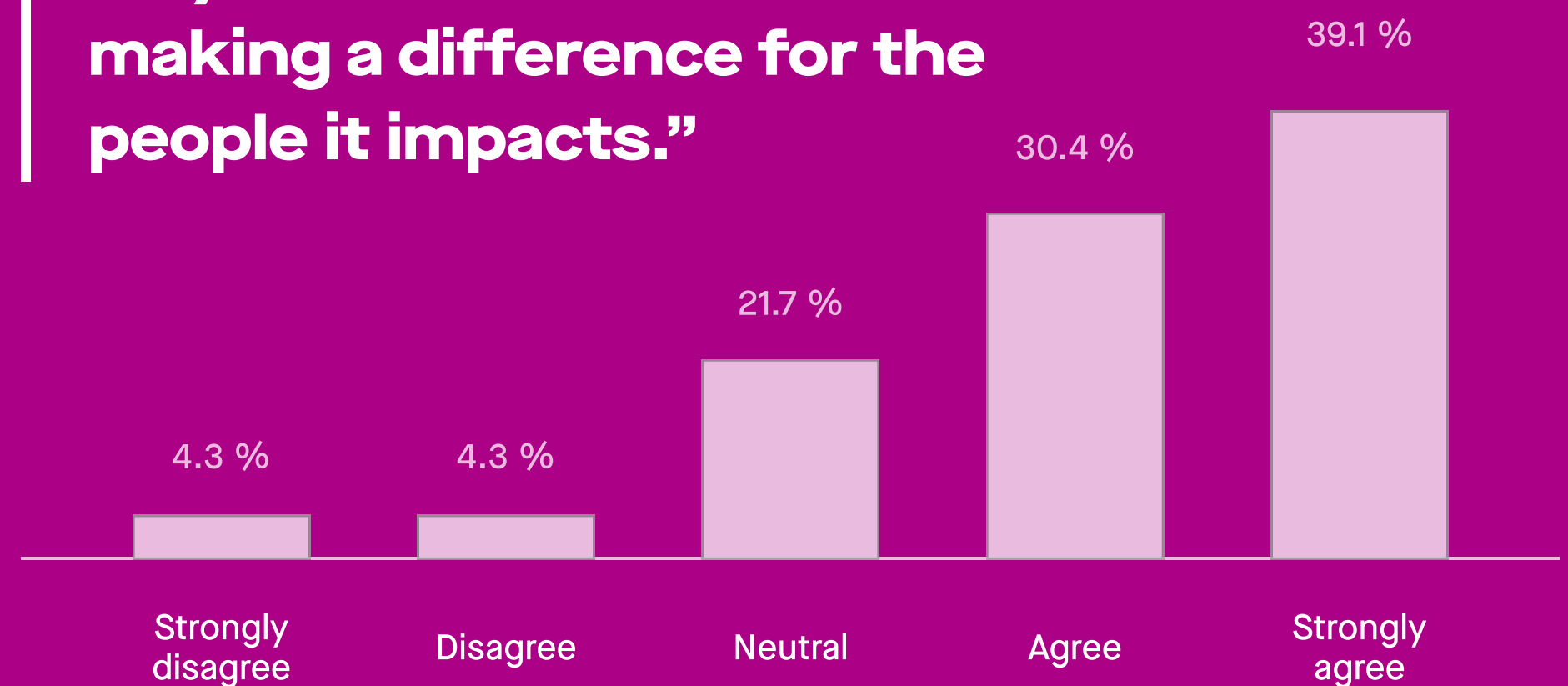
Nearly 70% of ML professionals surveyed feel their work is making a difference. It's no wonder, as some of the most visible use cases are also the most altruistic: agricultural technology that will one day feed the world more efficiently and sustainably, automotive models that enable safer driving, and medical technology that improves quality and speed of medical diagnoses.

But we're in the messy middle of how a model's ultimate success is measured.

