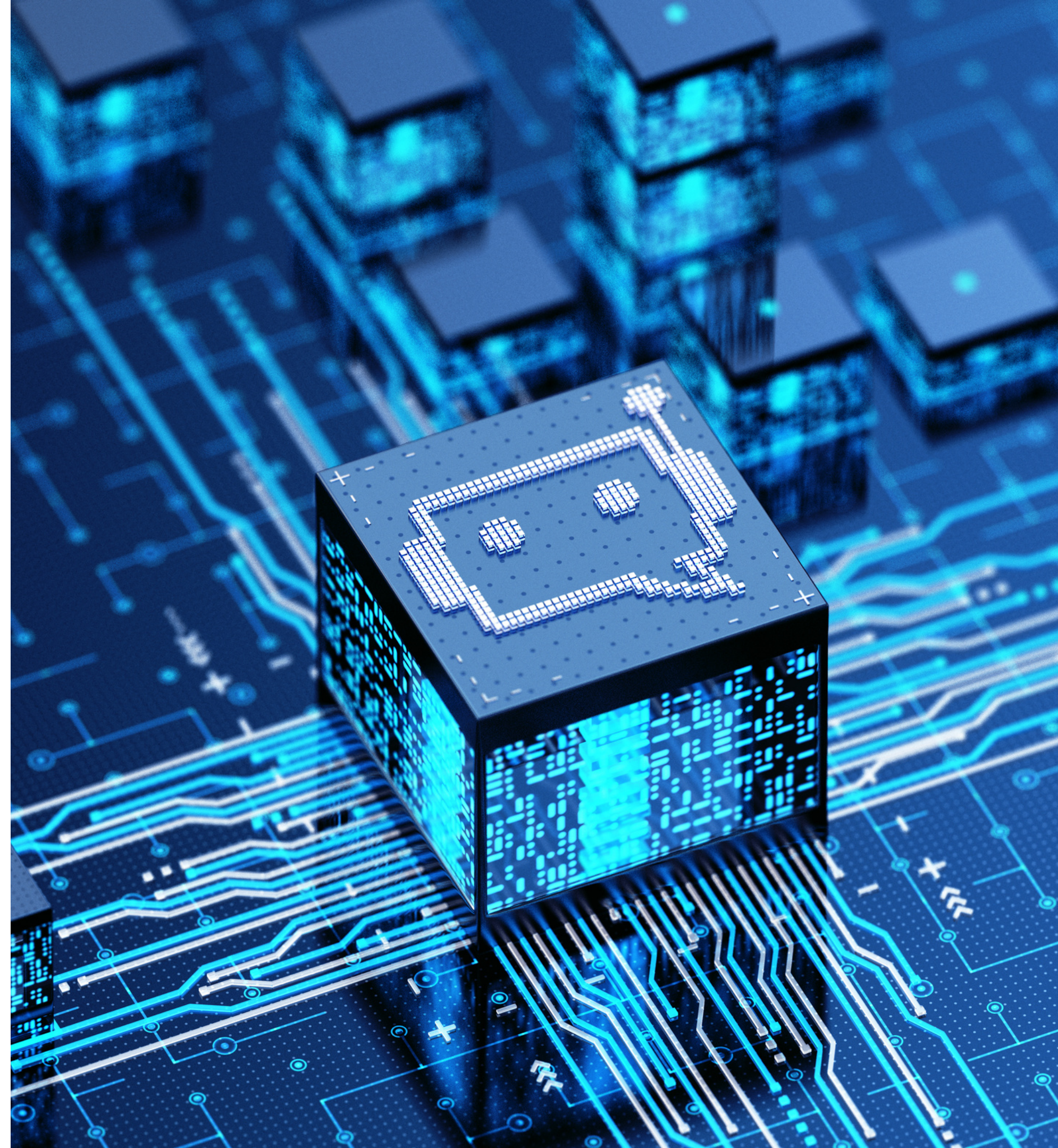


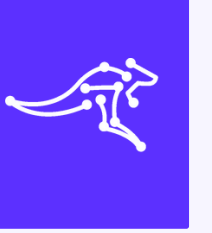
Unlocking the Power of Generative AI:

An Executive's Guide

May 2023



| Outline

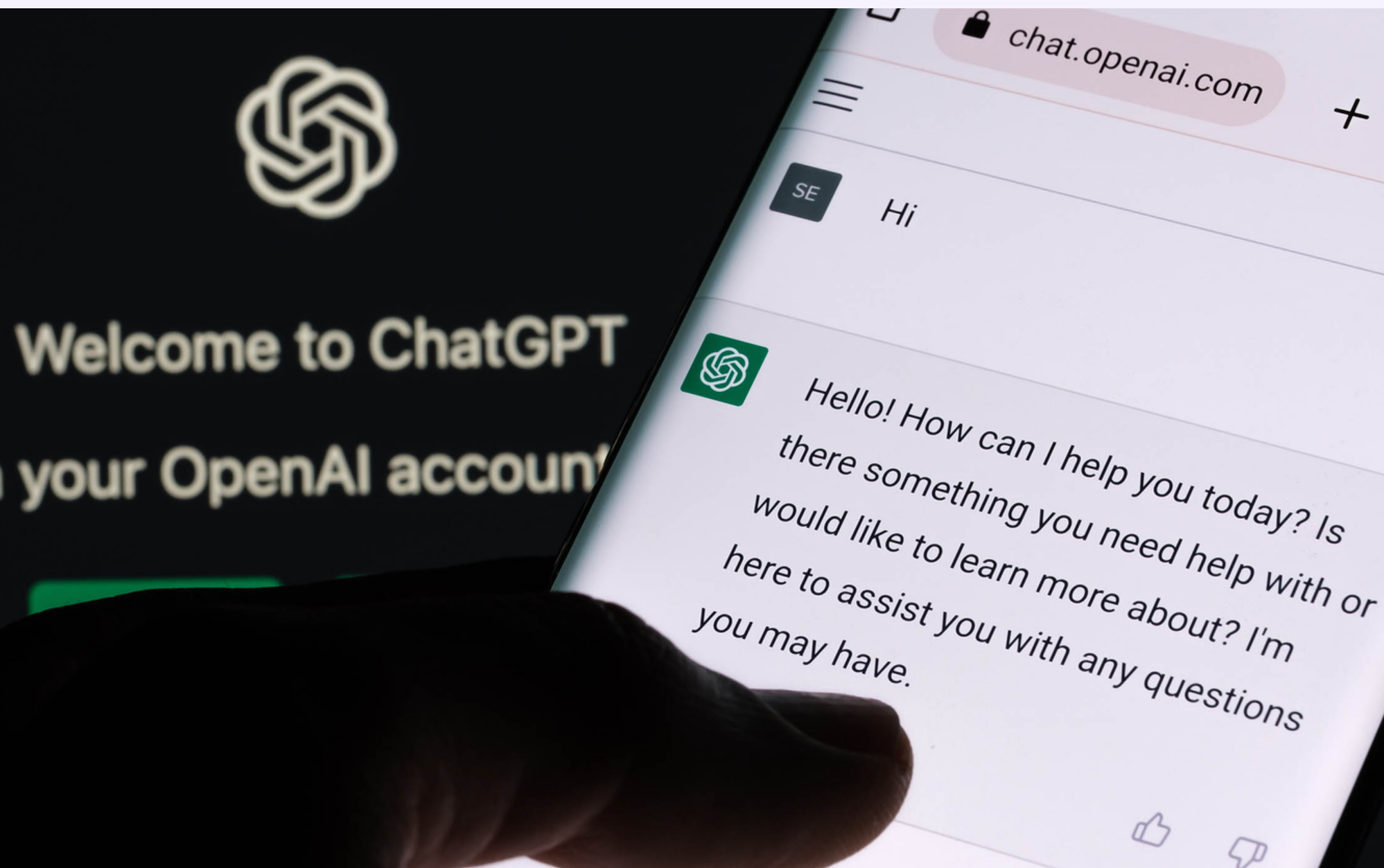


- **The Rise of Generative AI (Gen AI)**
- **Large Language Models Powering Generative AI**
- **Building with Generative AI**
- **Leveraging the Power of Generative AI for your business**
- **Using MLOps to create business value of Generative AI**
- **The Katomic advantage**



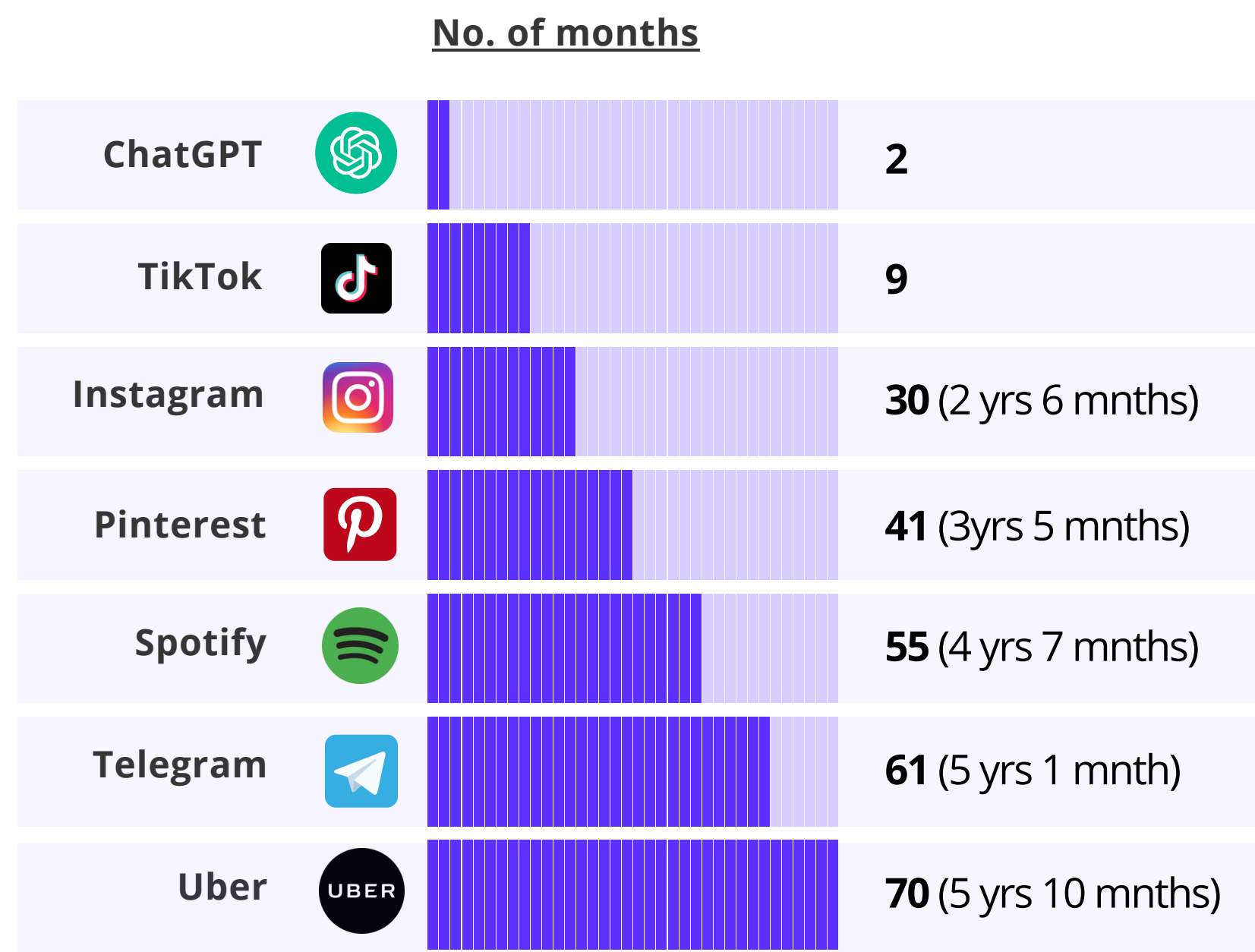
ChatGPT – The Big Bang Moment for Gen AI

