

ORACLE

ORACLE

Bringing Generative AI to the Enterprise

Name

Title

Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Recent AI
breakthroughs will
further accelerate
growth

\$16-17bn NLP market size in 2022

\$49bn forecasted by 2027

\$1.6tr AI Global market size forecasted
by 2023

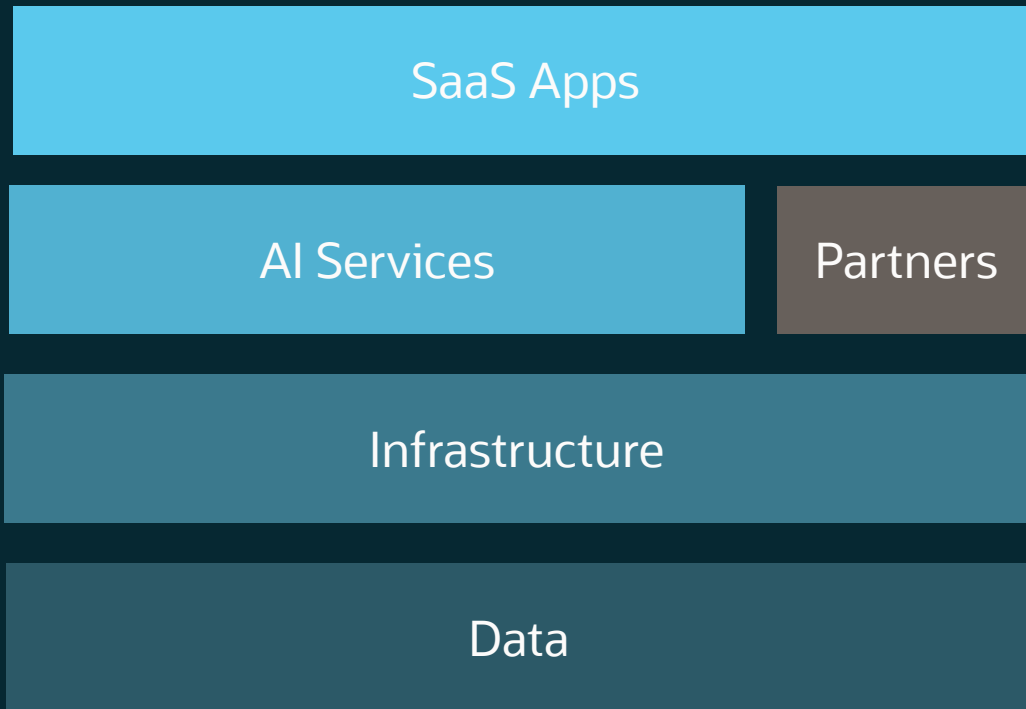
10% of worldwide code and tests
projected to be written by TuringBots by
2023

Some vendors are trying to move quickly ...

	Launched Azure Open AI service. Integrated ChatGPT into Bing, Office stack (Microsoft 365 copilot) and automatic code generation into Github copilot. \$10bn investment in Open AI.
	Google announced access to its PaLM model through Vertex AI (March 2023). Announced MakerSuite to make it easier to tune models. Announced Google Bard (a ChatGPT competitor). Invested \$300mm in Anthropic.
	Announced AWS Bedrock (Apr 2023) exposing models from AI21, Anthropic, Stability AI. Also exposes Amazon's own foundational models.
	Salesforce announced Einstein GPT (March 2023), using OpenAI for CRM.

But no one is addressing the **specific** end-to-end needs of **enterprise** customers with generative AI

AI for the enterprise



Oracle has been focusing on bringing AI to the enterprise at every layer of our stack

Extensive investment from infrastructure to SaaS apps

Generative AI and massive scale models are the more recent step – our focus and emphasis has not changed

Oracle and NVIDIA enable generative AI



- Oracle's AI Infrastructure with NVIDIA GPUs, including the A100 Tensor Core, support numerous LLM startups
- NVIDIA uses OCI for their AI infrastructure needs; Oracle delivered over 4000 GPUs for a single cluster

"Our partnership with Oracle will put NVIDIA AI within easy reach for thousands of companies."