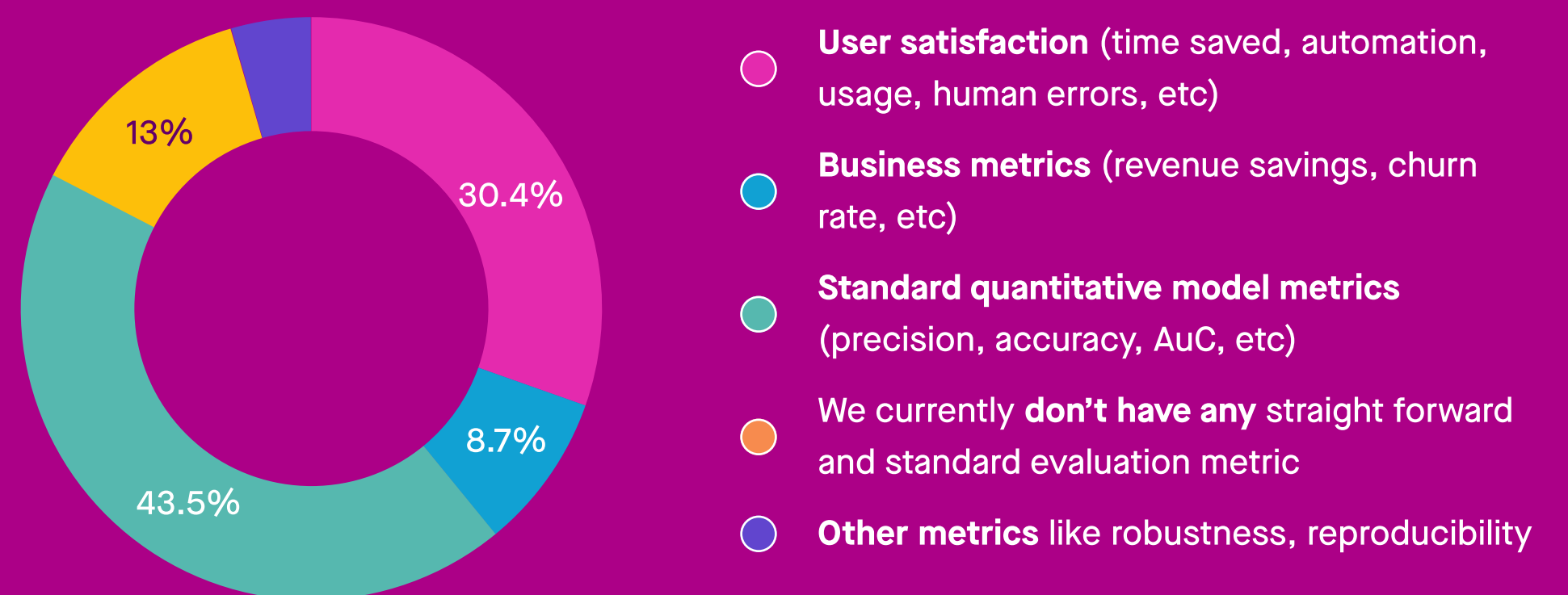
Most of our survey respondents use standard ML metrics like precision and accuracy, while 30% account for end user satisfaction including time savings and usage. Over 90% of respondents do not use business metrics as a measurement of model success. And 13% don't have consistent evaluation metrics at all.

Being committed to models that have a positive impact on the user experience should mean including user and business impact as core metrics. The field is still maturing, and hasn't yet learned how to link model development and business outcomes. Eventually, it will.

“My work in CV ML is making a difference for the people it impacts.”



Which indicators are the most used to measure success?





Model Metrics, User Satisfaction, Business Outcomes

There's an inherent disconnect between model efficacy and business value.

Model Metrics

Model metrics, such as precision or accuracy, measure how well the model was built; it's limited to how it performs in a specific environment

User Satisfaction

User outcomes and satisfaction—such as time saved, adoption of the end product, or reduction in human errors—become stronger indicators of the model's future success

Business Outcomes

Metrics such as revenue, cost savings, or customer churn rate should be layered onto a model's lifespan to provide context to the business

Many companies, particularly large ones, still have very ingrained silos. This can be a barrier to successful model development. When one group is responsible for data collection, another for model training, and a third for business value, there is a misalignment. Each group has its own metrics for success which do not necessarily translate into a holistic evaluation.

Model metrics won't disappear but will eventually be augmented by business metrics like cost savings, speed to market, and revenue increases. As more models go into production, the measurement parameters will improve to be more precise and related directly to business benefit.

