Alistair Clarke | Writing Samples

Three long-form pieces covering technical topics:

- [What Is AI Infrastructure and How Can You Build Yours?](#)
  - This guide explains the process of developing AI infrastructure, covering hardware requirements, data sovereignty, and legal compliance.

- [What Is Descriptive Analytics? Definition, Benefits, and Examples](#)
  - This guide defines analytics methodologies, helping readers understand core technical details and business implementations.

- [What Is Fine-Tuning? The Comprehensive Guide To Fine-Tuning LLMs in 2025](#)
  - This guide sets out the process of fine-tuning LLMs, considering the impact of different training approaches and how users can choose among them.

What Is AI Infrastructure and How Can You Build Yours?

Every enterprise-level business with any ambition is exploring ways to practically leverage artificial intelligence. The first wave of consumer-grade AI tools may have seemed gimmicky, but developments since then have made it abundantly clear that the technology has countless serious applications — and those who want to marshall those applications need AI infrastructure to get them there.

A business that invests in strong AI infrastructure is likely to end up in an excellent position to seize new opportunities and tackle emerging challenges. Conversely, proceeding without such a foundation is a recipe for encountering obstructive performance bottlenecks and long-term scalability limitations.

If you intend to build your business into an AI powerhouse, then, you need to take action. And if you're going to do that, you must first understand what exactly AI infrastructure entails, then form a solid idea of how you can create a framework to not only implement AI effectively but also sustain and expand it as needed.

In this article, we'll explore the concept of AI infrastructure, explain how it differs from traditional IT systems, and outline the key benefits a robust AI foundation can offer. We'll also detail the essential components of effective AI infrastructure and provide a clear step-by-step guide to building a setup that aligns with your business objectives, compliance requirements, and budget constraints.

What is AI infrastructure?

AI infrastructure encompasses all hardware, software and related resources committed to the cause of developing and using artificial intelligence. To serve that demanding cause, it requires high-performance computing, exceptional scalability, massive amounts of data, and careful oversight from skilled experts.

At the core of every AI infrastructure configuration lies processing devices capable of training and running sophisticated machine learning models such as the LLMs (large language models) so commonly used to provide high-level natural language processing and generation.

These processing devices have typically been GPUs (graphics processing units), but are increasingly likely to be TPUs (tensor processing units), chips designed specifically to deliver efficient machine learning. Accompanying them are systems for handling data storage, project management, compliance governance, integrations, user interfaces, and anything else relevant. These systems can be packaged in AI platforms for convenience.

Every business has a unique set of goals and challenges, so there's no universal standard for what's included in AI infrastructure or how exactly it's assembled. Regardless of the chosen configuration, though, an effective AI infrastructure foundation must keep varied

processes and technologies working together to squeeze maximum performance from AI tools.

Note that AI infrastructure is technically optional, but since competing without the use of AI technology is ever-more difficult, it's hard to make a compelling case (particularly in the long term) for doing without it. Businesses who try are sure to struggle with critical applications from predictive analytics and customer support to at-scale automation and risk assessment.

Consider the industries investing heavily in AI. Financial institutions are leaning on AI to spot fraud by poring through vast quantities of transactions and flagging abnormalities in real time. Healthcare organizations, meanwhile, are getting systems in place to further key processes such as diagnostics and treatment planning, reducing the stress on their workers and delivering superior outcomes.

In short, then, AI infrastructure is the base that allows enterprise businesses to make the most of AI's transformational potential. Businesses that make it a priority are thus far better positioned to move with the times.

AI infrastructure vs. IT infrastructure: Why does AI need dedicated infrastructure?

Since you're no doubt familiar with IT infrastructure, you may be wondering what sets AI infrastructure apart. Shouldn't a solid IT foundation have everything you need to achieve excellent AI tool performance? In actuality, AI workloads take specialized resources and system arrangements to accomplish. Here are the key reasons why dedicated infrastructure is essential for AI:

AI requires specialized processing power

Conventional IT infrastructure mostly uses CPUs (central processing units) that are designed for general computing and consequently not exceptionally good at any specific tasks. Typical CPUs can achieve decent results with AI tools through brute force, but only through racking up immense resource costs.

AI infrastructure, however, uses GPUs and TPUs optimized for the complex parallel processing that's so important for AI algorithms. GPUs were of course developed to accelerate 3D rendering by taking care of the many distinct processes needed to show realistic virtual worlds (with lighting and physics being particularly tricky), while TPUs were custom-made to efficiently accommodate the fundamental processes of neural networking at incredible scale.

And scale is a vital word here, because the demands placed on machine learning systems are extraordinary. The difference between using standard-issue IT infrastructure and AI infrastructure isn't a saved minute or two here or there. It's spending days training a model instead of spending weeks or even months, making it easy to see why the cost of setting up AI infrastructure is worth it.

AI needs scalable data storage

AI applications demand and generate enormous volumes of data, and that burden grows as they prove their value and are subsequently joined by further AI tools with wider remits. IT infrastructure from pre-AI times never needed to scale so dramatically. Relational databases would build up with system data, certainly, but it would never become anywhere near as weighty or essential as AI data.

Dealing with this demand requires one of two solutions. The first is building the type of data storage facility that's common for the largest of enterprise businesses, and the second is using an AI-geared cloud storage solution with the capacity and capability to keep data moving without major bottlenecks.

The former makes sense in rare situations, but the latter of those two is naturally easier to justify. Provided compliance is managed, performance is consistent, and functionality is steadily developed, a cloud-based storage system for AI will be the right option for almost any business.