Jensen Huang
CEO, NVIDIA



OCI Supercluster

OCI cluster networking with 1.6 Tb/s bandwidth, at microseconds of latency

Cluster Network for NVIDIA A100 GPU

Each GPU has its own RDMA connection

- 200Gb/s per connection
- 1600Gb/s bandwidth per server
- Only few microseconds of latency between nodes

Cluster networking powers OCI super clusters

- 40GB can scale to 512 GPUs
- 80GB can scale up to 32K GPUs

OCI Supercluster™

- A100 40GB: Up to 64 nodes
- A100 80GB: Up to 4096 nodes



Introducing the latest partner: Cohere

The Oracle + Cohere strategy is simple:

*Deliver the most impactful and intelligent generative solutions for your business. Oracle has partnered with leading generative AI company Cohere to create the industry's best generative AI solutions for the **enterprise**.*



Why Cohere?

- Cohere is a leading independent provider of LLMs
- Founded by one of the co-authors (CEO Aidan Gomez) of the famous "Attention is all you need" paper, which laid the foundation for generative AI
- Cohere's focus is to serve enterprises at scale. Cohere's GTM approach aligns well with Oracle's vision and customer profile

"Our chat models are focused more on business applicable tasks like answering questions than writing poems."

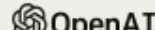
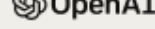
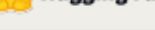
- Cohere CEO Aidan Gomez

Efficient, customizable AI—to be offered through Oracle

Cohere delivers top-tier LLM performance, outperforming peers in independent LLM benchmarks

(not included: GPT-4 from OpenAI)

Cohere's Command model is ranked very highly in HELM, **while being more efficient** (52B parameters compared to GPT-3 with 175B parameters) and **more easily customized**.

Company	Model	Model type	Mean win rate
 cohere	Cohere Command (52B)	Command	93.0%
 OpenAI	Davinci Instruct 002	Command	93.0%
 OpenAI	Davinci Instruct 003	Command	89.8%
	TNLG v2 (530B) <i>not publicly available or viable to serve given size</i>	Base	85.5%
ANTHROPIC	Anthropic v4 (52B)	Command	84.2%
 AI21labs	J1 Grande v2 (17B)	Command	80.6%
 ALEPH ALPHA	Luminous Supreme (70B)	Command	78.3%
 cohere	Cohere XL (52B)	Base	74%
 Meta	OPT (175B)	Base	67.8%
 OpenAI	GPT-3 Davinci (175B)	Base	62.8%
 AI21labs	J1-Jumbo (178B)	Base	59.2%
 ALEPH ALPHA	Luminous Extended (30B)	Command	58.2%
 Hugging Face	BLOOM (176B)	Base	52.9%

Three ways to transform the generative AI for business experience



SaaS: Oracle applications

Oracle will embed LLMs across Oracle's cloud applications, industry applications, and database portfolio.



PaaS: OCI Generative AI service

A new AI service through Oracle (with Cohere's LLMs) will enable OCI customers to custom train models and add generative AI capabilities to their own applications.



IaaS: GPU Infrastructure

LLMs, including those OCI created by Cohere, require large-scale AI Infrastructure on to build, adapt, and deploy its generative AI models in a cost-effective manner.

The Oracle AI Stack

Business applications, Oracle SaaS portfolio

ORACLE
Cerner

ORACLE
NETSUITE

ORACLE
Aconex

...

AI services



Digital Assistant



Speech



Language



Vision



Document
Understanding



Anomaly Detection



Forecasting

Machine learning services



OCI Data Science



ML in Oracle Database



MySQL Heatwave AutoML



Data Labeling

Data

AI infrastructure



Compute bare metal instances and VMs



Cluster networking



Block, object, and file storage; HPC filesystems

Coming soon: OCI Generative AI

State-of-the-art foundational models

We have partnered with Cohere to bring SOTA foundational models to both OCI customers and Oracle SaaS customers.