Until then, teams can use model validation workflows, model monitoring, and data quality assurance as important steps to get to the point where more accurate business metrics are available.

3 | Challenges

Biggest challenges and the year ahead

For those of you who are panicking about what will happen once their models go into production, you're not alone.

While some respondents are confident their models will survive in production, most aren't. In fact, 78% say they don't have the resources they need to produce high-quality models.

- **Only 22%** say they have the tools, services, and time needed to feel confident in their model development (see next page)
- **Only 13%** are very confident their models will survive in production; 35% say this is a concern—and 9% are very concerned

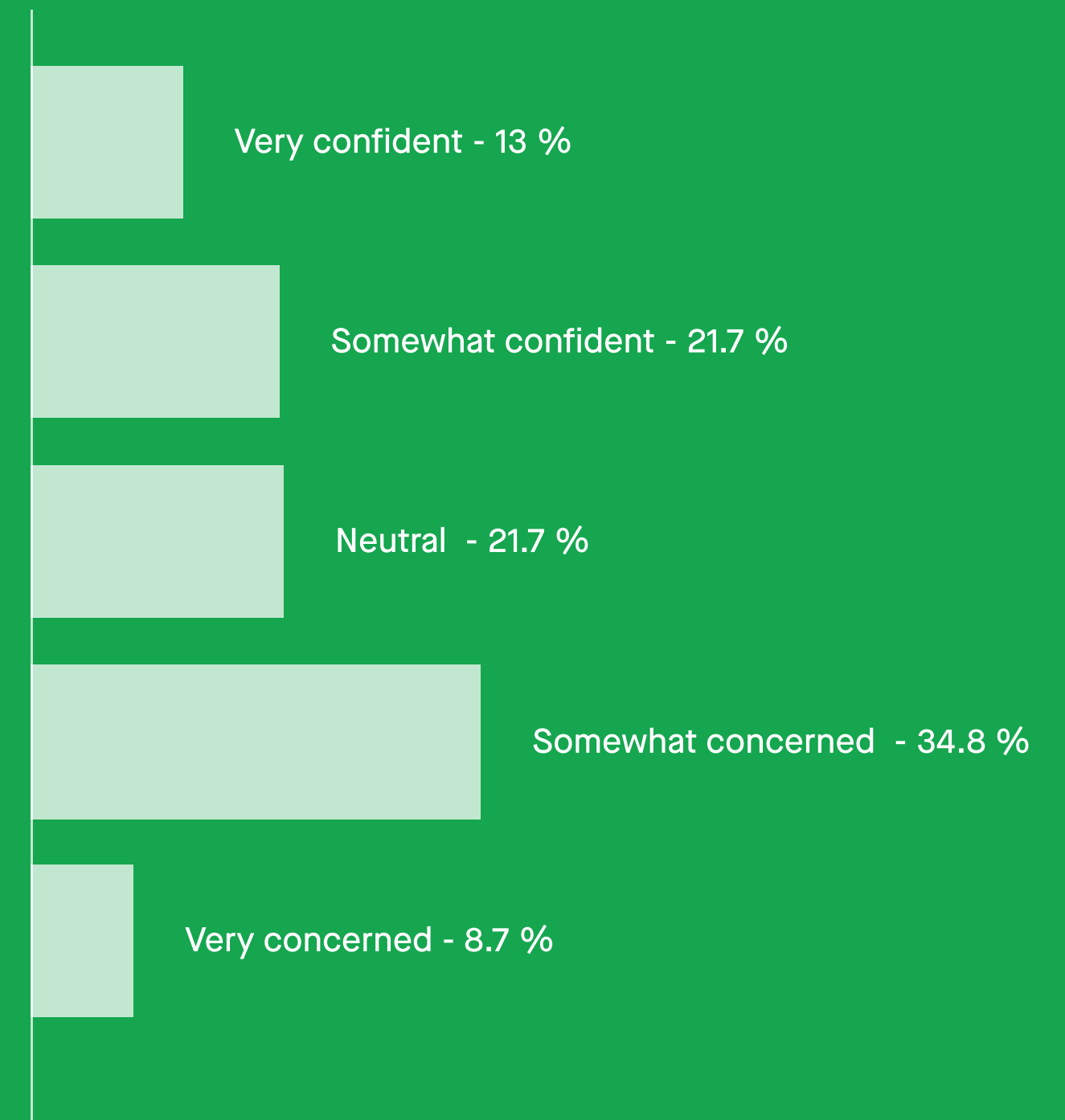
The main reasons they cite are the lack of data (20%) and the lack of time for sufficient quality assurance (40%). Their preferred solutions include better data and better data access (30%).

