\sifted/ Guides

How to get value out of data

Data can be your startup's biggest asset — if you know how to use it

How to get value out of data

Data can be your startup's biggest asset — if you know how to use it

Every company is a data company, but how do you become a data-driven company?

If you're wondering, you've come to the right place. This guide is here to lift the veil on extracting value from data, when it matters, in real time. We'll go through how to collect it, analyse it and use it effectively — especially on data-hungry projects involving AI.

Get in touch

Would you like to partner with Sifted on the next report on your sector? Contact us via **reports@sifted.eu**.

Looking for vital market intelligence? Visit **info.sifted.eu/subscribe-pro** to learn how Sifted Pro can help.

Contents

FOREWORD

The real-time revolution

In the digital age, where data is generated at an unprecedented pace, startups have a unique opportunity to build a strong foundation for future AI success. In my experience, the most successful startups are those that treat data as a strategic asset, not just a byproduct of their operations.

By prioritising high-quality, relevant data from the outset, and leveraging cutting-edge technologies for real-time processing, startups can gain a live view of their product usage and operations, enabling agile decision-making and a deeper understanding of their customers and the market.

This guide offers actionable insights on building the right data infrastructure, fostering a data-centric culture and utilising AI responsibly. But most importantly, it emphasises the transformative power of real-time data, which allows startups to track KPIs, identify emerging trends and adapt quickly to changing circumstances. By embracing the real-time revolution, startups can unlock their true potential and navigate the path to success.

Soumya Bijjal
*VP of Product,
Aiven*

Ready or not, it's time to update your data strategy

The world is drowning in data. It's estimated that 181 zettabytes will be generated globally by 2025. It's hard to picture what that means, but to give you an idea: one zettabyte alone is roughly the equivalent of streaming Netflix continuously for 36 million years.

Despite the vast quantity of data startups have at their disposal, much of it is underutilised in business decisions. In early-stage companies data can quickly become siloed, incomplete, inaccurate and obsolete, making it challenging for users to separate helpful insights from the surrounding noise.

"There's definitely more pressure to be making good use of the data that exists," says Ellen Chisa, partner at boldstart ventures.

Part of the problem is that startups, ever-constrained in terms of resources, don't invest in a robust and scalable data infrastructure upfront. "It's definitely something that a lot of startups neglect early on," says Meri Williams, CTO of Pleo. "They'll pull down a copy of a database, or pull everything into a spreadsheet, mess around with it and hope that that's good enough."

But if startups collect data intelligently from the start, the benefits can be significant, according to research by the Harvard Business Review and Google Cloud. It found the organisations that took a data-driven approach fared best over the past few years. Operational efficiency, revenue, customer loyalty and retention, and employee satisfaction all increased among data and AI leaders, when compared to other respondents.

At a basic level, being data-driven is about everyone pulling in the same direction, says Madiha Khalid, lead data engineer and founder of datavent.io. "It means each and every person in the company knows what the KPIs are and what they mean. Everyone across marketing, people, finance… they're all aware of what's going on."

With the right data infrastructure and strategy, startups can make better strategic decisions across all functions, pivot towards new opportunities and layer AI technologies onto their tech stacks with ease. In the future, it's those startups that will dominate their respective markets.
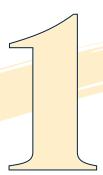
How to... collect the right data, at the right time

You can't manage what you can't measure

The foundation of a good data strategy is thinking about data from day one. Engineers should be instructed to make event tracking an integral part of the development process, collecting user feedback as soon as customers start engaging with the product.

Of course, collecting the data is only one part of the puzzle. Establishing a robust infrastructure to hold and analyse data is one of the most important investment decisions a data-centric startup will make, providing a single source of truth that spans every department.

So how is this done? We asked founders and technical leaders what decisions made the difference to them as an early-stage startup.

1

Build the right team

Data-centric startups need expertise — internally and externally — to plan and execute an effective data strategy

Having a strong data team that understands the company objectives is greater than the sum of its parts. But when to secure this expertise will depend on the company, says Francesco Tisiot, field CTO at Aiven. "As soon as you have something to measure, it's time to invest in internal or external help."