AI depends on low latency and high throughput

Most of AI's most notable applications (handling expansive real-time analytics, steering autonomous vehicles, monitoring cybersecurity, etc.) need to achieve low latency (the gap between calling for data and receiving it) and high throughput (the amount of data that can be delivered in a given amount of time). Common IT setups aren't nearly that demanding, so IT infrastructure is relatively limited.

A high-priority AI system needs a fit-for-purpose ethernet connection (Wi-FI won't suffice), software developed to operate as quickly and reliably as possible, and seamless integrations with systems supplying or receiving data. When lives are on the line, as is the case when handling autonomous vehicles or supporting hospital operations, even a tiny delay (one that would surely be insignificant in most other situations) has the potential to prove disastrous.

AI benefits hugely from software customizations

Think about the classic IT task of putting together a spreadsheet. Does it help to implement customizations within your spreadsheet software of choice? Macros, keyboard shortcuts, app integrations? Yes, it helps, but it doesn't help that much. You can open up a fresh installation of Microsoft Excel and get the job done almost as quickly and effectively (and with little added stress).

With AI tasks, however, any and all specialized customizations can drastically improve both the results and the pace of processing. Investing in software stacks using machine learning frameworks and languages devised to aid data analysis leads to better resource use, easier deployment, and superior iteration.

Achieving these customizations isn't easy, though, which is why AI infrastructure is something that must be carefully developed over time (rather than something completed and put into place as a static solution before any work can be done). Only through incremental

tweaking can you achieve an AI foundation that fully suits your business and supports you in everything you're trying to do.

AI demands enhanced security and compliance

For optimal results, AI systems must draw upon as much high-quality data as possible, and that data must be diligently protected to ensure total compliance with privacy and safety regulations. This is especially important given that the inherent variability of generative systems makes them prone to disclosing things they should keep under wraps (even if their instructions suggest otherwise).

Usual IT infrastructure deals mostly with internal data of limited sensitivity. It isn't ideal for a report never intended to be seen by those outside an organization to be disclosed, but it's probably not concerning. For an AI system aiding diagnostics or cybersecurity, however, the stakes are so much higher. A medical patient's history being leaked is a gross violation of their rights.

To keep things secure, AI infrastructure calls for AI-specific solutions employing strong security measures such as role-based access and end-to-end encryption. It also demands data sourcing and storage processes that follow the law to the letter by auditing data use on a regular basis, ensuring strict compliance with current regulations (such as GDPR and HIPAA) and promising adherence to future regulations as and when they're implemented.

AI infrastructure and operations fundamentals

If you hope to effectively implement and manage AI, you should first form a solid understanding of the four core components of AI infrastructure. Here's a breakdown of the basics to help you get there:

Data storage and processing

The more efficiently you can manage large data volumes, the more productively you'll be able to train and update your AI tools. Robust AI infrastructure pairs data warehouses (and data lakes, which are vast pools of unsorted data) with systems intended to make data management easier, such as compression algorithms capable of saving large amounts of space or deduplication programs developed to keep training data clean and effective.

For maximum impact, this data must be easy to move between systems and models, and that also calls for sophisticated stream processing and analysis. A robust pipeline and distribution process will deliver excellent responsiveness and scalability while keeping downtime at a minimum.

Compute resources

As we touched upon earlier, having the right compute resources in place is critical for performing the complex calculations involved in AI workloads. Modern GPUs (which often have some AI functionality baked in) and TPUs are ideal due to their relative proficiency at

parallel processing. CPUs are still necessary for related processes, but they shouldn't do the bulk of the heavy lifting for AI infrastructure.

HPC clusters (HPC meaning high-performance computing) are commonly made available through the cloud, allowing many businesses, even those below the enterprise level, to avail themselves of high-level AI tools. They allow dynamic resource allocation with accordingly variable pricing, allowing supply to scale neatly with demand and provide consistent value.

Machine learning frameworks

As machine learning has matured, huge developers have worked on frameworks that provide essential tools and guidance to accelerate the process of developing and deploying AI models. Google, for example, helped build TensorFlow, while Meta developed PyTorch. Taking advantage of these frameworks allows faster model creation, testing, implementation and optimization.

The extensive software libraries and pools of community-led resources within machine learning frameworks help experts stay up-to-date and give newcomers the support they need to start getting results. The frameworks also contain pre-built environments for experimentation and debugging, many of which can be rolled out directly (and thus rapidly) within popular cloud platforms.

MLOps

MLOps, standing for machine learning operations, is an assortment of principles and practices developed to streamline the use of machine learning workflows. Where frameworks contain software and informational resources to help people use it for model development and deployment, MLOps concentrates on aiding the use and improvement of models once they're in place.

Keep in mind that even the most finely built AI model won't return much value if those tasked with using it don't know what to do. MLOps guides teams to design reliable workflows and effectively collaborate on model maintenance and iteration, leading to consistent progress and a level of operational transparency that encourages course-correction when performance wavers.

How to build AI infrastructure in five simple steps

To make it easier for you to get started with building effective AI infrastructure, we've identified five easy steps you can follow. By making steady progress through the following process, you can move towards a positive outcome.

1. Clearly define your objectives and budget

The very first thing you must do is outline your business objectives, specify how AI can help you further those objectives, and set out your expected outcomes of implementing it. Note that your *desired* outcomes are distinct from what you can reasonably expect. If you achieve

your expected outcome, you can build on it. If you focus on a wildly optimistic outcome, you're likely to be disappointed.