For training models, specificity is everything. Humans learn constantly; sometimes we don't realize we're doing it. And we don't recognize the specificity of that new knowledge, either.

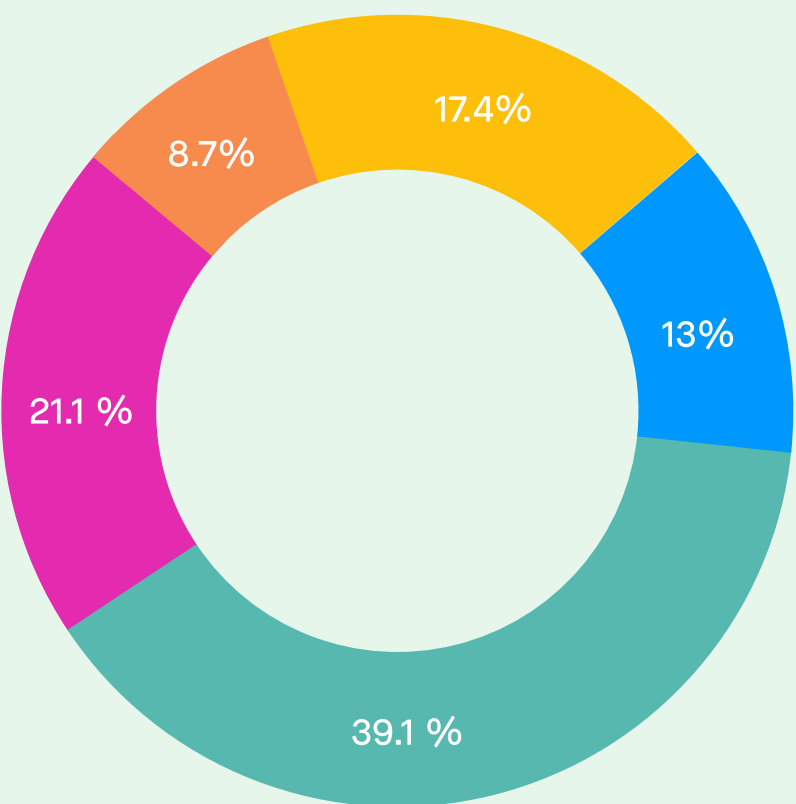
For example, within AgTech models are trained on all sorts of field conditions. But plants and weeds can look the same, or the same plants can look different depending on the location and conditions in which they are grown. Soil is different around the globe: in Australia, it's red due to high iron oxide content. Canada's northern Saskatchewan has rich, chocolate cake soil.

An AgTech AI model must be able to recognize both as "soils" and as "soils with different compositions." That level of detailed data is often missing and the resulting models inadequately quality assured to operate at that high a level.

How confident are you in your model's ability to survive in production?