> ## "As soon as you have something to measure, it's time to invest in internal or external help."
>
> **Francesco Tisiot,** *field CTO, Aiven*

While it can be tempting to run everything in-house, working with external partners can speed up the process and keep the team focused on what matters. "Because there is a lot of open-source tooling out there, you can be tempted to run everything in-house. But most of the time, you end up burning energy and resources just keeping the lights on, while you should be focused on innovation," says Tisiot.

Danish fintech unicorn Pleo recently added a VP of data analytics, Pri Nagashima, to its team. CTO Meri Williams says the seniority of the appointment reflects the strategic importance of data to the company, and adds that while external tools are helpful — and advancing all the time — founders still need internal expertise to stitch it together. "Startups might not need a full data-engineering team but you need a minimum level of skill to use those third-party products effectively. It's a mistake a lot of startups make — not having somebody who understands the landscape of data early enough to be able to put things together in a cost-efficient, useful way."

At Miro, a Dutch innovation workspace with more than 70m users, the data team has grown significantly too. "Four years ago, we were eight people, and now we're forty," says its head of product and growth analytics, Gabor Kiss. "For early-stage startups that want to be data-led or data-informed, having data engineers in the beginning to set up your infrastructure properly is well worth the investment. I would advise having two or three people at the beginning who are technically good enough to understand the vendor landscape and choose the best ones."

That doesn't always have to be expensive, Madiha Khalid, an experienced data and AI platform engineer and advisor, says, "Junior or mid-level engineers with some experience can be sufficient to set up your initial data platform." Some more senior data engineers may also be prepared to work for equity, she adds, and working with open-source data tools and platforms such as Airflow, Airbyte, Mage.ai, Data Build Tool and Python can significantly reduce costs.

2

Weigh up build vs buy

Build features that give you a competitive advantage but keep flexibility in mind when choosing tools

---

There is a plethora of tools and platforms available to help with data collection and analysis. The most suitable will depend on the startup in question but founders should choose cloud-based solutions that offer scalability and flexibility for the future. "I'm always mindful of the commercial strategy before I make a technical decision," Philipp Lorenz, CTO at Basecamp Research, says.

The biotechnology company is using AI to find and commercialise the world's unknown and undiscovered protein sequences at an unprecedented rate. "For me the two considerations are around IP and speed. Our engineers may want to build a solution from scratch but startups like us need to be very fast. If there's something off the shelf that works, I'm always in favour of integrating something like that."

Ellen Chisa, partner at boldstart ventures, agrees engineering teams should focus on building features that will become a competitive advantage. "The build vs buy question comes up all the time but you don't want to build all of your own data infrastructure if that's not the core of your company. You want to spend time on things that are going to help you iterate with your early customers and design partners."

Open-source software is growing in popularity as organisations look to save costs. OpenLogic's 2023

State of Open Source Survey found nearly 77% of businesses had increased their use of OSS in the previous year. Beyond cost saving, benefits include faster development speed and interoperability.

Having an expert community at your disposal has also been great for Miro, Kiss says: "They can develop and improve tools really fast. The tools we use to do analysis are most often open source and that has been a huge benefit for communal learning." Kiss has also been careful to avoid vendor lock-in, computing most of its metrics in the data warehouse so they can swap tools out. "If you do your metric definition in those tools, then migrating is really painful," he adds.

> **"**
> ## You don't want to build all of your own data infrastructure if that's not the core of your company."
>
> **Ellen Chisa,** *partner, boldstart ventures*

That's also been a priority for French digital health insurance unicorn Alan. Priorities are likely to change as a startup scales. "In the early days, your goal is to find something that's easy to change, rather than something perfect," says Charles Gorintin, cofounder and CTO of Alan. "Find something that's good enough so you can move forward. We decided to go with a business intelligence tool called Metabase in 2016. It's fully open source and we're still using it. We moved from Amazon RDS to Snowflake and we didn't have to change anything except for some queries. It's great to have that flexibility."