Fine-tuning support

Tune custom models with your own data to solve specific problems or target specific domains.

Fully hosted on OCI

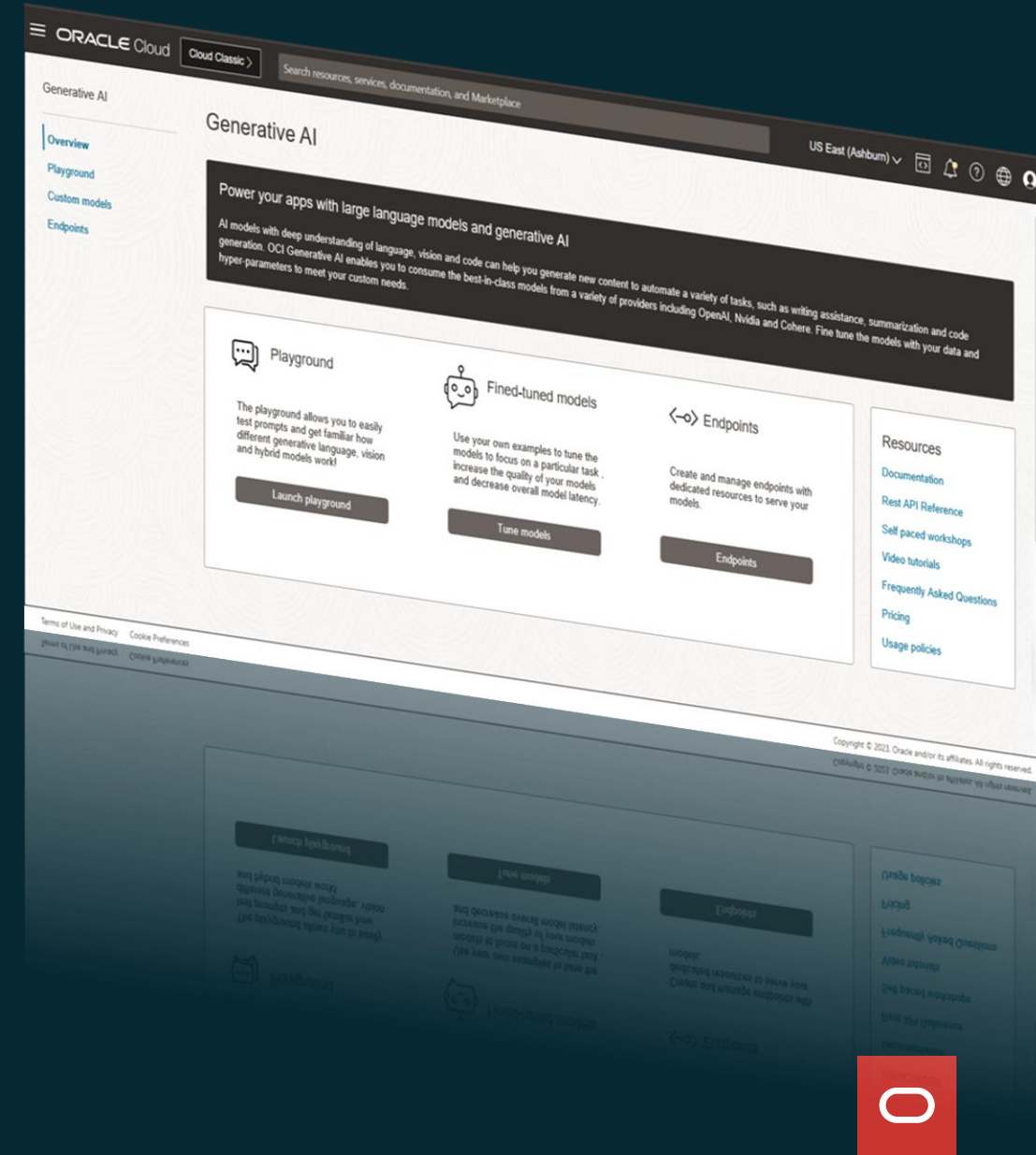
All processing and data storage happens inside OCI, no cross-region or cross-cloud communication.