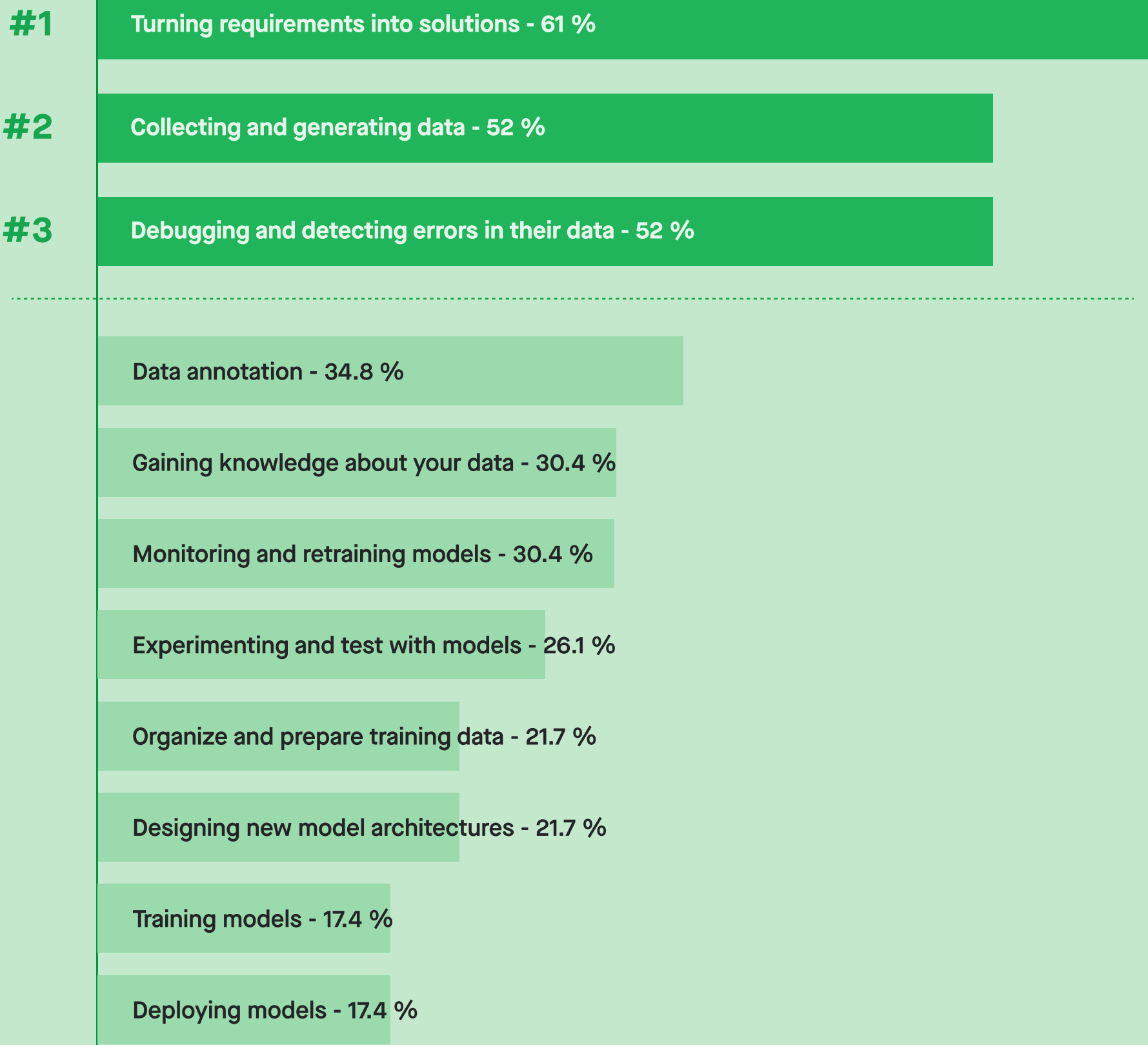


What is preventing you from feeling confident in your model development?



- Nothing. I have the tools, service, and time needed
- I don't have enough time to create or validate my models
- I can create models, but don't have enough time to QA to preferred level
- I don't have the tools or services to get my models into production
- I don't have enough data to get my models into production
- My models are not performing well in production

What are the biggest challenges for CV ML engineering for the next 12 months?



4 | Alleviating challenges

Requirements → Solutions

Turning requirements into solutions is an understandable pain point simply because the problems we're trying to solve using AI aren't simple problems. If they were, they'd have been figured out by now.

Questions relating to enough training data, achievable model performance, and the quality level of training data are all unresolved. Those answers depend on the particulars of any given use case. But an iterative approach using the right tools and subject expertise, including Human In the Loop supervision, will go a long way towards improving data quality and therefore model performance. And start simple: not every solution is complex. **Too often, inexperienced ML teams will use a sledgehammer to crack a nut.**

Data Generation & Collection

There's data, and then there's **data**, and then there's **data**. Not all data holds equivalent value; some data has more value in particular contexts than in others. Some data can greatly boost model performance, while other data can do the exact opposite.