3

Decide which metrics to track

Data storage is cheap. So keep your options open when finding a product-market fit

There are two approaches when deciding which metrics to track — take everything and work out what's helpful later, or optimise early by deciding on key indicators for long-term success.

Luke Talbot joined Flip, a German employee app empowering deskless work, earlier this year as CPTO. He says data is always an ongoing battle for any growing startup. "Data is just one of those things where, if you don't have a structure to it, you can drive yourself crazy because you could measure everything." He's a fan of Google's HEART framework, which stands for happiness, engagement, adoption, retention and task success. "It's a really good model for measuring the success of software platforms because none of the parts really work without the others," he says.

> ## You can drive yourself crazy because you could measure everything."
>
> **Luke Talbot,** *CPTO, Flip*

Others say there's no need to over-engineer metrics at first, particularly given the falling cost of data storage. Gorintin, who has previously worked as a data scientist for Facebook, Instagram and Twitter, says it's important to begin capturing information, even if you don't yet know what you'll do with it: "From the get-go, I wanted

to build a really solid data infrastructure and collect data even though we didn't know what was going to be useful. Having some data early on can save you a lot of trouble later. If you start measuring or logging too late, you lose a lot of history."

Chisa agrees: "Don't optimise too much," she says. "I often see too much AB testing. Early-stage startups usually don't have enough users for that to be helpful. Just try to think about the first few questions you're trying to answer." Look beyond vanity metrics, she adds, such as page views, downloads of a freemium version of an app, or number of social media followers.

Technical leaders also need to make sure they're keeping on top of that data as it balloons, Basecamp Research's Lorenz says. "We have extensive monitoring of our infrastructure and the database as it grows. One thing we did very early was be very comprehensive with our data annotation, building huge pipelines that are very scalable."

The number one reason organisations use open source software is access to the latest tech

Access to innovations and the latest technologies **38%**

Stable technology with community long-term support **37%**

Fast moving/constant enhancements, releases and patches **32%**

Community-orientated and transparent **26%**

No license cost, overall cost reduction **26%**

Open standards and interoperability **24%**

Makes it easier to hire or retain employees **11%**

Source: *The 2023 State of Open Source Report*

4

Don't forget about privacy

Organisations collecting data need to put guidelines in place to ensure they're compliant with GDPR

Startups collecting data must remember the legal responsibilities they have to safeguard that information. Amy Pierechod, partner at Gordons law firm, says it can quite often be overlooked by founding teams. "They want to show fast growth and have probably signed up a lot of customers, maybe on a free basis," she says. "Suddenly they find themselves holding a lot of data and don't have the structure in place to manage it safely."

Proper governance ensures that a startup's data is kept secure and compliant with legislation such as the General Data Protection Regulation (GDPR), and minimises the risk of large fines or reputational damage. Founders should get structures in place early to handle this, says Lauren Wills-Dixon, head of privacy at Gordons. That might include conducting data-protection impact assessments, training staff, obtaining consent and minimising the amount of personal data they collect. "One of the requirements under GDPR is that businesses have a privacy by design and by default approach," Wills-Dixon says. Companies that get this wrong could be fined up to €20m, and suffer wider reputational damage. It could even derail a funding deal with investors.

At UK-based parcel courier Packfleet, which was founded by three early Monzo employees, founding engineer Robin Bilgil says privacy was an early priority. "It can be quite difficult to manage later on if you don't put a framework in place early enough. We make sure we're using trusted providers that are based in Europe, and we have access controls and audit logs to limit the people who have access to different sets of data. People who work in our warehouse need access to different data to those in customer service." Data minimisation has had benefits for cost as well as privacy, he adds. "We had to limit the scope a bit and say we don't actually need that, or the amount of value it adds is so small."

> "
> **One of the requirements under GDPR is that businesses have a privacy by design and by default approach."**
>
> **Lauren Wills-Dixon,** *head of privacy, Gordons*

Having a comprehensive deletion policy is also important for when users request they're deleted from the system, says Talbot from Flip. "You need to have a really good concept in place, to run through everything and eradicate a user from the system. That's something a lot of people don't think through properly."

How to... analyse real-time data

Once you have your data, what can you do with it?