ChatGPT's explosive global popularity became an inflexion point for AI's public adoption.



ZOOMING AHEAD

Time to reach 100 million monthly active users



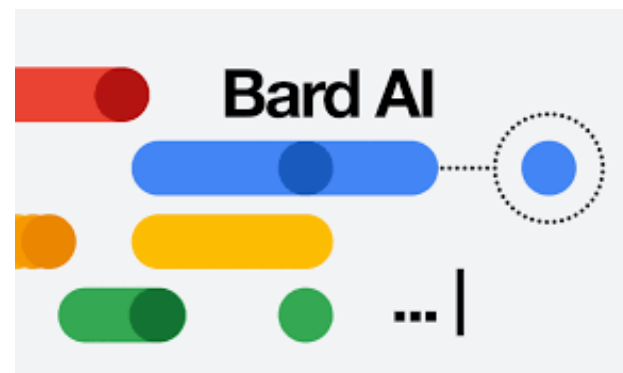


Large Language Models are not limited to just text

ChatGPT by Open AI



Bard by Google



BLOOM by BigScience



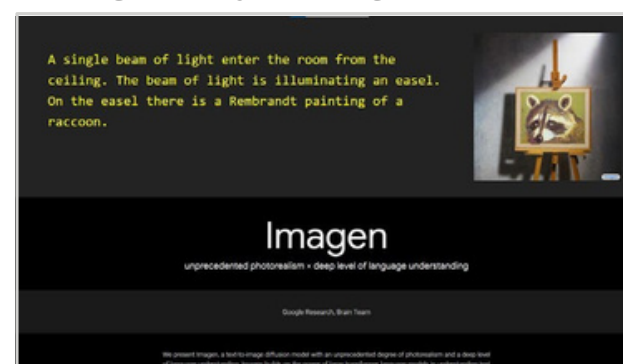
Galactica AI by META



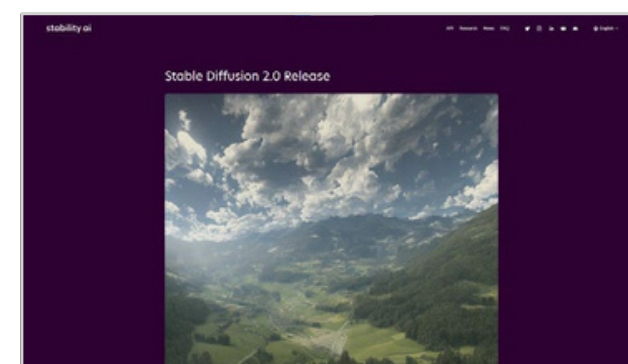
DALL-E2 by OpenAI



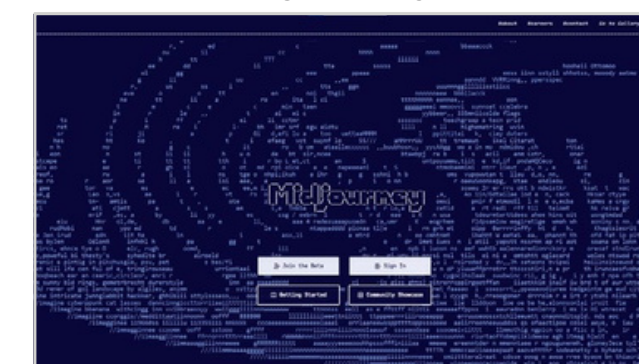
Imagen by Google



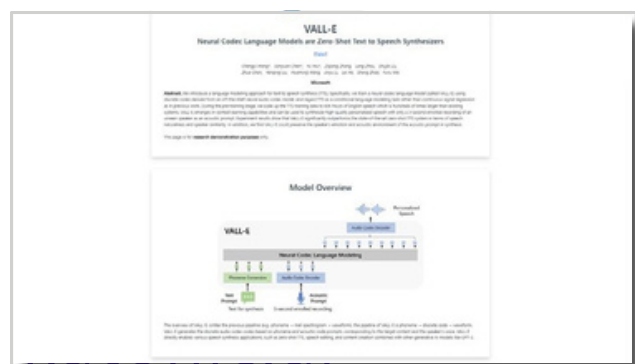
Stable Diffusion by Stability AI



MidJourney v4 by MidJourney



Vall-E by Microsoft for speech



PointE by OpenAI for 3D objects

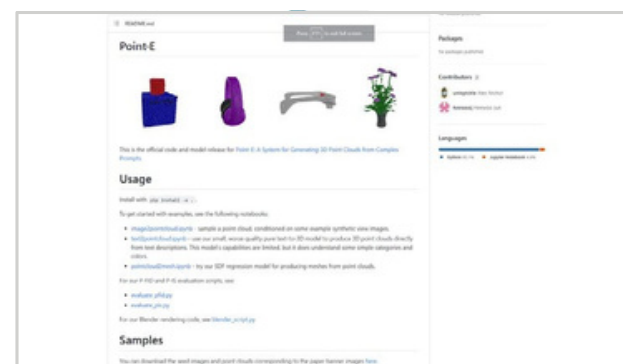
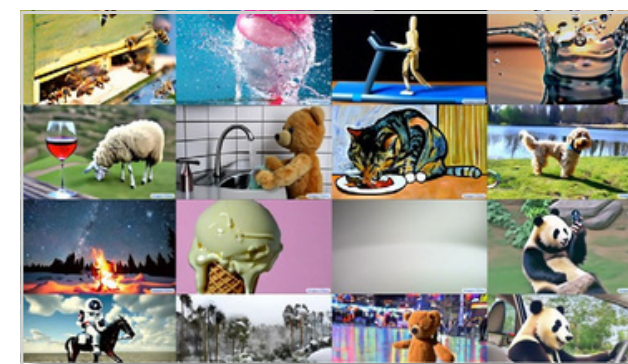