Due to how much an investment in AI infrastructure can cost, be sure to factor in both short-term and long-term ROI. It's a mistake to commit too much too early, but it's also a mistake to commit too little and miss out on the potential payoff. And check that you've taken all potential costs into account. Setup, software, processing, maintenance, scaling, consultation… There are so many things to keep in mind, so take the budgeting process seriously.

2. Choose between cloud and on-premises deployment

We earlier identified two approaches to putting together the resources needed for AI infrastructure, and it's here that you must choose between them. Can you meet your needs with the flexibility, convenience and scalability of cloud-based AI infrastructure? Or do you need the enhanced security, data protection and rich customizability of an on-premises private AI deployment?

The most likely scenario is that a cloud-based approach will fit you the best. It allows near-infinite scalability with no setup costs, and choosing a highly rated provider will give you confidence that you can rely on minimal downtime and a competitive rate of development and innovation. But if you're working with very sensitive data in a heavily regulated field, you should at least try to estimate the cost of taking the on-premises approach, as it may be right for you.

3. Invest in suitable hardware, software, and networking solutions

Selecting the right combination of solutions is absolutely crucial. Each part of the overall package should be exceptional in its own right, but the components must also fit together well to ensure smooth cooperation. Naturally, you should begin with the GPUs/TPUs that offer power and specializations suiting your unique needs. If you need rapid real-time functionality, for instance, confirm that your selected processing tier can accommodate that.

Once you've found a hardware lineup that does what you need and uses your budget well, think about the software you're going to use. Picking a standard machine learning framework will help here by giving you a straightforward template to follow, so take that route unless your AI-related intentions are so unusual that you can't do anything with a framework (this is very unlikely).

What you should choose for networking depends on your workload requirements and your chosen deployment model. If you're anticipating large-scale data transfers and/or real-time processing, you'll need ethernet networking that delivers extremely low latency and very high throughput. If you're not, you shouldn't suffer much from keeping your costs down here.

4. Establish data privacy and security measures

How much you need to do here depends on the sensitivity of the data you'll be using and the regulations active within your industry. Compliance isn't something to be risked here, as

falling foul of the law will damage (and potentially) ruin your company's reputation in addition to earning you punishing legal consequences. Due to this, it's better to do too much than too little. Aim to establish rigorous privacy controls and monitoring processes so attempted intrusions can be caught early and prevented from causing problems.

You must also put a lot of thought into the team you task with handling those processes. Everyone must know exactly what their responsibilities are and receive enough training to ensure they can get them done. Note that the training should extend past AI-specific elements to cover general cybersecurity. Even a superb set of security protocols can do little if a key member of your AI management team fails to secure their account and allows a malicious actor to access it.

5. Implement, maintain, and monitor

When you're ready to roll out your AI infrastructure, do so carefully with clear milestones and performance metrics to chart your progress. Don't expect everything to work perfectly right away, but be ready to act quickly if your setup isn't coming close to meeting expectations. And don't make the mistake of assuming that maintenance can wait because everything's newly installed. AI models are incredibly powerful and just as tricky to keep running optimally, so start the routine checks quickly and ensure continuous oversight.

Make good use of the monitoring and analytics tools you chose earlier to keep an eye on resource use, data quality, and output consistency. If everything's working as it should, don't make arbitrary changes. It's all about experimentation and steady development. It will certainly help to follow the AI industry closely, as this will help you learn about and draw upon fresh technologies and approaches.

AI infrastructure: The framework to harness the power of artificial intelligence

With AI infrastructure, companies have the foundational elements they need to optimally benefit from the power of artificial intelligence. By investing in dedicated infrastructure, enterprise businesses position themselves to use AI to good effect across the board, leading to transformative outcomes.

If you're eager to bring AI infrastructure to your business, consider that specialized providers can offer tailored solutions to suit different industries, and there's surely a solution out there that's right for you. A custom on-premises deployment can bring enhanced AI security and data governance if you need strict compliance, or you can embrace the ease and scalability afforded by a cloud solution.

Whatever course of action you take, this term (along with everything that accompanies it) is one to remember. As competition heats up across all industries, businesses that invest extensively in AI will lead the way in everything from data analysis to innovation, while those that don't will surely struggle.

What Is Descriptive Analytics? Definition, Benefits, and Examples

Descriptive analytics means analyzing data to answer questions of the form "what happened?" or "what is happening?" Answering questions of the form "what happened?" is specifically known as historical (or static) analytics, while answering questions of the form "what is happening?" is known as real-time analytics.

Descriptive analytics can be used either on its own to answer simple business questions like "Did overall sales go up or down last month?" or alongside more complex forms of data analytics to reach specific company goals.

This page explains what descriptive analysis is, describes how you can use descriptive analytics to inform strategic business decisions and organizational choices, provides examples of descriptive analytics in practice, and discusses its benefits and limitations.

Descriptive analytics definition

Descriptive analytics involves interpreting historical or real-time data to better understand trends or relationships between online customer-driven events. An event is an action taken by a customer online in which they interact with products or services, and events can be recorded for the purpose of analytics.

Descriptive analytics can help your company understand its market better and determine whether it's on track to achieve its goals. It is the most basic form of data analytics in that it aims to answer simple descriptive questions. Importantly, descriptive analytics does not aim to answer "why" something happened.