The goal of data collection, of course, is to put that data to work. But too many startups don't make the most of the insights they have. One survey by Deloitte found while 49% of businesses are using data analytics to drive process and cost efficiency, only 16% are using it to drive strategy and change.

Without the right analytics, data can be overwhelming. If startups want to make better decisions based on real-time evidence, they will need a robust pipeline that can transform the raw, messy data into actionable insights. But how is this actually done? We asked founders to weigh in with advice based on their experiences.

1

Build trust in the data

Creating one source of truth creates better insights

As startups experience rapid growth, more data can lead to complexities such as data volume, inconsistencies and accessibility. When Gabor Kiss joined Miro as head of analytics in 2020, it was a time of rapid scaling for the business, which led to real problems for the data team. "All of our metrics went through the roof but our data infrastructure was not ready for the scale. We had hundreds of dashboards and so many parallel realities of how our business was performing. We also didn't have good observability or experimentation velocity. That hindered our growth."

> ❝
>
> ## All of our metrics went through the roof but our data infrastructure was not ready for the scale."
>
> **Gabor Kiss,** *head of product and growth analytics, Miro*

Today, Miro's analytics team focuses on building reliable and consistent data foundations. It has spent time understanding how each department's metrics relate to the other, and aligned them to the overall business objectives as set by the executive team. "We spent time deconstructing the business, understanding the growth loops and prioritising those core metrics with our executive team. We decided on a set of metrics the business really cares about and have migrated everyone to using those metrics. That's helped us focus on the right things."

It's a common problem, agrees Francesco Tisiot from Aiven. "You always need to balance the innovation and span of data usage across companies, alongside having a good way of defining metrics and KPIs. Evaluate how much the same numbers are being crunched in different ways and create pipelines that allow the company to align."

Startups have to harmonise metrics so that everyone knows what they mean. Often harmonising metrics is taken for granted, explains Pri Nagashima, VP of data analytics at Pleo. "If it's a product business, you might have a metric like lifetime value or churn. If it's not widely agreed how that's calculated, it can become a moving target. As the business grows and you have more data and more ways of analysing it, that's going to cause a foundational problem for data quality and underlying governance and compliance. Take the business metrics you care most about and make sure everyone really understands what they mean."

2

Guard against complexities that come with scale

Keep things as simple as possible and build in capacity when analysing data in real time

Thanks to cloud technology, data storage has reduced significantly in price. But it can still be expensive to access and compute the data in real time. "That's where a good data scientist and engineer comes in," says Flip's Luke Talbot. "The most common problem is your analytics will be connected directly to your database, which will become hugely slow, problematic from an architecture point of view and really limiting. A dashboard will take 20 seconds to load, which becomes completely unmanageable."

" As you add more features, you need to make sure they're structured properly."

Luke Talbot, CPTO, Flip

There are steps you can take to limit the impact, he adds, by ensuring data pipelines are nice and clean and only capturing data that you need. But the biggest difference comes from designing the data platform properly. "I've seen platforms that have hundreds of millions of data points and the analytics is blistering

fast because it's been well designed. It's something you need to keep on top of. As you add more features, you need to make sure they're structured properly."

Techniques such as parallel processing, in-memory computing and stream-processing can also help. "Trying to do operational workload and analytics workload on top of the same technology will hit performance," Tisiot says. "You will never have a technology which is capable of doing both perfectly at the same time. That's where streaming technologies can help decouple where you analyse the data from where you store the data. It's a scalable way of having the best of both worlds."

Keeping things simple as much as possible also pays off, Kiss says. "It really helps to swap out metrics where they no longer matter. And do you really need 150 events, for example, to determine whether or not someone is an active user?"

Who oversees analytics initiatives?

Business unit or division head **23%**

CFO **18%**

CIO **15%**

CEO **9%**

CMO **5%**

Other C-suite exec **5%**

Chief analytics officer **4%**

COO **0%**

No single exec **20%**

Source: Deloitte

3

Use data to spot opportunities

Data should be central to how leaders and departments collaborate and grow the business

---

Once data pipelines are established and routine reporting is in place, founders can better use analytics to identify new market opportunities, emerging customer trends and other areas for growth. With proper feedback loops, data can help founders refine a customer profile, design better experiments and estimate the appeal of a new feature.