Imagen Video by Google



Chinchilla by Deep Mind for text

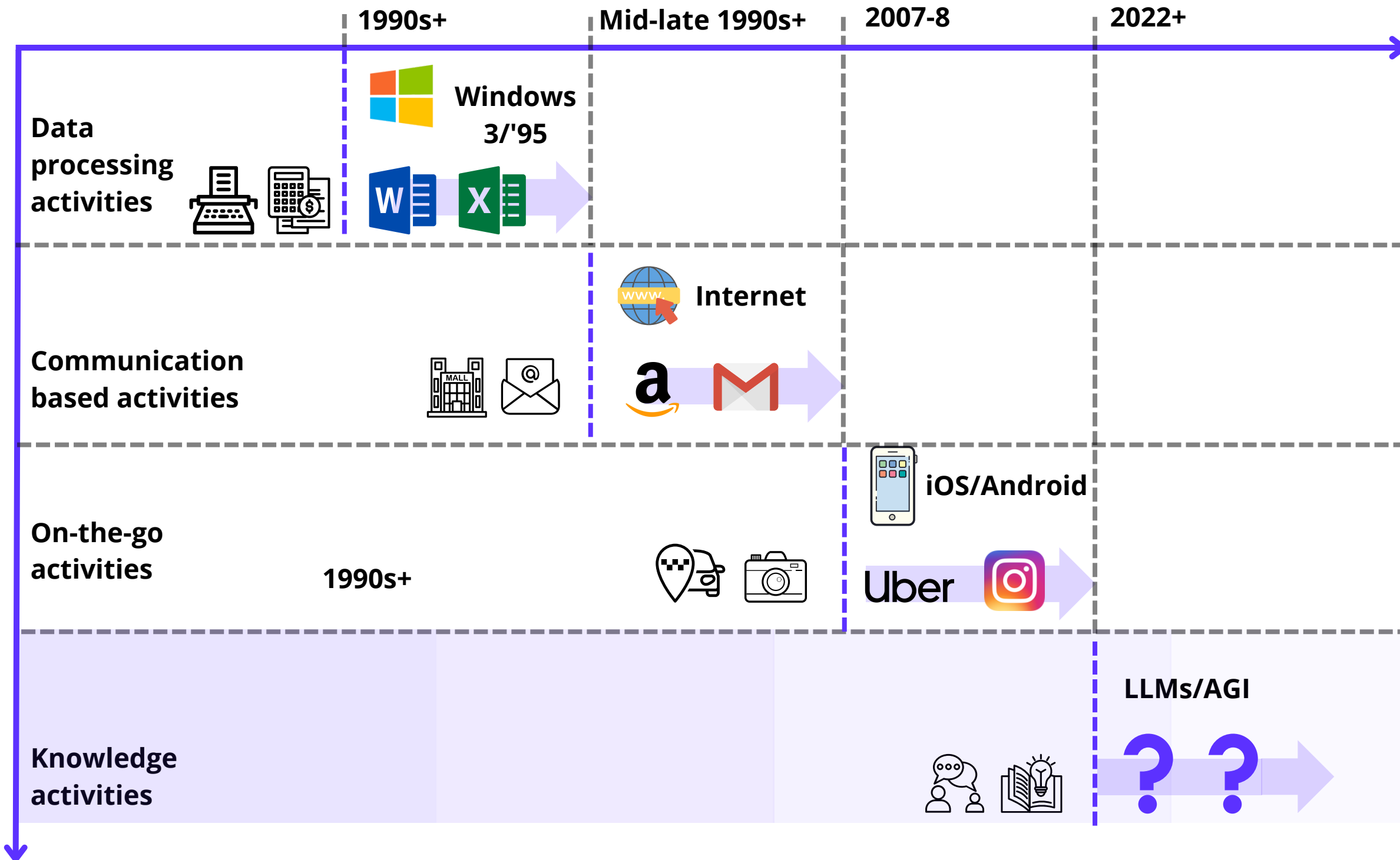




LLMs Are The Engine | AI Applications are the Product

Major platform launches that have enabled new types of applications, over time

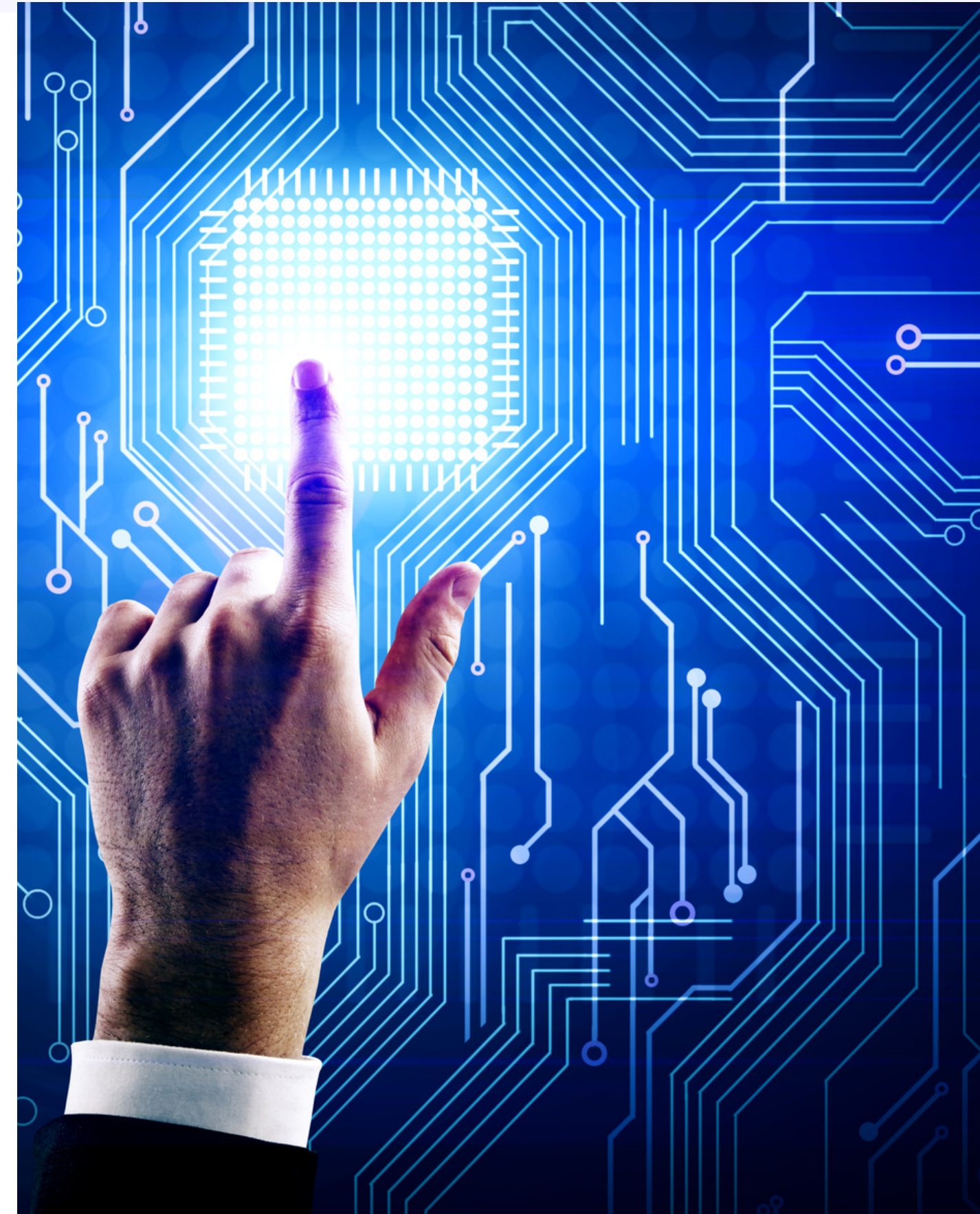
Range of human activities that software has "eaten"





Potential opportunities and use cases

- ▶ **Finance & Legal:** Draft and review documents, patents and contracts; find, summarise and highlight important points in regulatory documents; find and answer specific queries from large documents; scan through historical data to recommend a course of action.
- ▶ **Marketing & Sales:** Automate SEO-optimised content generation, enhance ad bids, hyper-personalise communication and deployment, create product user guides by persona, analyse & segment customer feedback, hyper-capable chatbots for upsell and cross-sell .
- ▶ **Customer service:** Natural-sounding, personalised chatbots and virtual assistants can handle customer inquiries, recommend swift resolution, and guide customers to the information they need.
- ▶ **HR & Recruitment:** Smart-shortlist of candidates, risk assessment of candidates, self-service of HR functions via chatbots and automation
- ▶ **Information technology:** Advanced code writing code and documentation, code review and error detection, and accelerated software development, auto-complete data tables, generate synthetic data.





Data Security , privacy and cost is a BIG hurdle

The screenshot shows the top section of The New York Times website. At the top right, it says "The New York Times" and "MLOps community". Below that, there's a market ticker showing "US MARKETS OPEN" with various indices like Dow Jones, Nasdaq, S&P 500, META, and TSLA. The main headline is "Disinformation Researchers Raise Alarms About A.I. Chatbots". Below this, there's a section for "THE WALL STREET JOURNAL" with a sub-headline "Amazon warns employees not to share confidential information with its answer 'clo' inside the com". Another article is titled "JPMorgan Restricts Employees From Using ChatGPT" with a sub-headline "Verizon and other organizations have also blocked access to the popular AI chatbot".

THEY DIDN'T EVEN ASK —
ChatGPT is a data privacy nightmare, and we ought to be concerned
 ChatGPT's extensive language model is fueled by our personal data.
 URI GAL, THE CONVERSATION - 2/8/2023, 6:12 AM

Australian Financial Review
<https://www.afr.com> › Companies › Professional Services
PwC warns staff against using ChatGPT for client work
 6 Feb 2023 — The consulting firm is encouraging its staff to experiment with the AI chatbot but forbids them from using such tools in any work projects.



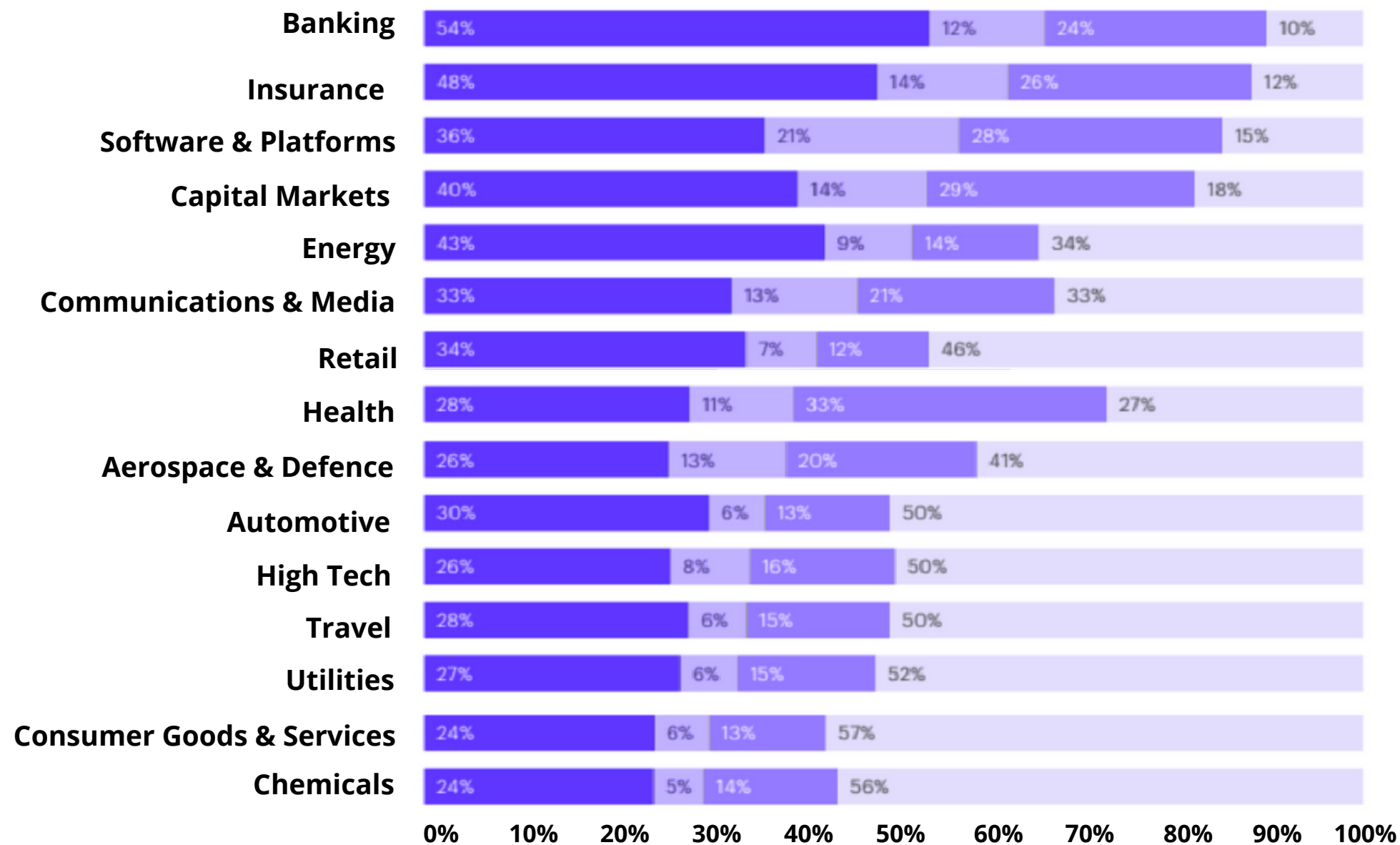
Adapt or be left behind

Companies must reinvent work to find a path to generative AI value. Business leaders must lead the change, starting now, in job redesign, task redesign and reskilling people.