As such, descriptive analytics is often an organization's entrypoint into the data analytics space. It is common to begin with descriptive analytics and then, with more experience, move on to more complex types of data analytics such as predictive analytics, customer data integration or the 360-degree customer view.

As noted, this type of analytics doesn't give reasons why something happened (this is covered by diagnostic analytics), but it also doesn't attempt to forecast the future (predictive analytics) or identify which actions should be taken (prescriptive analytics). Regardless, descriptive analytics is an essential part of data analytics, as a solid understanding of the past can provide useful insights. If paired with other forms of data analytics, it can give a rounded view of a business.

What is descriptive analytics used for?

Descriptive analytics provides businesses with clarity on their current and potential performance, offering insights that help them shape their strategies and make adjustments to their operations. By presenting historical data in convenient reports and/or dashboards, it draws attention to key trends and patterns and shows success relative to targets.

Here are some effective use cases for descriptive analytics:

Tracking business performance

Companies that wish to benchmark their performance over time (comparing successive years, for instance, or specific sales quarters) can use descriptive analytics to do just that. They can break their metrics down however they like to uncover insights that may help them improve.

Highlighting operational bottlenecks

A manufacturing business, for example, might wish to pinpoint minor obstacles within its production lines, as such obstacles can have compounding impact. A customer service team, meanwhile, might want to determine when support requests spike, allowing the appropriate allocation of staff. Using descriptive analytics can aid this process.

Evaluating marketing campaigns

Marketing teams can use descriptive analytics to gauge how different efforts perform based on metrics like click-through rates, conversion rates, or social media engagement. This retrospective view enables them to refine their campaign ideas to better align with audience preferences and behaviors.

Managing risk

By examining historical incidents through descriptive analytics, businesses can detect recurring dangers (such as equipment failure or delivery delay) and take proactive steps to mitigate their impact.

In general, descriptive analytics is a core tool for understanding the state of a business, how it achieved that state, and how it can improve. It positions teams to make informed decisions, bolstering stability and supporting growth.

Descriptive analytics vs diagnostic, predictive, and prescriptive analytics

Descriptive analytics provides a foundation for understanding what has happened in a business, but it's only one way to look at things. Other forms of analytics (namely diagnostic, predictive, and prescriptive) contribute further to business assessment, yielding additional insights and recommendations. Here's how these analytics models compare to one another:

Diagnostic analytics: understanding why things happen

While descriptive analytics details what's happened, diagnostic analytics focuses on *why* they've happened. For instance, if descriptive analytics highlights a drop in sales, diagnostic analytics might indicate that the drop resulted from a supply chain disruption or a

competitor's promotion. And when a business understands the causes behind its metrics, it can address its underlying issues.

Predictive analytics: identifying what's likely to happen

Predictive analytics uses historical data to project future outcomes. By identifying patterns in data, it forecasts scenarios such as changes in customer demand or shifts in market trends. For example, a retail business might use predictive analytics to anticipate which of its products will see increased sales during a holiday season, allowing it to optimize its inventory investments.

Prescriptive analytics: deciding the right steps to take

Prescriptive analytics takes the insights from descriptive, diagnostic and predictive analytics and uses them to recommend suitable actions. If predictive analytics indicates an upcoming spike in demand, for instance, prescriptive analytics might prompt an increase in production and the rollout of a targeted marketing campaign to take full advantage of the opportunity.

How these approaches to analytics work together

These forms of analytics are most effective when used together. Descriptive analytics highlights what's happening, diagnostic analytics explains why it's happening, predictive analytics forecasts what could happen next, and prescriptive analytics recommends what actions to take. Together, they provide a complete picture that enables optimally well-informed decisions.

How does descriptive analytics work?

It's important to follow a structured process when using descriptive analytics, as this will allow you to achieve the best results for your business. Adhere to the following sequence of steps and you'll get where you want to go:

Define your question

First, your company must decide on a question that can be answered using descriptive analytics. This means it must be about an event or trend that happened in the past or is currently happening. It is common for companies to use key performance indicators (KPIs) to guide their questions, as these are their most important company goals.

The people who want to know the answers to the questions will be your project stakeholders. Once your stakeholders know what questions they want to ask, you must ensure that you have access to the data you need to answer them. If you don't, you need to find a way to collect this data before you can begin your descriptive analytics work.

Aggregate your data

Next, aggregate all the data that may be relevant to answering your particular question together in one place. Using data purchased from third parties (alongside your own

first-party data) can be helpful, as it can provide you with more information to answer your question.

Aggregating your data into one place may involve an ETL (extract, transform, load) process: extracting data from various different locations, transforming it so that all the relevant data goes into a single table, view, or dataframe for ease of querying, and then loading it into a new location (usually inside a data warehouse).

Perform exploratory data analysis

Now that your data is aggregated together, you can do some exploratory data analysis. This involves investigating your data set to look for any useful information or anomalies, which may give insights that will help when cleaning your data.

For example, if you find that a disproportionately high percentage of people have input "Mr." for their title, you may need to investigate, as this is often the default value for a title field. You could compare their first names to a list of names that have a definite gender attached to them, enabling you to fix some of these values (you won't be able to fix all the values, as some names are gender-neutral).

Choose your model

Exploratory data analysis can also help you decide on the most appropriate model for your main analysis work. This could be a SQL model, or a statistical model such as linear regression or a clustering algorithm. However, for descriptive statistics, by far the most common model to use is a SQL model, as SQL lends itself very well to performing groupings and aggregates like sums, counts, averages, and percentiles. Combining that with SQL being far easier to learn than complex statistics or machine learning, you can see why descriptive analytics and SQL often go hand in hand.