Respect for customer privacy

Customer training and inference data is owned by the customer and only used to serve the customer.

PAYG and single tenant, dedicated hardware

Pay per token or host the models on dedicated hardware for stable performance.



Oracle customers for generative AI



“Generative AI is an amazing opportunity to support our business, delight our customers, and boost productivity. We’re excited to partner with Oracle and Cohere to achieve those goals together.”

- Kamran Zargahi, Senior Director



“Oracle has been a valuable partner for Altair and with Altair’s focus on computational science and artificial intelligence (AI), Altair has begun integrating large language models (LLM) into our solutions because of their transformative promise. We believe generative AI will transform the way we work.”

- Sam Mahalingam, CTO



“University of Missouri Health Care is looking forward to simplifying physician and nurse workflows by incorporating generative AI into their Oracle electronic health record system, which will help them spend more time with patients to improve the quality of care and healthcare experience.”

- Richard J. Barohn, MD, Executive Vice Chancellor

Guiding principles for successful generative AI

...more than just the models



**Data privacy, security
& governance**



**Powerful and high-
performing models**



**Embedded
generative AI**



**Generative AI where
needed**

With its foundations in enterprise data management, industry data, AI infrastructure, and business-critical apps, Oracle is the best positioned company for delivering enterprise-grade generative AI for any company.



Oracle AI's commitment to data privacy and security

With decades of experience helping organizations secure mission-critical data, Oracle understands the sensitivity and criticality of enterprise data and is developing generative AI solutions aligned with the most stringent privacy and security requirements.



Unrivaled data security, privacy, and governance

OCI ensures customers have complete control and ownership of their data

Maintain superior data security and privacy

Customer data must never be used by Oracle to train public models, to preserve security and privacy.

Gain complete control of data

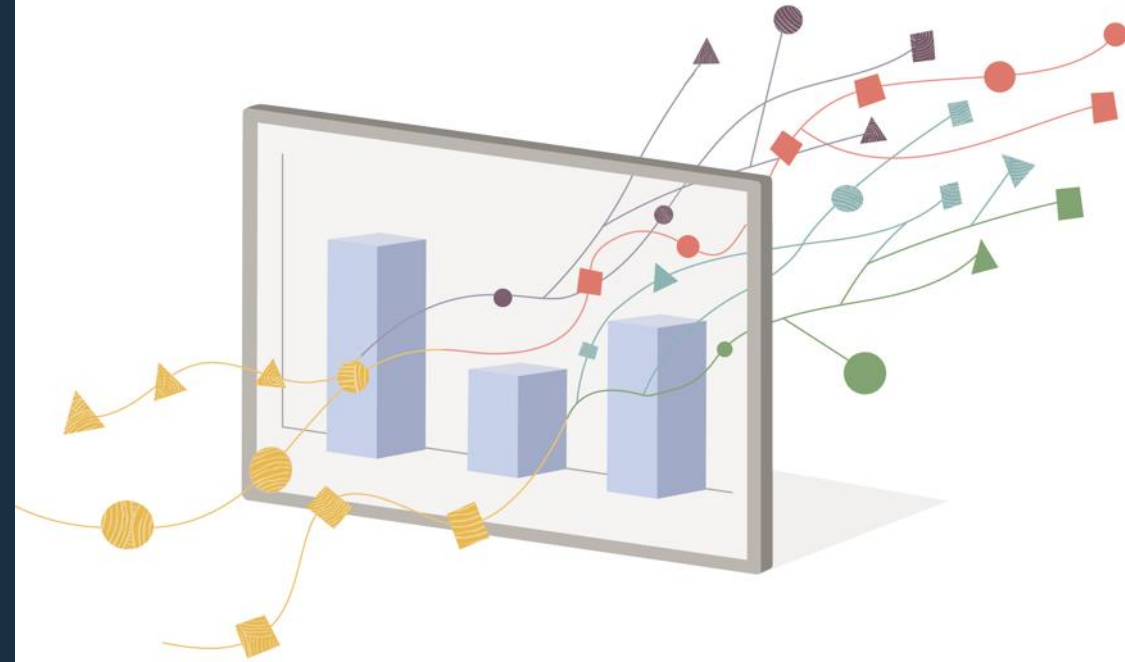
Customer data stays in OCI, is entirely owned by the customer, and is only used for custom training their private models.