Nearly 6 in 10 organisations plan to use ChatGPT for learning purposes and over half are planning pilot cases in 2023. Over 4 in 10 want to make a large investment.

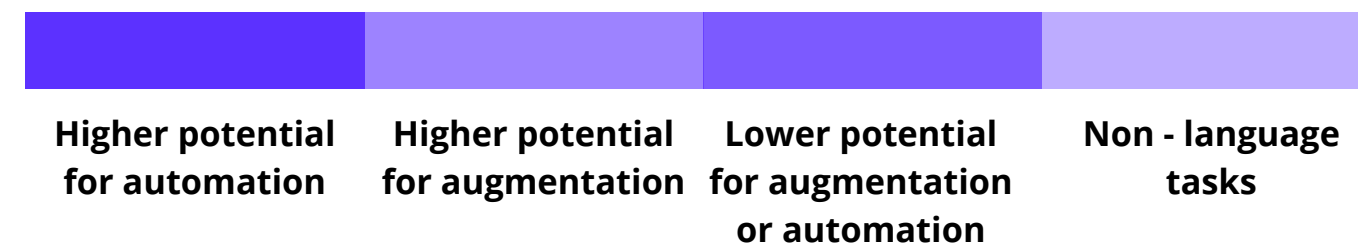
40% of working hours across industries can be impacted by Large Language Models (LLMs)

Why is this the case? Language tasks account for 62% of total worked time in the US. Of the overall share of language tasks, 65% have high potential to be automated or augmented by LLMs.



Work time distribution by industry and potential AI impact:

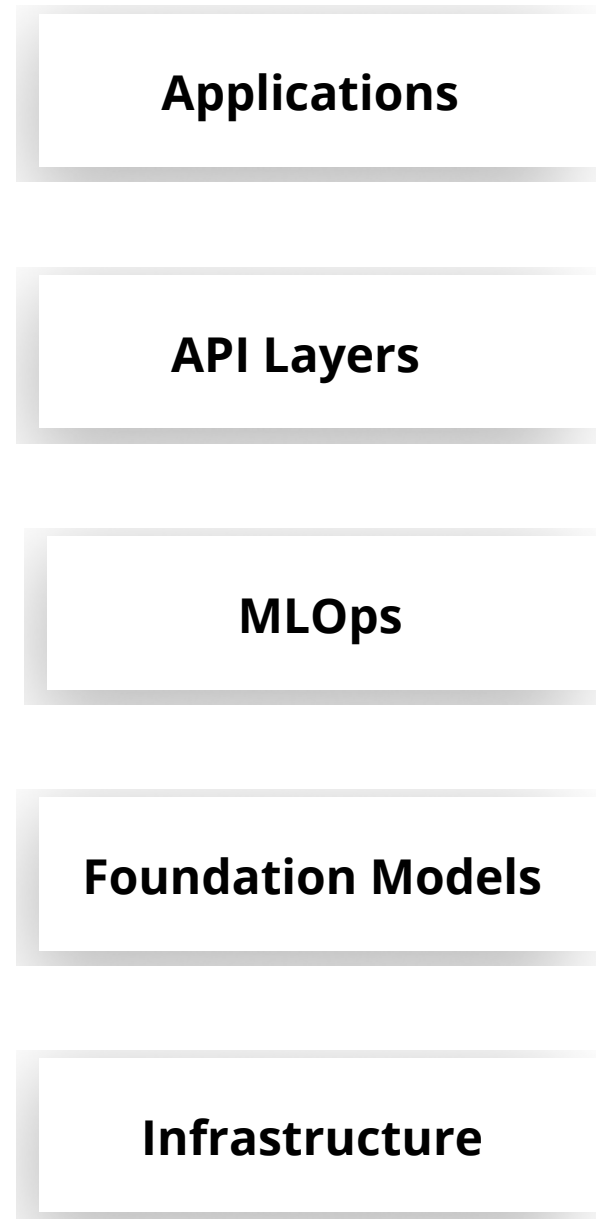
Based on their employment levels in the US in 2021





How Can Businesses Use Generative AI, Today?

Generative AI Tech Stack



Generative Models

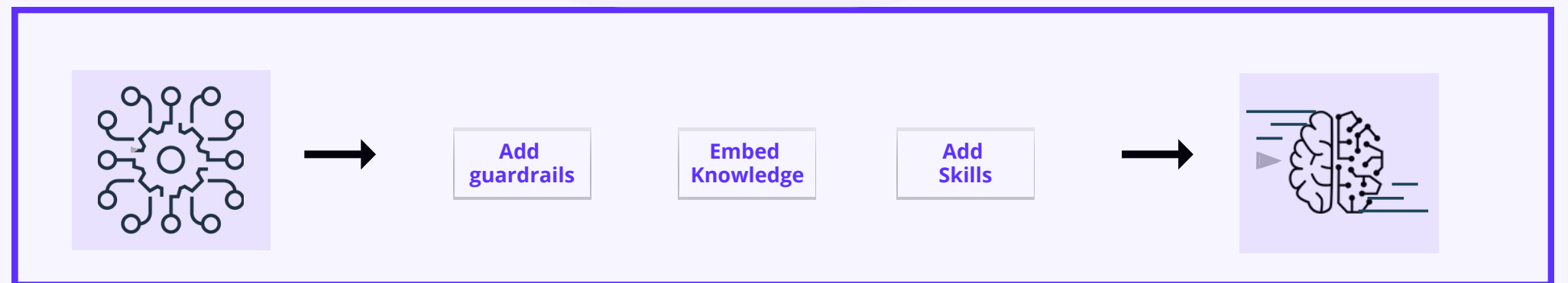
Closed Source

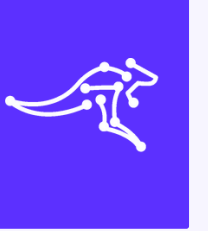
GPT-3.5 DALL.E 2
Codex LaMDA

Open Source

CLIP DALL.E 2
Stable Diffusion BLOOM

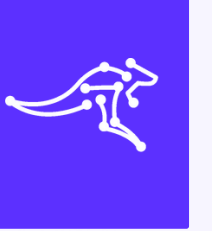
Build Your Own





What are the Options?

	Option	Explanation	Cost
Off the Shelf	<i>Subscribe and Use</i>	Embracing off-the-shelf tools leveraging LLMs that are already available.	Minimal
Customise	<i>Consume with Guardrails</i>	Build Guard Rails by adding pre and post-processing restrictions to off the shelf LLM's	\$
	<i>Augment</i>	Use database lookups to tailor LLMs to an organisation's needs.	\$\$
	<i>Fine Tune</i>	Using fine-tuning to tailor LLMs to an organisation's needs	\$\$\$
Train your own	<i>Build your own</i>	Build and Train your model from scratch with your data	\$\$\$\$



Off the Shelf - Benefits and Limitations

Using paid subscriptions or corporate user plans of Generative AI tools like ChatGPT, Jasper, Notion etc. for trial and training of employees **without exposing confidential company data**. Use-cases limited to the generation of low-quality and low-risk content.



BENEFITS

- ▶ Requires the least LLM training technical skills.
- ▶ Cost limited to subscription fees
- ▶ Fastest turnaround time
- ▶ Cost limited to subscription fees
- ▶ Can leverage the best-performing LLMs in the market



LIMITATIONS

- ▶ Limited to publicly available info
- ▶ Cybersecurity Concerns
- ▶ Fabricated Information.
- ▶ Copyright Issues
- ▶ Data Privacy
- ▶ Deepfakes