The solution is the right data recipe—a mix of the data in the right proportions with the highest value for a use case. **Generative AI will be part of solving this issue, along with ML model tools that collect and correlate real and synthetic data.**

Error Solutions

Spot the error—how difficult can it be? Plenty difficult, actually. Debugging models requires:

1. Knowing there's an error in the first place
2. Finding the cause

The larger and more complex a model, the more difficult it will be to detect errors and debug it. And, right now at least, tools to do that aren't mature and standardized. While we wait for those tools and standards to improve, the importance of robust in-house tools, high levels of consistency, and precise documentation cannot be overstated.

Managing Total Cost of Ownership (TCO)

When we asked which resources would be most impactful in solving the challenges practitioners face, the top answer was **better data and access to data** (34%). But three categories tied for second place with 17% each:

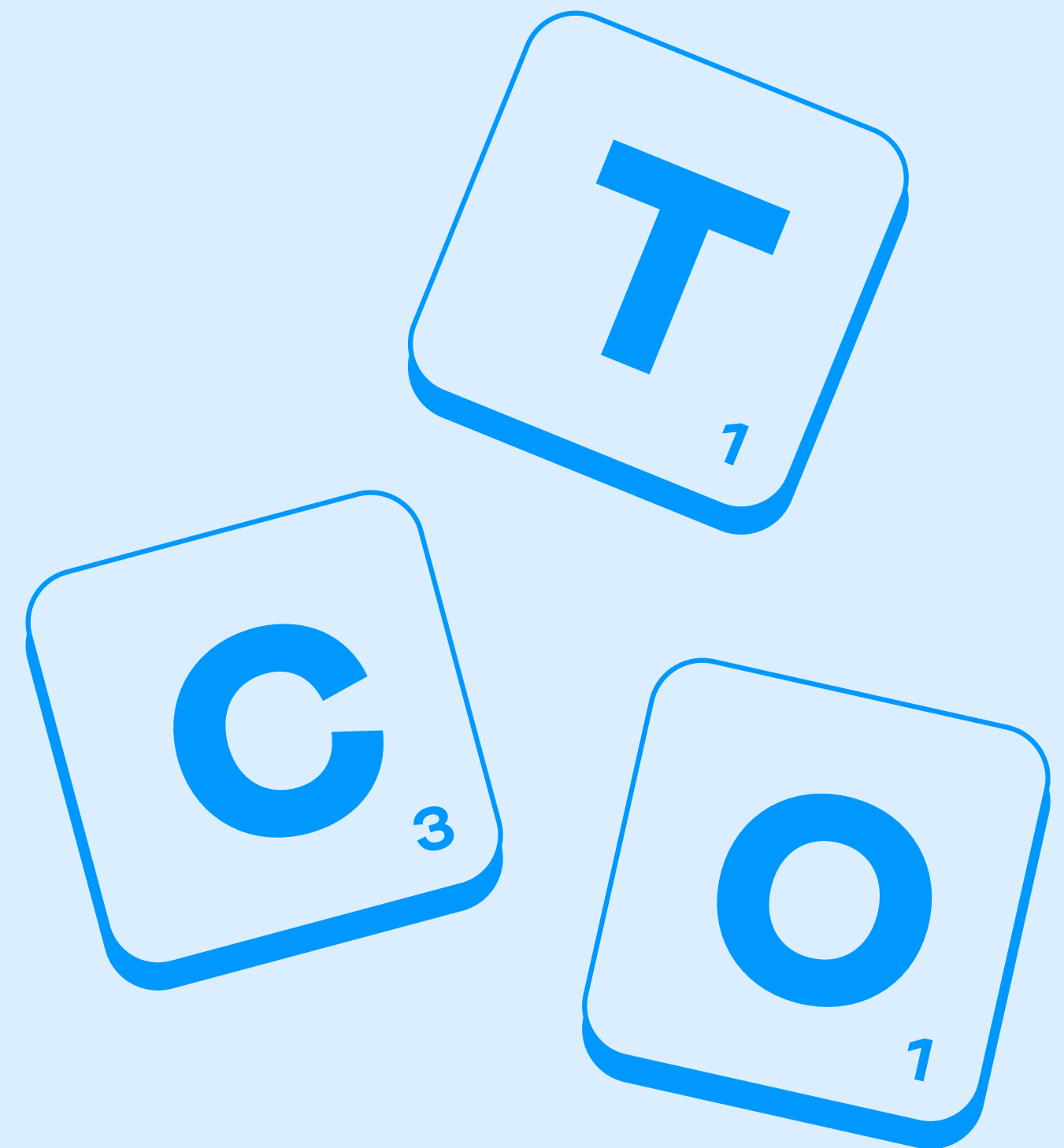
- more in-house team members
- better validation tools
- a bigger budget

The equal weight of these is fitting, as there is a strong relationship between them. At Sama we refer to this as the **total cost of ownership**.