Track data provenance and lineage

To validate the accuracy of the data, mappings, and transformations, tools for data provenance and lineage will be available.

Enterprise-first AI: Models that understand your business

- Oracle's industry knowledge enables model training that provides a high degree of accuracy for specific use cases
- Oracle stores the world's valuable data. Custom train high-performant, domain-specific models with data already on OCI's leading modern data platform



Flexible Generative AI training options

Bring your own model (BYOM)

Bring your own large language models to OCI

Train foundational models

Train generative language and vision models from scratch on OCI infrastructure

Pre-built domain specialized models

A host of vertical domain specialized models (i.e., customer support, healthcare, finance) to meet your needs more quickly.

Generative AI is coming soon to Oracle's business apps ...



ERP



SCM



HCM



Sales



Sales



Service



Marketing



OCM



Cerner



NetSuite

... and Oracle's industry-specific apps too



Automotive



Communications



Construction and
engineering



Consumer goods



Energy and water



Financial services



Food and beverage



Government and education



Healthcare



High technology



Hospitality



Industrial manufacturing



Life sciences



Media and entertainment



Oil and gas



Professional services

And more



Available where customers need it

- Leverage public cloud to scale solutions on demand, customize models, and create private model endpoints for their business.
- Or, access generative AI through your own data center with Oracle to combine generative capabilities with on-premises data and applications

Oracle's generative AI advantage

Unrivalled for data security, privacy, and governance

- ✓ Oracle's generative AI offerings won't store customer data, nor mix customer data.
- ✓ Tools for accessing data provenance and lineage will be available as well.

Powerful and high-performing models

- ✓ Use your industry-specific data to custom train generative AI models for increased accuracy, on top of Oracle's enterprise-grade data management.
- ✓ Cohere generative AI models are state-of-the-art but still efficient enough to be highly customizable, for domain adaption and full fine-tuning

Embedded generative AI services

- ✓ Offer generative AI capabilities where end users work (Fusion HCM, ERP, Netsuite, Cerner, and others)
- ✓ Mask the complexity and detail of using these models to power easy and innovative user experiences

Generative AI where customers need it

- ✓ Leverage public cloud with OCI, or take advantage of private cloud with Cloud@Customer

The Oracle difference in Generative AI



The best models based on unique data and industry knowledge.



Tailored to your data. Oracle's Gen AI service allows you to refine the models using your own data



Deploy your models securely and privately on dedicated, single-tenant hardware for predictable performance and pricing



Generative AI where you need it. Use GenAI on OCI commercial regions, scale your solution on demand, or deploy in your own private cloud in Cloud@Customer.

Thank you



ORACLE



Oracle customers for generative AI, continued



“We are excited about the potential of combining our Oracle and other enterprise data with the power of large language model AI to accelerate numerous workflows and improve access to our enterprise project knowledge; these approaches have potential to unlock tremendous value in the AEC [architecture, engineering and construction] industry.”

- Nitesh Alagh, Business Lead - Digital Engineering, Sustainability & Emerging Tech



“Through the thoughtful development of impact-driven generative AI applications clinicians will be empowered to maximize their potential, increase productivity, and gain access to deep insights transforming the way clinical decisions are made. The partnership between Oracle and Cohere will help us bring these innovations into Ronin’s clinical intelligence platform with increased velocity and performance.

- Christine Swisher, Chief Scientific Officer



Elevating customer support and services with the language comprehension and Generative AI capabilities of large language models is a top priority for NetSuite. [Oracle|Cohere]’s models will streamline customer communications, provide customer solutions, and deliver exceptional customer experiences. With [GenAI], we will be able to revolutionize our most important customer touch points.

- May Lam, Director, Machine Learning Program, Product Management