"

Focus on the customer. How can you use data to solve pain points for them?"

Pri Nagashima, *VP of data analytics and AI, Pleo*

At Packfleet, data insights are helping the team optimise its operations. "We're focused on becoming profitable on a per package level," Bilgil says. "This year we went back to the fundamentals and questioned everything we'd set up two years before. We made changes around driver pay and performance incentives, and shifted how we're making predictions as well." Like Alan, Packfleet uses Metabase, a business intelligence tool that the whole company has access to. "It's a no-code solution that helps people answer their own questions about the business. They're coming up with solutions to problems we may not have even spotted otherwise," Bilgil adds.

When trying to identify new opportunities for growth, Nagashima recommends cross-departmental working. "I am an advocate for experimentation and seeing what value it brings. Where it works really well is when you have data people working closely with designers, product and engineering people. Make sure you've got data as a core part of the development cycle. And focus on the customer. How can you use data to solve pain points for them?"

Having insufficient data is a problem, but too much data can lead to analysis paralysis, where it's difficult to understand what data you should be looking at, or what questions to ask. Clarifying your goal, taking incremental steps forward and setting a deadline can all help build momentum again.

4

Invest in data literacy

Employees need training to make the most out of the data they have access to

Data shouldn't be a separate part of the business but embedded across all departments and workflows. But that's not always straightforward. Research by NewVantage Partners found cultural challenges are the biggest obstacle for organisations aiming to become data-driven. Organisations need to democratise access to insights as much as possible and leaders need to lead by example, being transparent about the data they're using to back up their proposals and plans, and investing in data-literacy training for employees.

> ❝
> ## A lot of places aspire to have more of a self-service data culture, but with that you have to teach people how to interpret data"
>
> **Meri Williams,** *CTO, Pleo*

"The lack of general data and mathematical literacy is a real problem," says Meri Williams, CTO of Pleo. "It's very easy for people to misinterpret data. A lot of places aspire to have more of a self-service data culture, but with that you have to teach people how to interpret data, and about things like statistical significance. It's very tough even getting basic data-privacy training in some companies. So helping people really understand this is a challenge." Having good dashboard design can help, she adds, which also prevents your expensive data scientists spending time purely pulling data for other teams in the business.

"We are on that journey," Kiss at Miro says. "We have so many product teams that naturally ask questions about their specific features, so we spend a lot of time on core data models which explain a lot of our feature usage and business performance. Rather than allowing every event to be used for all kinds of metrics, we expose a smaller subset to the business and try to make sure it's understandable."

Packfleet has recently run some Lunch and Learn sessions around data knowledge, which Bilgil says have helped employees understand how to interpret dashboards better. "I think everyone can upskill themselves even without mathematical or statistical knowledge," he adds. "The first thing to be aware of is that data can be misleading, and you need a certain amount of it before you can conclusively draw conclusions. People are also often quite right in their intuitions, so if the data is saying something completely opposite to what you believe, the chances are you've misinterpreted it. You've still got to listen to your gut."
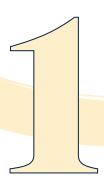
How to... be data-driven in an AI era

The infinite potential of AI starts with good data foundations

It is a big investment for a startup to get their data foundations right, but those who do are best placed to take advantage of new AI technologies. Those that don't will find their problems are magnified, especially as the speed and volume of AI innovation intensifies.

Once a startup is AI-ready, it will be able to plug-and-play new tools with a raft of benefits. "When implemented carefully, AI has the power to transform business performance while prioritising safety, governance and human oversight," says Marc Warner, cofounder and CEO of Faculty, which helps organisations make better decisions using human-led AI. "There's a sensible path to reaping benefits while managing risks, but we will only get there through smart choices."