While many say they would prefer additional internal staff to improve their models, that is often not an option. But finding a good external data curation, annotation, and validation partner presents a “cost versus quality” trade-off: an inexpensive product or service is unlikely to substantially improve quality—and may even make models more prone to errors with faulty data.

Similarly, annotation and validation tools see inverse relationships between up-front investment, effort, and efficacy. The fewer resources you spend on one, the more you’ll pay for it with the other two. Cheap, simple, or good quality; choose two.

Unsurprisingly, the temptation to spend less on staff or tools is a result of squeezed budgets or underfunded departments. Looking at the total cost of ownership can be a way to evaluate which service is right for your organization.



No matter the industry, the data struggle is real

Given the scale of most ML datasets, we inevitably see practical data overlap across industries. But each industry faces unique hurdles.

Compare an autonomous car to an autonomous tractor: four wheels, a motor, and a dozen cameras—with nearly inverse training. One should definitely stay on the road while the other operates almost entirely off it.



Retail struggles with data privacy

Retail is a growing sector for the ML industry. But there's an important balance between efficacy and data privacy. Using ML to improve the consumer experience is acceptable only as long as it's not perceived as surveillance. That limits partnering with external companies that outsource: it's too big of a risk when privacy is important.



AgTech lacks subject matter expertise

Is that a young plant, or a weed threatening to take root? One of nature's go-to moves is mimicry, adding a high level of nuance to visual recognition—a challenge for any computer vision application. Annotated datasets must have high levels of accuracy, but finding or training annotators to the same level as an agronomist takes time.



Robotics must blend machine function with human safety

Robots are being used in more and more real-world environments. To function correctly and safely for their human co-workers, their models must be trained to recognize and respond to human movements.



AV and ADAS need quality datasets at huge scale

AV and ADAS ML models run on enormous data sets, and there must be a high level of data accuracy and an ability to effectively manage edge cases. Unlike other industries, vehicle manufacturers must view model failures in production as life or death issues—because in the worst-case scenarios, it's true.



Manufacturing simply lacks data

Specific manufacturing processes also require specific data, which is often lacking. Emerging generative AI services can be used to augment training data—but must be used correctly and responsibly.

To sum up

Our respondents gave us a lot to consider: Rethinking model success metrics, finding that model confidence, and preparing for Generative AI.

Whatever you think GenAI's impact will be, double it and triple it. GenAI will touch many fields to varying degrees—**and computer vision applications are right on target.**

With some imagination and experimentation, its uses will proliferate.

Key takeaways

01

Generative AI is hyped and still maturing—but **yes, it is a big deal.**

02

Model success metrics are varied and not always linked to business outcomes; this will be an area that will develop as the ML field matures and businesses become more cohesive.

03

There's an overall lack of model confidence—many we surveyed said they have concerns that their models won't survive production and scale correctly—and the root cause is lack of resources. No high-quality quality assurance, no confidence.

04

With some imagination and experimentation, the uses of Generative AI will proliferate. But it is, at the end of the day, still a tool. Approach it as one.

About Sama

Sama is a global leader in data annotation solutions for computer vision that power AI and ML models. Our solutions minimize the risk of model failure and lower the total cost of ownership through a unique combination of ML-powered platform and SamaIQ™, actionable data insights uncovered by proprietary algorithms and a highly skilled workforce of 5,000 data experts. 25% of Fortune 50 companies, including GM, Ford, Microsoft, and Google, trust Sama to help deliver industry-leading ML models. Sama follows strict security protocols. ISO certified delivery centers, GDPR and Tisax certification protects your data and ensures compliance.

Driven by a mission to expand opportunities for underserved individuals through the digital economy, Sama is a certified B-Corp corporation and has helped lift more than 65,000 people out of poverty through training and employment. For more information, visit www.sama.com.