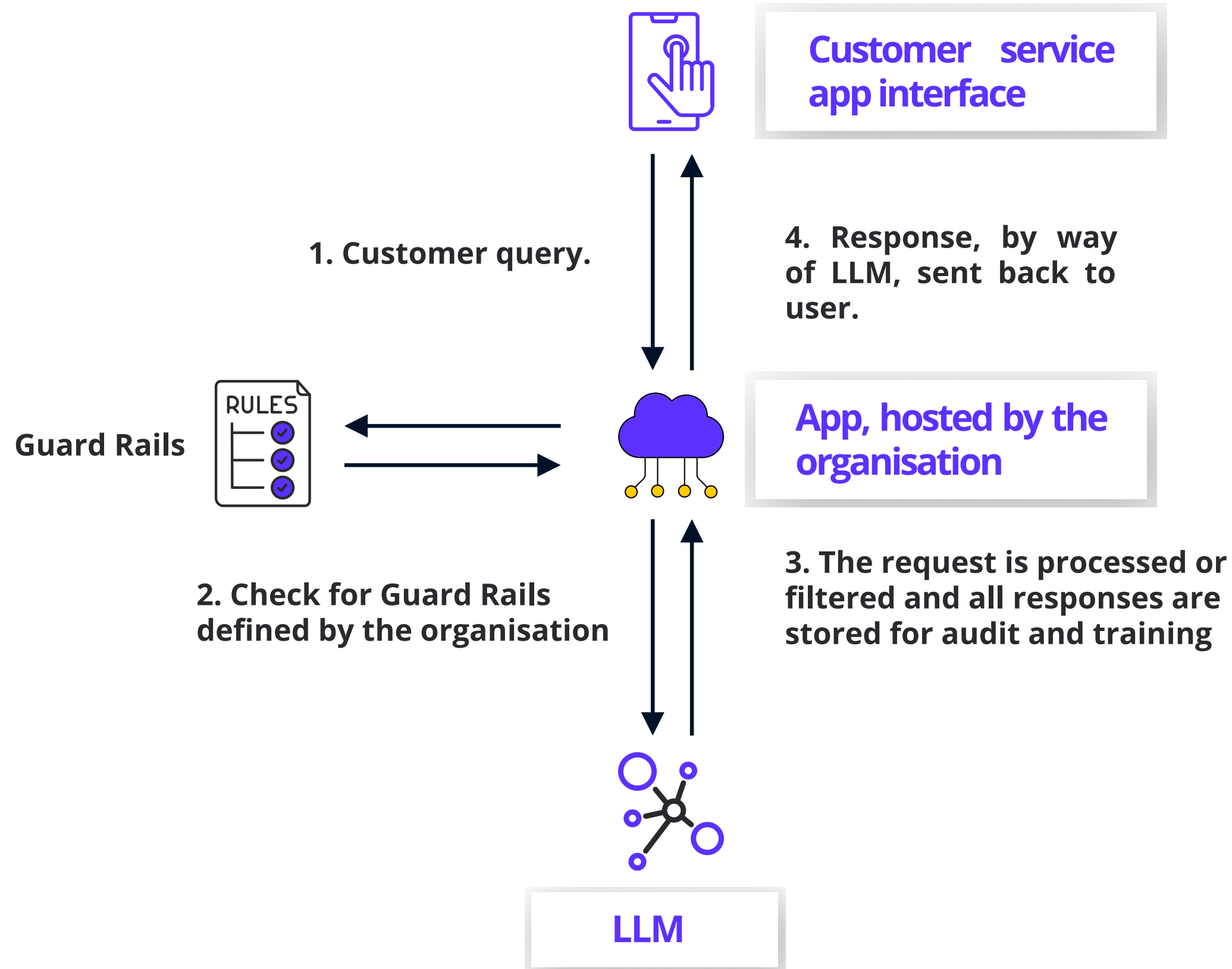


RECOMMENDATION

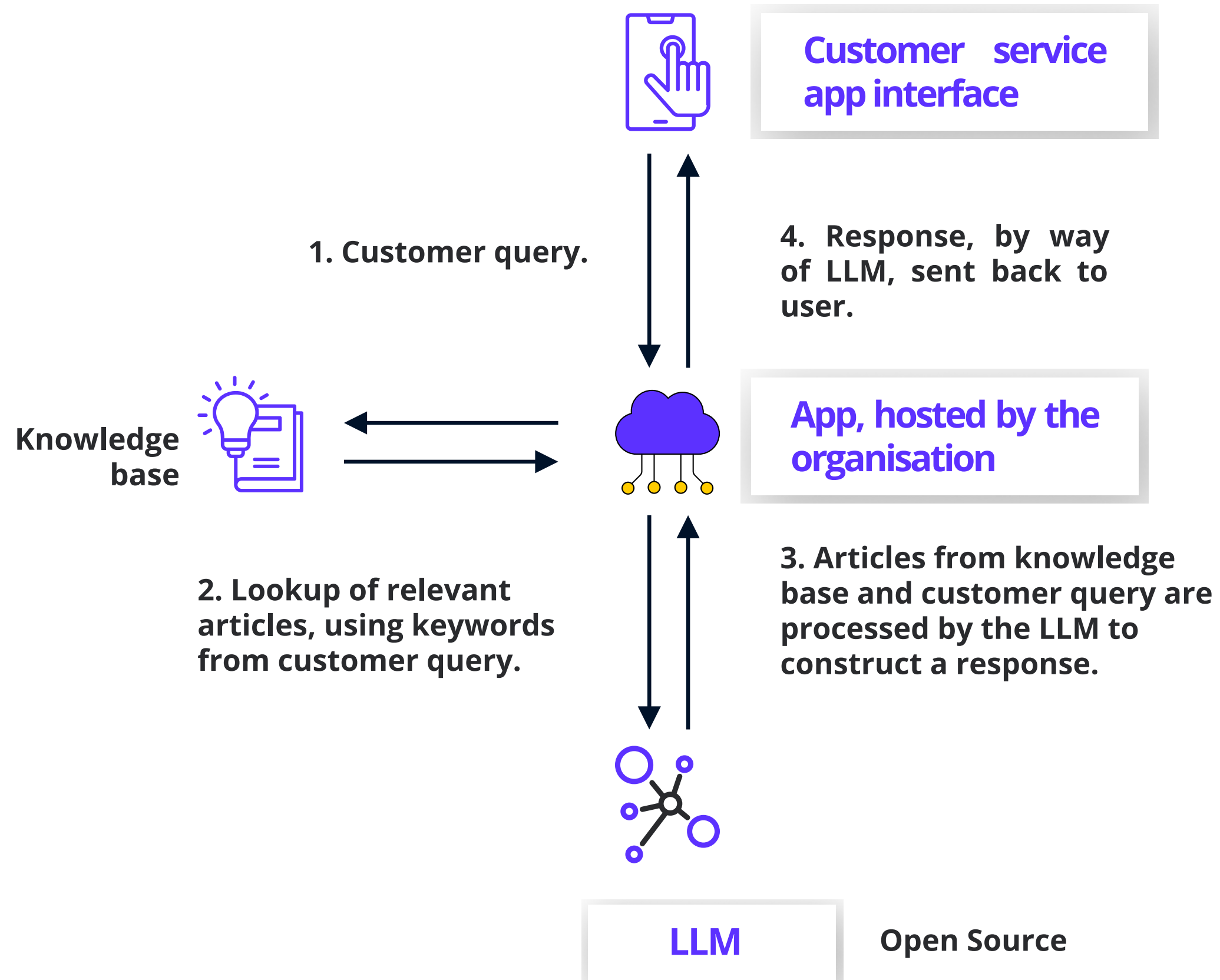
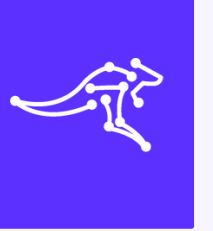
- ▶ Acceptable only for trial and training of employees.
- ▶ Strongly recommend avoiding sharing of any confidential information.
- ▶ Good for prototyping apps and exploring what is possible with LLMs.



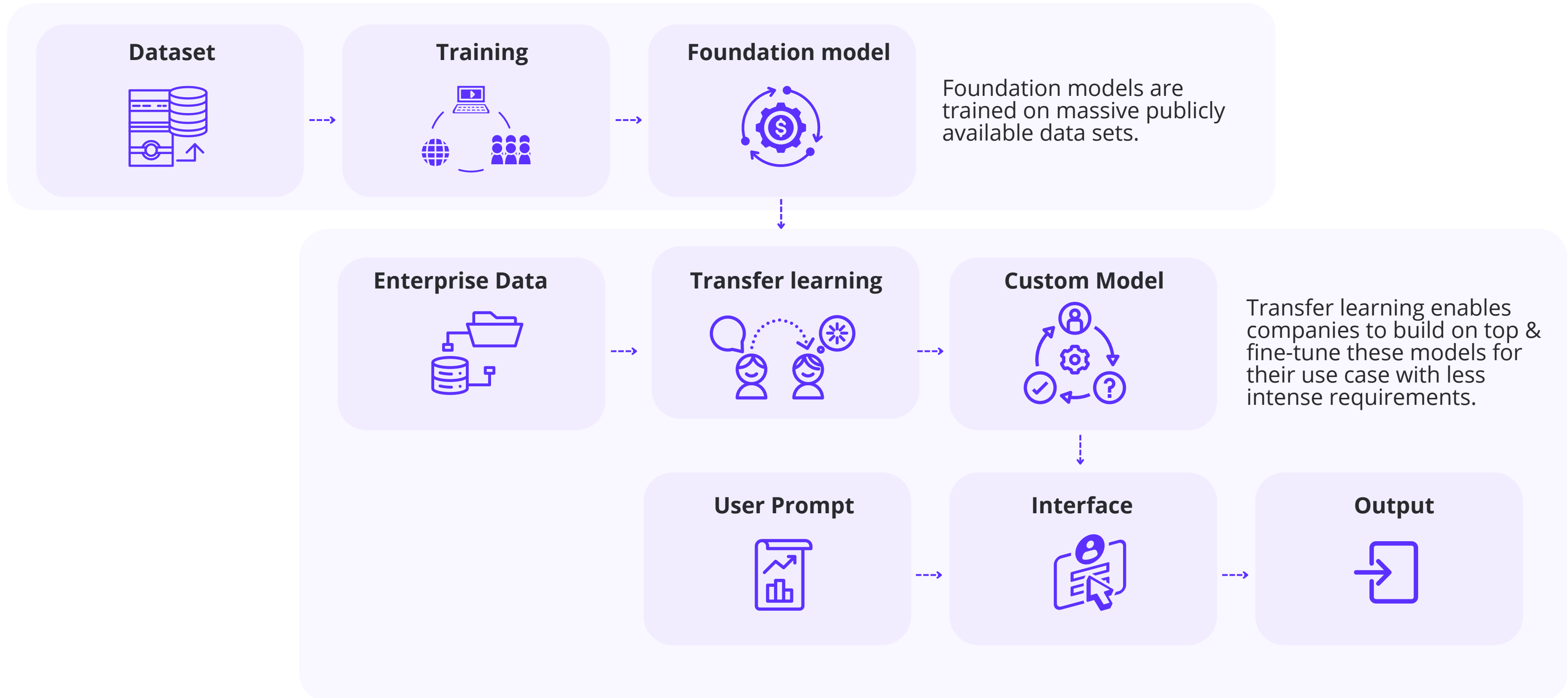
Consume with Guardrails

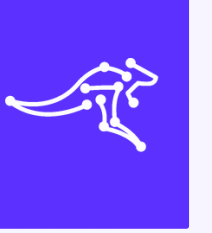


Augment



Fine Tune





Customise - Benefits and Limitations

Organisations can boost the capabilities of their applications by integrating them with LLMs by consuming Generative AI and LLM applications through APIs and tailor them, to a small degree, for your own use cases through prompt engineering techniques such as prompt tuning and prefix learning.