Build your model

Once you've decided on your model type, it's time to build it. If you've chosen a SQL model, this is best built by a data analyst. If it's a statistical model, you'll need to use your judgment of your and your team's skill sets to decide if a data analyst or a data scientist is best placed to do this work. Either way, you will need an experienced and knowledgeable person to choose a suitable statistical model from the many that exist.

Share your insights

The final stage is to produce the requested outcome to your stakeholders in the form of a report or data visualization. It's a great idea to make use of charts or other visualizations to form a data narrative – storytelling is a great way to ensure that your stakeholders are engaged and that they understand the importance of your analysis. Remember that some charts are more appropriate than others for certain types of data, and you should familiarize yourself with the best data visualization techniques for different types of data analytics questions.

Depending on the scope of your research, and your organization's budget, a descriptive analytics project can consist of answering a single question, where the project ends once the requested answer has been communicated to stakeholders. Alternatively, your descriptive analytics project can serve as a jumping-off point for more complex projects such as prescriptive analytics.

Benefits of descriptive analytics

Using descriptive analytics is advantageous in a number of ways, making it an important consideration for the average business. To expand upon this, here are some of the specific benefits it brings to the table:

Easier performance analysis and improvement

Using descriptive analytics allows workers to understand what's been happening within their companies, which then enables them to change what they're doing if their KPIs aren't being met. Since even a minor improvement can become very consequential over time, this is significant. And because descriptive analytics is easier to implement than other types of analytics, it cuts down the time needed to quickly spot trends and patterns or detect problems.

Convenient automated alerts that can steer workflow tweaks

The information that descriptive analytics provides can often be used to make small but efficient changes to business workflow, and this can be done very efficiently through automation. Setting up automated alerts that trigger off events within descriptive analytics (e.g. severe decreases in web traffic or conversions) is a great way to keep track of worrisome developments.

Greater information accessibility

Descriptive analytics typically uses simple reporting measures and analysis techniques, including line, bar, and pie charts, which a wider business audience can easily understand. It can also be used together with other data analytics techniques (diagnostic, predictive, and prescriptive analytics) to provide a more comprehensive picture of a business.

Simplified cross-departmental collaboration

Descriptive analytics makes data more accessible to teams across an organization, helping departments align their efforts. A marketing team, for example, can use trends in social media engagement to refine campaign strategies, while a customer support team might use website traffic reports to anticipate seasonal spikes in inquiries.

Enhanced strategic planning

By summarizing historical trends and patterns, descriptive analytics helps organizations refine their strategies. Businesses can use this information to assess the impact of past decisions and adjust their plans to better align with long-term objectives (such as improving efficiency or increasing customer satisfaction).

Challenges of descriptive analytics

While descriptive analytics offers significant advantages, it also comes with certain obstacles that must be overcome to use it effectively. Below are the key challenges you can expect to encounter if you seek to implement it:

Inconsistent predictive power

Because descriptive analytics can only consider what's already happened or is currently happening, it can't fuel reliably accurate predictions. There are simply too many things that can change as time goes by, and too many variables that can't be fully accounted for. It takes broader technical and industrial knowledge to make better predictions, and only relevant expertise can deliver that.

A need for further analytics

While descriptive analytics highlights patterns and trends, it stops there, limiting the utility it can offer when it isn't paired with other forms of analytics. Any business that wishes to implement it, then, needs a corresponding plan to carry out diagnostic analytics and predictive analytics.

Dependence on data quality

The accuracy of descriptive analytics depends heavily on the quality of the data being analyzed. Incomplete, duplicate, or erroneous data can lead to misleading insights. Due to this, prioritizing data cleansing and validation is crucial for ensuring the analysis produces reliable results.

Risk of measuring the wrong metrics

Descriptive analytics is only as useful as the metrics being analyzed. If the wrong metrics are chosen (or even if the right metrics are picked but aren't suitably aligned with business goals), the resulting insights may lack relevance or bring confusion rather than clarity.

Potential for information overload

With even a modest modern business gathering vast volumes of data, it can be challenging to figure out which areas to focus on. To avoid getting completely lost, it's vital to start with clear goals and steps for choosing and interpreting data, and to guard against scope creep causing issues down the line.

Bias in data selection and framing

Stakeholders may choose to highlight metrics that confirm their assumptions or make performance look better than it really is so they can solidify their positions and/or make more money. All the while, the business as a whole can suffer. This calls for a firm commitment to choosing and analyzing the data fairly, regardless of whom the results flatter or what the overall performance seems to be.

Reliance on human interpretation

Because descriptive analytics only lays out the data, it calls upon stakeholders to provide appropriate interpretations. If they fail to do so, the analytics project can lead to poor decisions, and potentially even a complete disconnect between the state of the business and how it's perceived.

Examples of descriptive analytics

Descriptive analytics provides tangible benefits to your company by asking well-formulated questions that have a clear, achievable answer, such as "How many users did our website have last month?" or "What was the conversion rate for our most recent marketing campaign?" Some examples of how descriptive analytics can be used across different sectors of a business are explained below.

Financial metrics

You can use descriptive analytics to produce financial reports on month-over-month sales growth or year-over-year product price changes. These reports allow you to monitor your company's financial health over time.