1

Make sure the data is clean

As the old adage goes, garbage in, garbage out. AI tools will only ever be as good as the data you feed it

Before layering AI on top of a data platform, founders should be sure the quality is high. Inaccurate or incomplete data can lead to flawed analysis and misguided decision-making. "Do you know where the data is coming from and how reliable it is?" asks Philipp Lorenz, CTO of Basecamp Research. "Is it data you generate yourself or are you extracting it from somewhere else? Can you be sure this is good data? If you're using data for deep-learning applications, it has to be representative. There are lots of data resources that are very biased and very small, which won't be as effective."

Meri Wiliams, CTO of Pleo, agrees: "You need to know your base data is clean, well formed and well understood." Be careful about the sensitive information you're loading into large language models (LLMs), or take advantage of the enterprise options which most of the big LLMs have now added. Organisations want to upload sensitive business information without worrying they're going to surface in a competitor's query. "Nobody wants their data included in training models," she adds.

AI isn't plug and play, says Aiven's Francesco Tisiot. That misconception is why a lot of organisations are struggling to realise the benefits of this technology: "ChatGPT and other large language models create the fake illusion that AI is simple. You just need to open your Pandora box of data assets and AI will magically pick them up. It's not like that," he adds. "AI is not different from any analytical workload, so we cannot expect it to be smarter or dumber than any other. We have to go back to the simple rules of cleaning data, removing any personally identifiable information (PII) and preparing our data for AI usage."

> ## "Nobody wants their data included in training models."
>
> *Meri Williams,* CTO, Pleo

It may be necessary for technical leads to build an interim step for an LLM to access, rather than giving it freedom to roam through the entire data lake. That's what Kiss has done at Miro, which has also improved the answers it's able to give: "It would not be cost effective or computationally effective to let it into the entire database," he says. "We have built an interim step that parses all of our data and puts it into a format which is easy for the LLM to consume. That makes it more reliable."

2

Maintain transparency and human oversight

Leaders should not layer AI without a thorough understanding of its data inputs and connections

A startup's AI readiness needs to be built from the ground up, and integrate with its existing data infrastructure. "Today, most organisations deploy AI models in silos. Standalone AI tools are not only ineffective but also potentially dangerous," Warner says. "Organisations can't be sure what AI tools are doing and whether the results are rife with bias, misinformation, or are vulnerable to security breaches. Connected AI, on the other hand, means not building models in isolation and ensuring central governance across your systems. Its goals should be set by humans and output monitored by them too."

Pleo is using AI within its applications but only to provide suggestions rather than entirely automate the user experience. "We have a lot of AI that suggests the right category for an expense, for example, or the date on a receipt, but it doesn't make the decision for you. We want to make efficiency gains for users without reducing accuracy," Williams says.

There are ethical considerations to be made too. Lorenz says Basecamp Research has put great emphasis on traceability of its data, obtaining consent from the National Parks they work with, and sharing a portion of the profits with them. "In the age of generative AI, a lot of biodiversity-rich countries feel screwed over by public databases. We've prioritised engaging with stakeholders from where the data originates and getting consent from them, which is already a revolution in this space. We can trace every single data point back to the landowner. So far, we believe we're the only life science company thinking about consent for training data at that kind of scale."

> ## Standalone AI tools are not only ineffective but also potentially dangerous."
>
> **Marc Warner,** *cofounder and CEO, Faculty*

That transparency is likely to become more important as legislation catches up with AI, so it's best to get it right now, says Lauren Wills-Dixon, head of privacy at Gordons. "We're yet to see a lot of regulatory guidance on it, but it's coming. You could have a model ingesting huge amounts of personal data, but you need to be able to tell an individual where their data has gone. If you're an organisation processing personal data using AI, you would still need to meet the transparency principles under the GDPR."

3

Use AI tools to solve genuine problems

The goal is not to introduce AI for its own sake but to provide value to users

The proliferation of AI tools on the market has made it easy for anyone to test the technology without investing significant time and money. Startups should begin with the simplest AI model that can solve your use case, especially if it's off the shelf. "The nice thing about GenAI is it's suddenly made this technology very accessible for anyone and everyone," says Pleo's VP of data analytics, Pri Nagashima. "Encourage people to try things out, see what's possible, mock something up and put it in front of customers to see if something has value. Having that kind of design thinking is really key."