BENEFITS

- ▶ Model trained on organisations data which is publicly not available .
- ▶ More affordable than organisations further training (“fine-tuning”) an LLM
- ▶ Data security as data resides in your own environment.



LIMITATIONS

- ▶ Not appropriate where the model needs to have a wide-ranging understanding of the content in the knowledge base, as only a limited amount of data can be passed to the LLM.
- ▶ The LLM will only use the data passed to it, along with the user’s original query, to construct a response.



RECOMMENDATION

- ▶ An affordable and powerful way to quickly leverage the power of generative Ai for your business
- ▶ An intermediate step for most businesses.



Build your Own

LLM Training Stack

 Hugging Face

 DeepSpeed

 TensorFlow  PyTorch

 RAY

 kubernetes

 NVIDIA



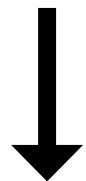
Add guardrails

Embed Knowledge

Add Skills



Your Enterprise Model



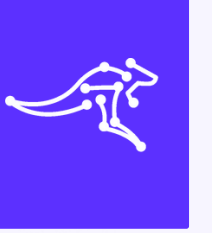
Monitoring

Enterprise Source Systems



Continuously Improve

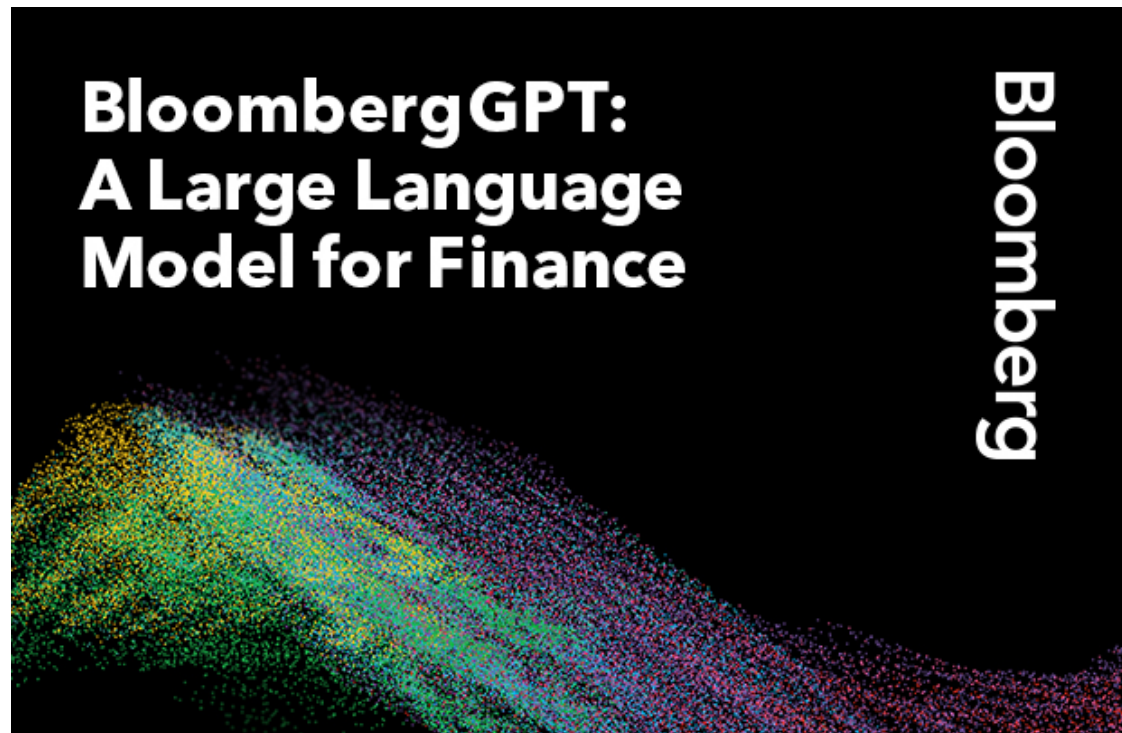




Build your Own - Examples

Example: BloombergGPT

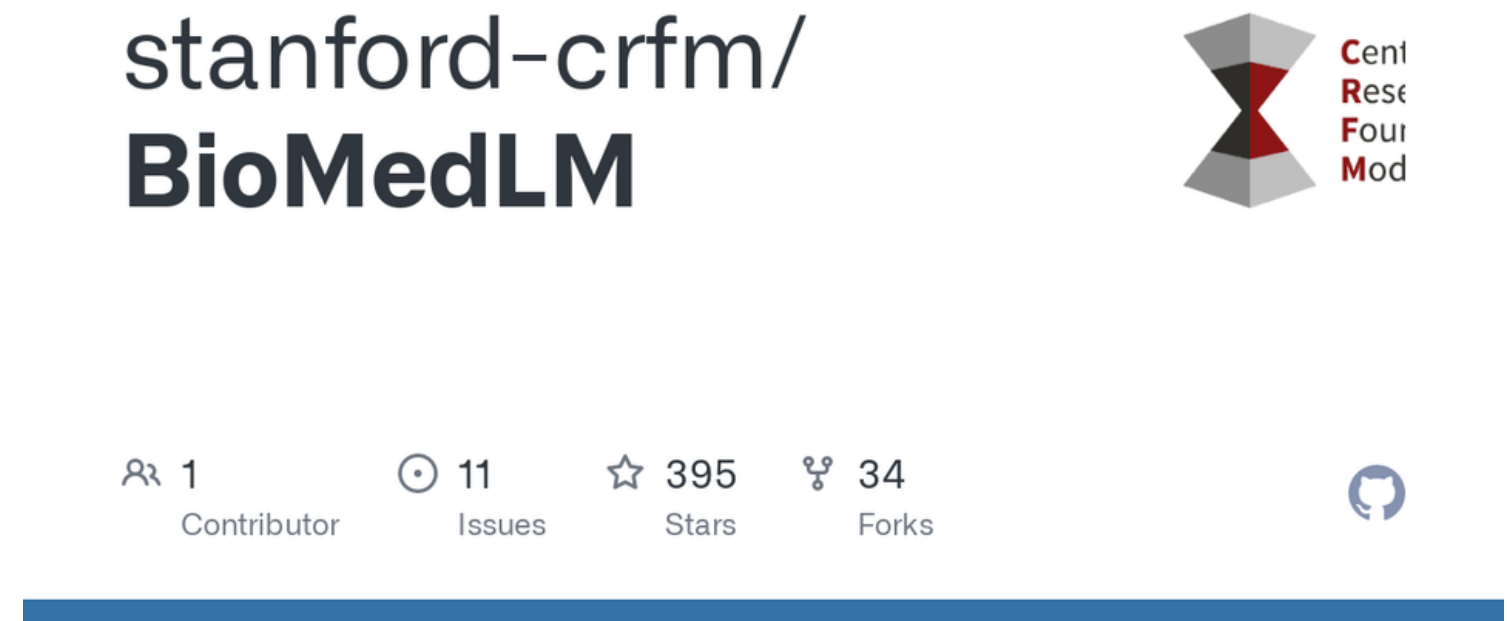
- ▶ Bloomberg trained a 50B LLM on combination of web data + internal Bloomberg data
- ▶ Outperforms existing open source models on financial tasks



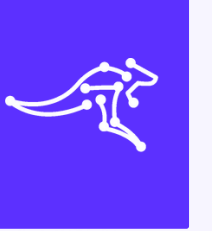
Trained on a combination of web data that's already out there and internal Bloomberg data.

Example: BioMedLM

- ▶ Growing evidence that training domain - specific LLMs (medical, legal, etc) are more accurate than using a generic LLM
- ▶ Example: BioMedLM, a 3B parameter LLM only on PubMed publications



Smaller 3B- 7B but specialised models can have a strong business value.



Build your Own - Benefits and Limitations

Organisations training their own LLM gives them a deep moat: superior LLM performance either across horizontal use cases or tailored to your vertical, allowing you to build a sustainable advantage, especially if you create a positive data/feedback loop with LLM deployments.



BENEFITS

- ▶ Specialised models are smaller and can be deployed on significantly cheaper hardware
- ▶ Specialised models are significantly more accurate for the same resource budget
- ▶ Gain full control of training datasets used for the pre-training,



LIMITATIONS

- ▶ Very expensive endeavor with high risks. Need cross-domain knowledge spanning from NLP/ML, subject matter expertise, software and hardware expertise.
- ▶ Less efficient than Customise option as it leverages existing LLMs, learning from an entire internet's worth of data and can provide a solid starting point