Social media engagement

Your marketing team can use descriptive analytics to help them find out what content is most popular, and they can use this data to work out what kind of content they should share in the future. They could start by generating reports on how their social media likes, shares, and clicks change each month.

Web traffic reports

If your web team wants to improve your company's website, a web traffic report is an example of descriptive analytics that can show the popularity of different pages on the site (by seeing how many users are clicking on each page). The team can then analyze their successful web pages to work out what they are doing right.

Trend identification

Descriptive analytics can be used to identify trends in customer preferences and behavior regarding your product. You can use this to make a reasonable guess about future demand for particular features of your product. Netflix, for example, uses descriptive analytics to power the "trending now" part of its recommendation system. The popularity of this system is evidenced by the fact that 80% of all content streamed on Netflix comes via its recommendation system.

General business reports

You can use descriptive analytics to generate internal business reports on metrics like stock or cash flow to get a snapshot of how your company is operating. Creating these reports can help you work out ways to improve efficiency, such as a new strategy for restocking products.

Automated alerts

Your company can use descriptive analytics reports as the basis of automated alerts; for example, when web traffic or key events decline by more than a specified amount, an automated email alert could be sent to the appropriate member of staff, enabling them to catch issues as early as possible.

Are there descriptive analytics tools?

It's fairly easy to find tools that can aid descriptive analytics. They may not have been developed specifically for it, but the following types of tools tend to be good at streamlining the lengthy process of collecting, processing and presenting data:

Business intelligence tools

Business intelligence (BI) tools such as Power BI, Looker Studio or Tableau are highly useful for descriptive analytics. Offering convenient user-friendly dashboards and solid reporting options, they help teams share data clearly and communicate more effectively.

Statistical tools and libraries

Tools such as SPSS (Statistical Package for the Social Sciences), the SAS (Statistical Analysis System) suite and the free open-source alternative Jamovi offer more advanced statistical analysis that can help here. Because they include various relevant features (from generating basic summaries to producing multivariate statistical models), they empower businesses to uncover deeper data insights.

Spreadsheet tools

Industry-standard spreadsheet tools such as Google Sheets and Microsoft Excel can carry out numerous operations that can aid descriptive analytics. They can calculate averages, frequencies, and ranges, and turn them into basic visualizations that will suffice in many situations.

Database management tools

When dealing with huge datasets, choosing the right management tools can make it much easier to retrieve relevant entries, aggregate them usefully, and create summary reports. SQL is the most popular system for handling structured datasets, so tools based on it (such as MySQL and PostgreSQL) are solid choices here, but there are tools using other query languages (such as Cypher or N1QL) that offer further flexibility by catering to different data structures.

Data preparation tools

Data going into descriptive analytics should be prepared properly beforehand to highlight file-type issues and reduce formatting errors. Tools like Alteryx, Talend, and Datameer can smooth out the process of cleaning, reformatting and integrating data drawn from different sources.

Descriptive analytics and customer data platforms

Descriptive analytics deals with historical data and involves detecting patterns and trends to better understand your business. When used alongside diagnostic, predictive, and prescriptive analytics, the information from descriptive analytics allows you to make even more informed decisions on your marketing strategies and other business functions.

Customer data platforms can help save time collecting this historical event data in the first place, or at the ETL stage during descriptive analytics projects. Instead of data analysts having to extract data from various sources, transform the data, and load it into a data warehouse, a customer data platform can automate much of this process.

Importantly, it allows you to set up a data pipeline between your different data sources and your data warehouse, continuously transforming your data into a useful format as it is loaded into the data warehouse. This allows data analysts to focus on the other steps in the descriptive analytics process.

What Is Fine-Tuning? The Comprehensive Guide To Fine-Tuning LLMs in 2025

How can you give your generative AI models the context and direction they need to be most useful to your organization?

A Large Language Model (LLM) doesn't simply appear fully developed. Instead, it requires a collaborative effort from developers using vast datasets and advanced technologies to bring it to life, especially for enterprise-level applications. Model fine-tuning is just one critical part of this larger ecosystem, helping to tailor the model for specific tasks and industries within the broader development process.

Each model undergoes a pre-training phase in which it's exposed to extensive datasets, teaching it to predict the next word in a given sentence and grasp the statistical relationships between words and phrases. This foundational training equips a model to generalize across a vast array of data, enabling it to identify and reproduce underlying patterns.

Once a model has shown it can grasp grammar, syntax and even contextual understanding from exposure to a wide range of human-written content, it's ready for the next time. To make it more than just a generic LLM, it's necessary to do some fine-tuning.

What is fine-tuning?

Fine-tuning, meaning the process of adjusting a pre-trained LLM to better suit a specific task, dataset, or use case, helps a pre-trained model adapt to nuances. It's part of what McKinsey calls the "Shaper" approach to generative AI: a way to scale generative AI capabilities, create greater accuracy, develop more proprietary capabilities, and meet higher security or compliance needs.

Fine-tuning vs. training

So, what is fine-tuning in AI relative to training? It's simple enough. Training refers to the process of teaching a large language model (LLM) from scratch by feeding it huge amounts of data in which it can detect patterns. During training, an LLM develops its core ability to predict the next word in a sentence, allowing it to analyze and respond to natural-language prompts.

Fine-tuning, however, is a more specific process. What does fine-tuning a model mean? It involves taking a pre-trained model and adjusting it to make it better suit a specific task. An LLM that was fed many types of content to serve a variety of purposes, for instance, can be fine-tuned to optimize it for a given task, such as translation, customer support, or even something niche like poetry writing.