> ❝
> ## AI solutions should be designed to serve business outcomes.❞
>
> **Marc Warner,** *cofounder and CEO, Faculty*

The team at Moneyhub uses a system called Credal AI, which allows it to layer on any of the LLMs. "Our data is ring fenced and we're in a perfectly secure environment to try it and see how it works," says

its CEO, Samantha Seaton. "The developers are loving it because it's helping them to code and debug. Our CTO thinks it's improved productivity by around 30%."

The pace around the development of AI is so fast that it pays to not get locked in with one tool or another, Aiven's Tisiot explains. "Today the top leader is OpenAI and Microsoft. Tomorrow it might be Amazon or Google. Being able to port your data and not rely on tooling by a single provider helps a lot. It allows you to experiment, and port the data to where it's needed in a safe, scalable and cost-effective way.

And while experimentation is useful, founders will get the most out of AI if they focus on solving specific problems, Faculty's Marc Warner says. "AI solutions should be designed to serve business outcomes. Leaders must consider the critical decisions that AI can help with and devise a plan with this at its core. These narrow applications are safer because they have predetermined goals. They also tend to be more successful as they're laser-focused on solving a set problem — not something loose or undefined."

Marketers are quickly adopting AI

70%
of marketers used AI to help streamline data analysis last year

44%
of marketers say AI is very effective at performing data analysis

Source: 2024 AI Trends for Marketers

4

Put safeguards in place

As with privacy, startup founders should remain up to date with legislation around AI

On May 21 2024, the Council of the European Union approved the EU Artificial Intelligence Act. The world's first standalone law governing the use of AI imposes new obligations on companies developing and deploying AI systems. Businesses need to assess whether AI systems are considered high or limited risk, develop an AI-governance programme, and maintain transparency and human oversight, among other responsibilities. Wills-Dixon from Gordons says it's something they're advising a lot of business owners about at the moment. "Everyone wants to use ChatGPT but while chatbots on websites, for example, can be a really helpful tool for something like customer service, businesses need to be aware the AI model could be ingesting a load of personal data."

Security needs to be a high priority and startups should establish an AI framework to provide teams with guidelines. That should give them confidence to experiment within set parameters from the start.

But there are moral questions too, Nagashima says. "I think that needs to be taken a little more seriously. If you're processing personal data, you also need to be thinking about the implications to a human being and treating that data with respect."

AI models can quickly become biased, adds Pleo's Meri Williams, particularly if companies aren't cleaning the training data used. "Famously in the US, they trained some models on datasets from the past 50 years of banking decisions. They included some very racist decisions about not giving people of colour mortgages and things like that, which unsurprisingly created a model that was just as racist. We've got to be careful about the potential for money laundering for bias."

> **"**
>
> **Businesses need to be aware the AI model could be ingesting a load of personal data."**
>
> **Lauren Wills-Dixon,** *head of privacy, Gordons*

As for the changes required off the back of new AI legislation, Charles Gorintin, cofounder and CTO of Alan, is pragmatic. "I will take the same approach I had when the GDPR was launched. Let's wait and see and learn from others. At the moment the regulation has a lot of uncertainty in it. We trust we'll be able to move faster than most of the other players."

How to... build a data-centric company

Ten of our experts share advice for entrepreneurs looking to make data the focal point

"Choose three metrics that are the right fit for your company, and that you want everyone to understand. Galvanise the entire company around them. Most importantly, do not overcomplicate them. A metric which is understandable and directionally right is much better than something that needs a 10-minute video explainer. The more you reference a metric, the more top of mind it will be and the easier it is for people to be data-informed."

Gabor Kiss, *head of product and growth analytics, Miro*

"There's a lot of people who overdo it in the beginning because they've read what Google or Meta does. They imagine problems they don't have yet. As long as you're doing a good enough job of capturing raw data so you can transform later, that's enough. Get your foundations right, get them clean and you can build anything on top."