RECOMMENDATION

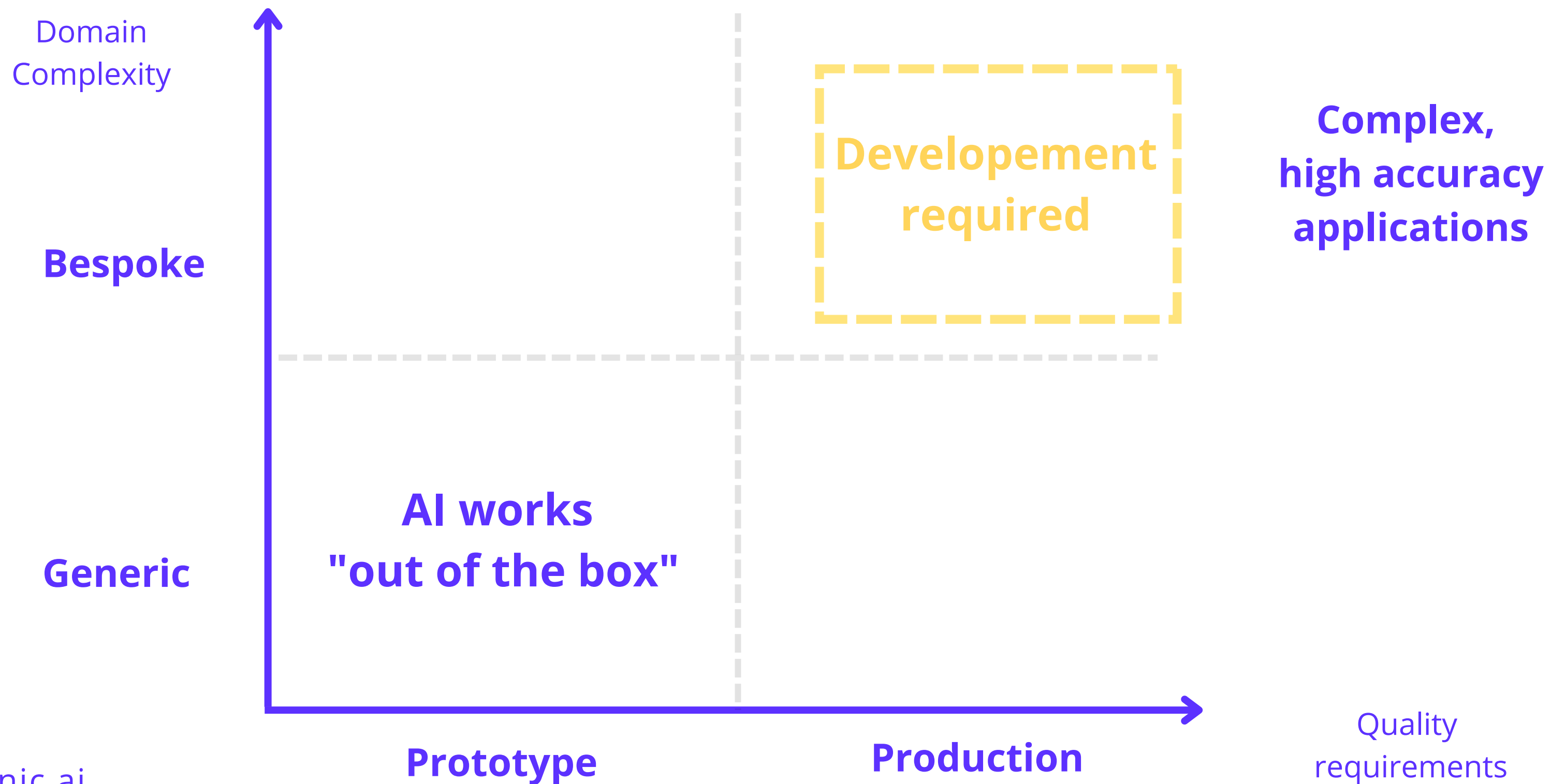
- ▶ Best if you need to change model architecture or training dataset from existing pre-trained LLMs.
- ▶ Typically, you have or will have lots of proprietary data associated with your LLM to create a continuous model improvement loop for sustainable competitive advantage



| What is the role of MLOps?







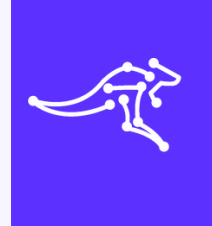
AI Does not work out of the Box









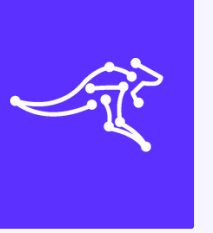
Complex Use Cases Requires Significant Development

	 GPT-4	 BERT	 Clip	 GPT-4
Domain	Fortune 500 pharma	Top US bank	Global ecommerce	Legal data case study
Use case	Information extraction	Chat intent classification	Image classification	Document classification
Foundation model performance	60%	60%	43%	59%

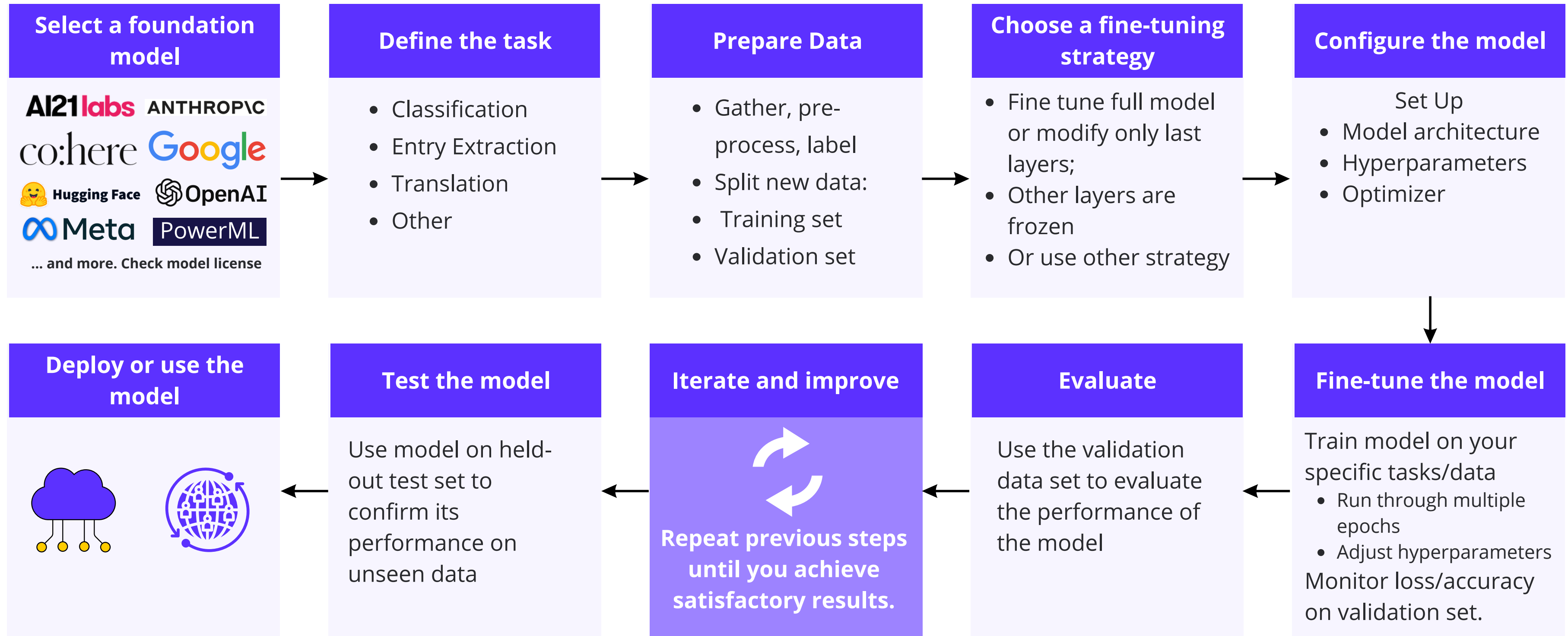


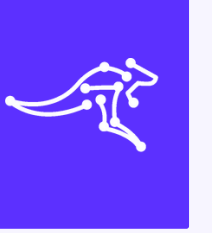
Complex Use Cases Requires Significant Development

	 GPT-4	 BERT	 Clip	 GPT-4
Domain	Fortune 500 pharma	Top US bank	Global ecommerce	Legal data case study
Use case	Information extraction	Chat intent classification	Image classification	Document classification
Foundation model performance	60%	60%	43%	59%
Fine Tuned model performance	86%	85%	71%	83%*



Key Steps To Derive Value Out Of Generative AI

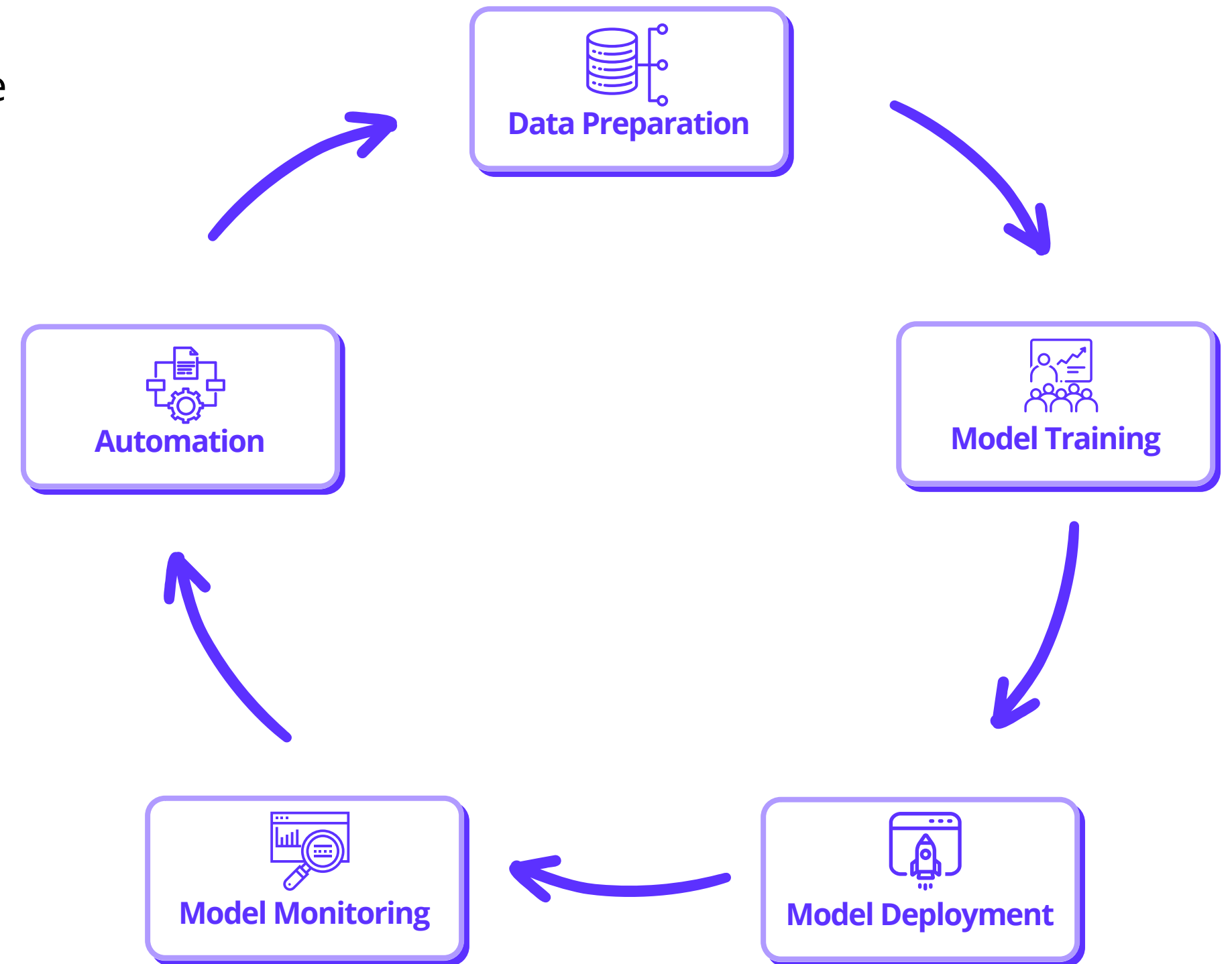




Role Of MLOps

- ▶ **Data Preparation:** Create training data & continuously update data.
- ▶ **Model Training:** Feed the data into a model for training
- ▶ **Model Deployment:** Deploy trained models into production (live).
- ▶ **Model Monitoring:** Monitor models for performance, accuracy, data sways or data drifts.
- ▶ **Automation:** Automate model for retraining, version control, rollback or update basis performance.