When a business needs an LLM to accommodate its unique requirements, it can opt to skip the training stage by finding a pre-trained model and simply fine-tuning it as required. This is a way to save time, money, and effort.

Why do you need fine-tuning?

Fine-tuning plays a critical role in maximizing the performance of LLMs in real-world applications like enterprise tech stacks. It helps an LLM to understand and complete *specific* tasks. Such a task could be as simple as translation or sentiment analysis or as complex as Retrieval Augmented Generation.

By running a fine-tuning process, you improve a model's ability to perform targeted tasks and deliver high-quality results without the expense of training a custom model from scratch. Fine-tuning can help transfer learning, use time and resources efficiently, adapt to specific tasks, and mitigate bias and the risk of data leaks.

The process of fine-tuning can also help the model better understand the internal data of an enterprise. When a business adds an LLM to its tech stack by purchasing access to a generic model or rolling out its own, it will initially rely solely on training data that almost certainly won't have contained any data about the business, and it certainly won't know anything about the company's strengths, objectives, or desired uses for generative AI.

Fine-tuning a model by exposing it to a narrower dataset related to the desired application helps get the model fit for the enterprise's specific purposes, optimizing it so it can be released to the workforce with the information it needs to function effectively.

Common fine-tuning use cases

Some common use cases where fine-tuning can improve results include:

- Setting style, tone, format, or other qualitative aspects.
- Improving reliability in generated outputs.
- Correcting failures to follow complex prompts.
- Handling edge cases in specific ways.
- Performing new skills or tasks that are hard to describe.

Benefits of fine-tuning your LLM

Fine-tuning helps boost the capabilities of a large language model, especially when you're building on top of an LLM trained on vast amounts of data. Fine-tuning that model allows you to tailor the knowledge base to your specific needs while also ensuring compatibility and optimization with your distinct dataset (and doing so without the need to create a new model for each new use case). It also serves AI alignment, the process of encoding human values and goals into LLMs to make them optimally helpful, safe, and reliable.

This process is particularly useful for LLMs intended to take on critical tasks using sensitive data. Through fine-tuning existing models, businesses can achieve higher accuracy and better results on specific tasks than they could if they trained custom models from scratch. After all, they're leveraging broad capabilities and simply refining them as needed. With further training tweaks, a model can develop over time as new data becomes available or business requirements change, offering superb adaptability.

Prominent fine-tuning methods

When you undergo fine-tuning for your LLM, you are in effect adjusting the parameters of a pre-trained model on a task-specific dataset to improve its performance on that task. This can be done in a variety of ways, depending on business needs.

Prompt engineering

According to Google, prompt tuning is the "quickest way to extract domain-specific knowledge from a generic LLM without modifying its architecture or undergoing retraining. It involves crafting prompts that steer models towards certain domains of operation. For example, listing stylistic or tonal parameters can shape output ("use medical terminology and bullet-pointed lists"), and the identification of a specific role for a model to assume (e.g. "you're a senior robotics engineer") can make a huge difference.

General fine-tuning

As noted, general fine-tuning allows you to push an LLM towards a specific domain of expertise to make it more adept in that particular area, leaving it capable of providing more accurate and informative responses. It can be done by adding extra dense layers or by unfreezing the model to allow all weights to be updated.

Gradient-based parameter importance ranking

Gradient-based parameter importance ranking helps rank the importance of features in a model and is determined by how much the accuracy decreases when a parameter is excluded. This may not produce significantly more precise results, but it can improve model speed and ensure that the right elements are prioritized.

Reinforcement learning

Reinforcement learning aims to emulate the way that human beings learn by giving models feedback about the outputs they produce and having them iterate accordingly. Rather than drawing upon added training data, then, it has a model draw upon comments provided by relevant experts to refine its output until it can consistently meet their standards.

How does model fine-tuning work?

Here's how to do fine-tuning in 5 steps:

1. Select a pre-trained model.
2. Prepare a targeted dataset by preprocessing and formatting it to match the input requirements of your pre-trained LLM.
3. Fine-tune the model through your preferred process, ensuring the model's parameters are adjusted based on the new data with the objective to minimize a chosen loss function.
4. Evaluate and validate your results, ensuring the model generalizes well to unseen data and produces reliable predictions.
5. Establish a system of iterative refinement. There will likely be multiple rounds of experimentation spent adjusting parameters, trying different architectures, and considering if more training data is needed.

You might also consider fine-tuning hyperparameters (such as learning rate, batch size, and regularization strength) to improve a model's performance before deploying it for real-world use by integrating it into software systems or services.

Limitations of fine-tuning

Unfortunately, fine-tuning a model is a long and compute-intensive process. It takes a vast quantity of high-quality data (data that is representative and diverse enough to capture the variability of the target domain) as well as a huge budget and a lot of infrastructure to even begin a fine-tuning process.

According to a Google Research blog, "serving a single 175 billion LLM requires at least 350GB of GPU memory using specialized infrastructure", with a powerful LLM containing over 500 billion parameters. Not all companies can afford to accommodate such processing requirements, and that's before you factor in the data.

You need large amounts of unlabelled data to even reach the fine-tuning stage, and acquiring that data can be complicated. It's difficult to gather such amounts of data without running the risk of including sensitive information, and having such information emerge in outputs (particularly when they're public-facing) can be hugely problematic.