Meri Williams, *CTO, Pleo*

"Be intentional and make sure you know what question you're trying to answer when you're looking at the data. Have a hypothesis around what you're trying to test — whether it's changes you're making to the website or the user flow — and check the data to see the impact it had. Most casual business users are not strong enough data analysts to decipher correlations that have real statistical backing to them. You need to know what you're looking for."

Ellen Chisa, *partner, boldstart ventures*

"Think about technologies that give you options in the long run by having eyes on the current need and the long term. If you're working with a single solution by a niche vendor and they raise their prices, you'll have to pay them. On the other hand, you may be able to leverage that same technology in different places and mitigate that risk."

Francesco Tisiot, *field CTO, Aiven*

"Set up the pipelines that you need as early as possible. It shouldn't be an afterthought. Data should be collected almost automatically, with the relevant safeguards in place. The other big thing that's helped us is improving transparency, and giving users access to answer their own questions."

Robin Bilgil, *founding engineer, Packfleet*

"We're a big advocate of not building what we don't need to build for two reasons. One, because it distracts you. But secondly, we want to stay quite lean as a business ourselves. We want the people we hire to spend time doing stuff that no one else can do, rather than peripheral work."

Samantha Seaton, *CEO, Moneyhub*

"Big datasets are an amazing asset to have but prioritisation can be a big challenge. If you have a large context dataset that lends itself to many tasks and applications, the question really becomes about combining your technical abilities with your product and your commercial strategy. Experiment early and learn lessons fast. You don't want to invest a crazy amount of resources only to find out down the line there's no product-market fit and you should've built something completely different on top of your database. As a startup, the speed at which you learn in the early days impacts your potential success later on."

Philipp Lorenz, *CTO, Basecamp Research*

"I definitely would encourage anyone who is looking to build a business around data to explore beyond GenAI. There's a lot out there around supervised and unsupervised machine learning, computer vision, hybrid AI, using knowledge graphs of deep learning etc, that can solve problems and in a more reliable way. You might find that a very simple statistical model will solve the problem in a faster, cheaper way."

Pri Nagashima, *VP of data analytics and AI, Pleo*

"Some people are scared of analytics tools but it's much easier for everyone to be a data scientist now. If I were to do it again, I would train everyone to be able to use an AI that can write SQL so they can get answers to queries more easily. It wasn't possible in the past and all of the requests would go to the same people. But today, you can get a lot of things done with GenAI."

Charles Gorintin, *cofounder and CTO, Alan*

Acknowledgements

ROBIN BILGIL
Founding engineer,
Packfleet

PHILIPP LORENZ
CTO,
Basecamp Research

LUKE TALBOT
CPTO,
Flip

ELLEN CHISA
Partner,
boldstart ventures

PRI NAGASHIMA
VP of data analytics and AI,
Pleo

FRANCESCO TISIOT
Field CTO,
Aiven

CHARLES GORINTIN
Cofounder and CTO,
Alan

AMY PIERECHOD
Partner,
Gordons

MARC WARNER
Cofounder and CEO,
Faculty

MADIHA KHALID
Lead data engineer and founder,
datavent.io

SAMANTHA SEATON
CEO,
Moneyhub

MERI WILLIAMS
CTO,
Pleo

GABOR KISS
Head of product and
growth analytics, Miro

LAUREN WILLS-DIXON
Head of privacy,
Gordons

Aiven is a global data and AI platform company, helping organisations gain more value from their data. Aiven's cloud platform combines open choice services to stream, store, and serve data, simply, securely, and rapidly across major cloud providers. Aiven is trusted by thousands of customers, from start-up to enterprise, to create next-generation applications incorporating AI/ML confidently and quickly.

Through their Cluster Startup Program, Aiven supports early stage startups with up to $100,000 in credits that can be spent on services available on the Aiven platform over a year.

Headquartered in Helsinki with offices globally, Aiven is backed by private and institutional investors.

You can visit us at aiven.io and follow us on LinkedIn and X at @aiven_io.

Get in touch

Would you like to partner with Sifted on the next report on your sector? Contact us via **reports@sifted.eu**.

Looking for vital market intelligence? Visit **info.sifted.eu/subscribe-pro** to learn how Sifted Pro can help.