An MLOps platform allows you to manage this complete process end to end with high accuracy, reliability and efficiency

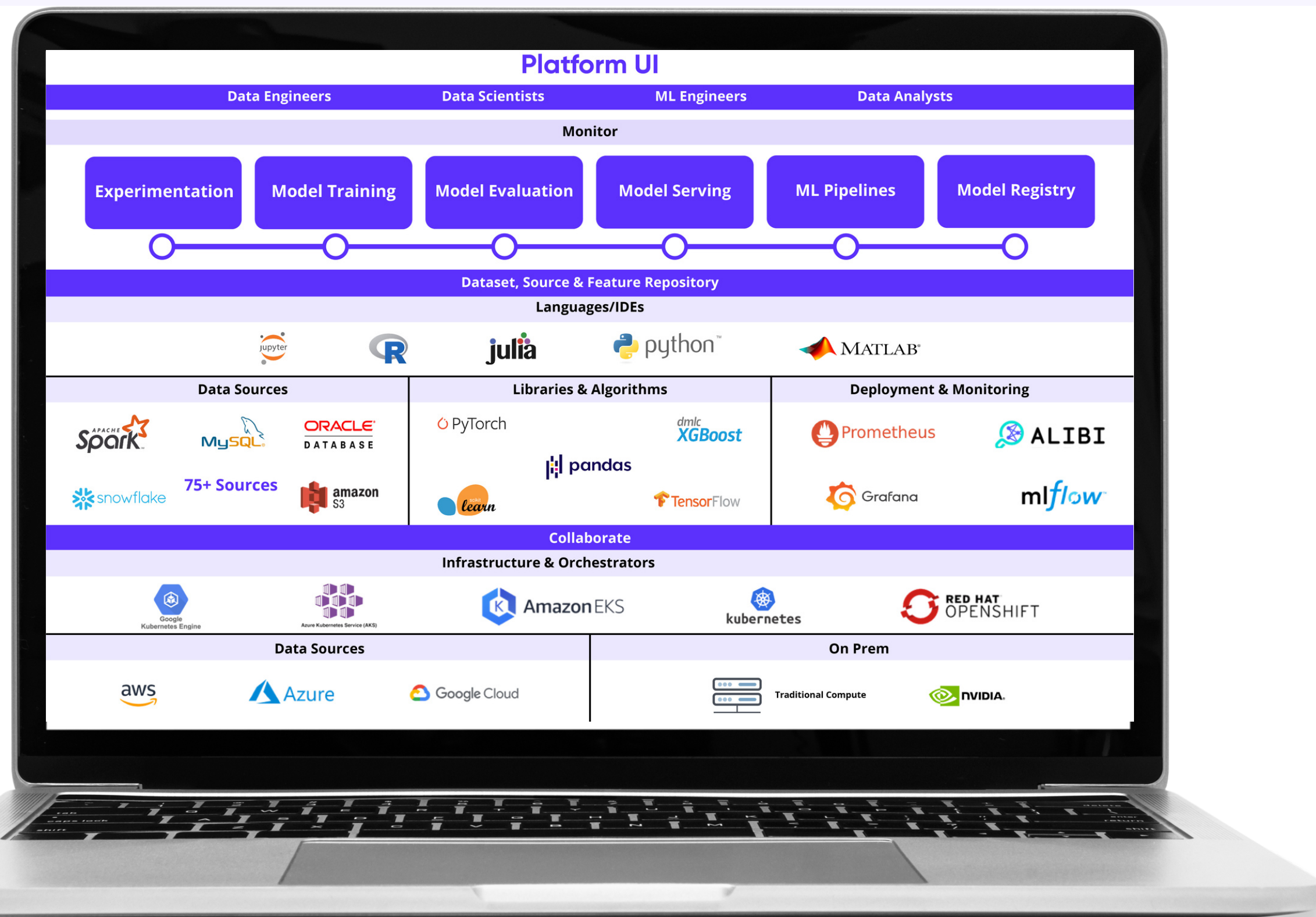




| How can Katomic Help?



Unified Platform powering your Generative AI strategy



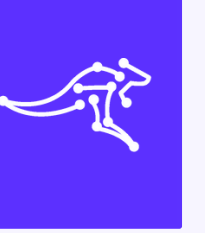
Creative scientific process of data scientists

+

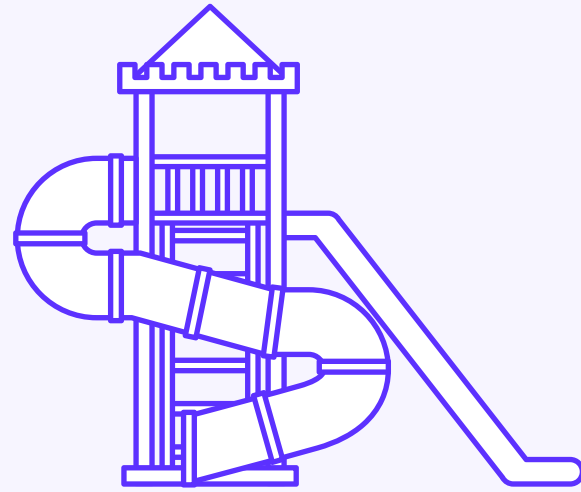
Professional software engineering process

=

Releasing ML Models into production safely, quickly, and in a sustainable way.

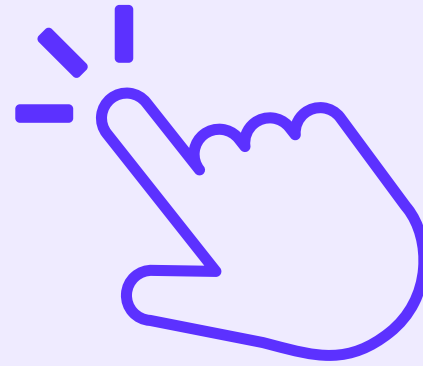


How Katomic can help accelerate your LLM Journey



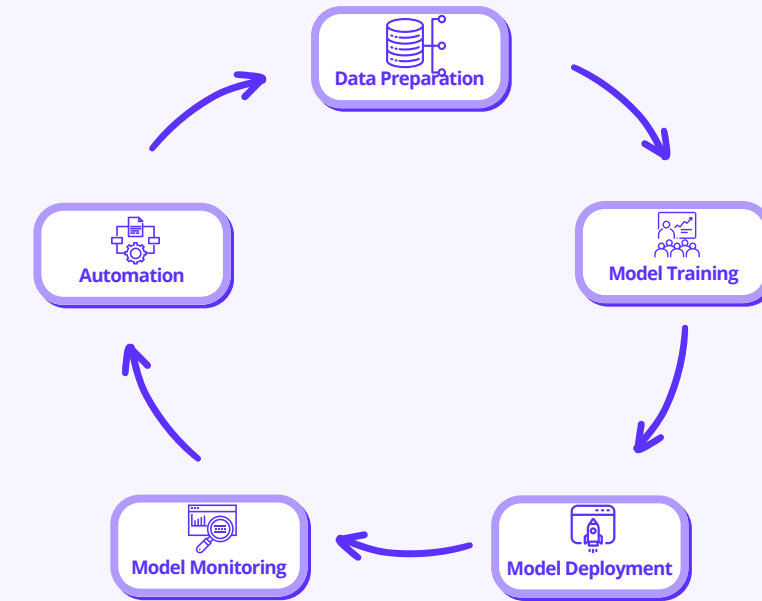
Katomic LLM Playground

Test best-in-class foundation models for your business and your specific data to build sustainability. Katomic has curated over 70 + popular LLMs for you to effortlessly experiment and prototype flows with drag-and-drop components.



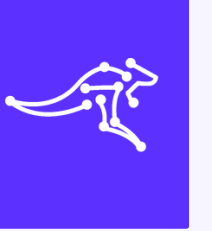
Deploy with few Clicks

Katomic partners or integrates with all of the leading AI models, from open-source to closed-source. Deploy from our curated list of LLMs or any open-source LLM in a few clicks securely on your infrastructure



One Unified Platform

Develop, test, and productionize - all in one platform. The only full-stack platform for powering your Generative AI strategy—including Data preparation, Distributed training, fine-tuning, security, model safety, model evaluation, and enterprise apps.



Katomic Advantage



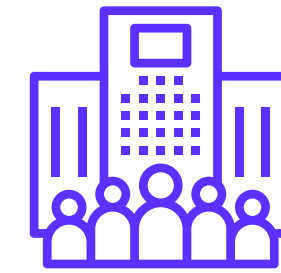
Achieve Higher Accuracy

Easily parallelize and distribute workloads across multiple nodes and GPUs.



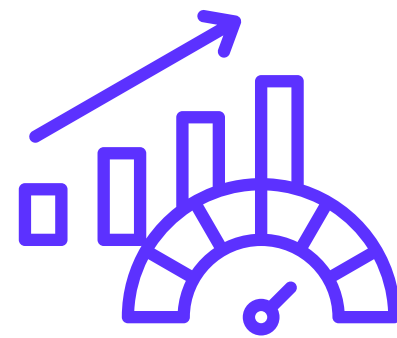
Run Anywhere

Public cloud, private data centres, bare metal, Kubernetes cluster — Katomic runs anywhere



Enterprise Support

Fully Supported by Katomic Experts every step of the way.



Fastest Performance at Scale

Bundled with State of the art tools and techniques



Ease of Use

Easily access the capabilities of your custom LLM through Intuitive UI.



Thank You

Automate your **cycle of intelligence**