Data that's been sufficiently anonymized must also be sorted and categorized, a process that's expensive and tedious. Labeling can also introduce bias through human error and/or personal views, making it necessary to give ethical considerations serious priority.

Complicating things further still, there's no defined end to fine-tuning. There will always be new data becoming available for a given field, and models have no reliable access to real-time information so keeping a model competitive requires periodic work. Also note that updating training data without disrupting existing productivity is no easy feat.

Finally, if you can handle all of that, the effort might not even prove worthwhile. Recent developments in generative AI involving vector databases may make fine-tuning enterprise LLMs unnecessary by delivering contextual retrieval. Retrieval Augmented Generation allows models to dynamically draw upon independently updated knowledge resources, ensuring that they never become outdated.

Fine-tuning process and best practices

Understanding how fine-tuning works can help ensure your model fine-tuning efforts are effective and efficient. The process begins with selecting a pre-trained model, after which you must prepare a dataset that caters to the task you wish to optimize the model for. (Note that the data format must match the input demands of your selected model, so make adjustments if needed.)

For optimal fine-tuning, follow these best practices:

- **Choose your model carefully.** Pick a pre-trained LLM with a strong foundational knowledge that aligns well with your industry or desired application.

- **Prepare high-quality training data.** The better and broader your dataset is, the more effective your fine-tuning will be. Making an effort to learn how to prepare data for fine-tuning will pay off. Be sure to label your training materials effectively and

check them for accuracy, because miscategorized or inaccurate data will lead your fine-tuning to introduce major flaws.

- **Carry out extensive validation.** After fine-tuning the model, you must thoroughly evaluate its performance against validation sets. If it hasn't significantly improved, that's a clear indication that the fine-tuning process didn't work (possibly due to issues with the training data).

- **Periodically add fresh data.** Fine-tuning isn't something you should do once then forget about. Regular retraining with fresh data is key to maintaining relevance and accuracy, especially when the model is geared towards a task within a rapidly changing discipline.

If you adhere to these best practices and pay close attention to what you're doing, you'll soon know how to manage fine-tuning efficiently, setting you up to achieve superior LLM results while keeping your costs down.

Fine-tuning applications

The applications of fine-tuning AI models span a wide range of industries. By applying AI fine-tuning techniques, organizations can customize their chosen LLMs to perform specific tasks more effectively. Here are some of the most common applications of fine-tuning:

- **Healthcare.** Fine-tuning embedding models (which can effectively capture the semantic connections between words and concepts) can be very effective for rapidly analyzing medical records or research documents. Adding hyper-relevant training data covering specific medical disciplines can allow these models to deliver more accurate diagnostics.

- **Legal and compliance.** What is fine-tuning a model in a legal context? It means teaching a model to handle large legal datasets and understand essential compliance protocols. This can result in faster and more precise document review and risk assessment capabilities (with the precision being particularly important in an industry dominated by technicalities).

- **Customer service.** Fine-tuning models can make them better at handling specific customer queries, following brand guidelines for tone, and offering elements of personalization within conversational settings (such as chatbot exchanges). Additionally, training a model on technical product documentation can allow it to provide effective technical support.

By fine-tuning an LLM for specialized applications, a business can leverage the power of a pre-trained model while ensuring it meets the required standard.

How much does fine-tuning cost?

How much does fine-tuning cost for an average enterprise implementation? In truth, there's no useful answer because the cost of fine-tuning varies dramatically based on the complexity and scope of the project. For small-scale niche use cases, fine-tuning can cost

mere dollars, while costs for fine-tuning the most expansive LLMs can escalate into the hundreds of millions of dollars.

It makes more sense to consider the cost of fine-tuning in comparison to creating a model from scratch. Fine-tuning a pre-trained model is typically far more cost-effective than building an LLM from the ground up, which requires vast amounts of compute resources, data collection, and infrastructure. And since fine-tuning improves performance by reducing token usage, it also serves to significantly reduce computing costs in the long term.

FAQs about fine-tuning

Still curious about fine-tuning? Here are our answers to some of the most commonly asked questions on this topic.

What is AI fine-tuning?

AI fine-tuning is the process of refining a pre-trained model to improve its performance on a specific task or dataset. Taking an existing model and tweaking it as needed offers a strong combination of economy and efficacy.

How to prepare data for fine-tuning

To prepare data for fine-tuning, ensure your dataset is high-quality, broad, and correctly labeled. Your training data must meet the input requirements of the model, so reformat it as needed.

What is an example of fine-tuning LLM?

An example of fine-tuning an LLM is customizing a pre-trained model to provide personalized customer service interactions. Exposing the model to customer interaction data and customer purchase histories will teach it how to field tricky queries and make customers feel valued.

When should you go for fine-tuning LLMs?

Fine-tuning LLMs is ideal when it isn't economical to invest in training bespoke models but existing models aren't quite right for performing the tasks you need to get done. Whether you need a level of accuracy you can't find, a range of knowledge no existing model possesses, or to parse input data current systems aren't familiar with, fine-tuning can get you the results you're looking for.

What is the difference between fine-tuning and transfer learning?

The difference between fine-tuning and transfer learning is that fine-tuning adjusts an existing model's parameters to improve performance for a specific task while transfer learning adds new training to an already trained model without altering its core structure. Transfer learning is economical when an existing model is already well-suited for a task very similar to yours, as you only need to add enough training data to cover the differences.

However, if you're using a general-purpose model for a more niche task, you'll need fine-tuning to achieve the desired level of accuracy.