The Dawn of the Agentic Era

Satinder Singh
Director, Tech- AWS India

The next big thing is the one that makes the last big thing more usable

- Blake Ross (Co-creator of Mozilla Firefox

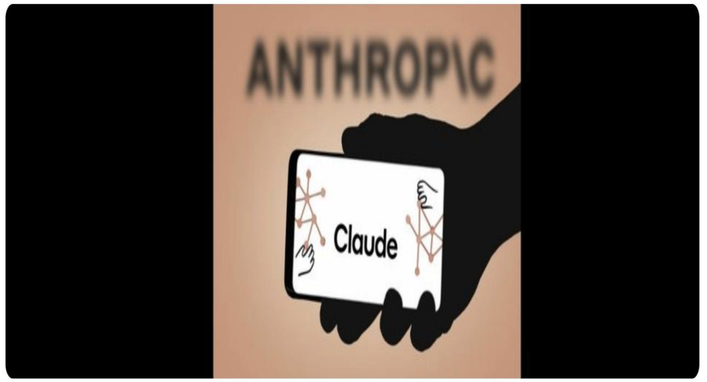
Building a C compiler with a team of parallel Claudes

Published Feb 05, 2026

We tasked Opus 4.6 using agent teams to build a C Compiler, and then (mostly) walked away. Here's what it taught us about the future of autonomous software development.

Anthropic unveils Claude legal plugin and causes market meltdown

February 3, 2026



Gen AI | Knowledge Management | Latest News

Generative AI vendor Anthropic has unveiled a legal plugin that helps customise its large language model Claude for legal tasks such as document review, sending public legal software stocks into an ensuing spin today (3 February).

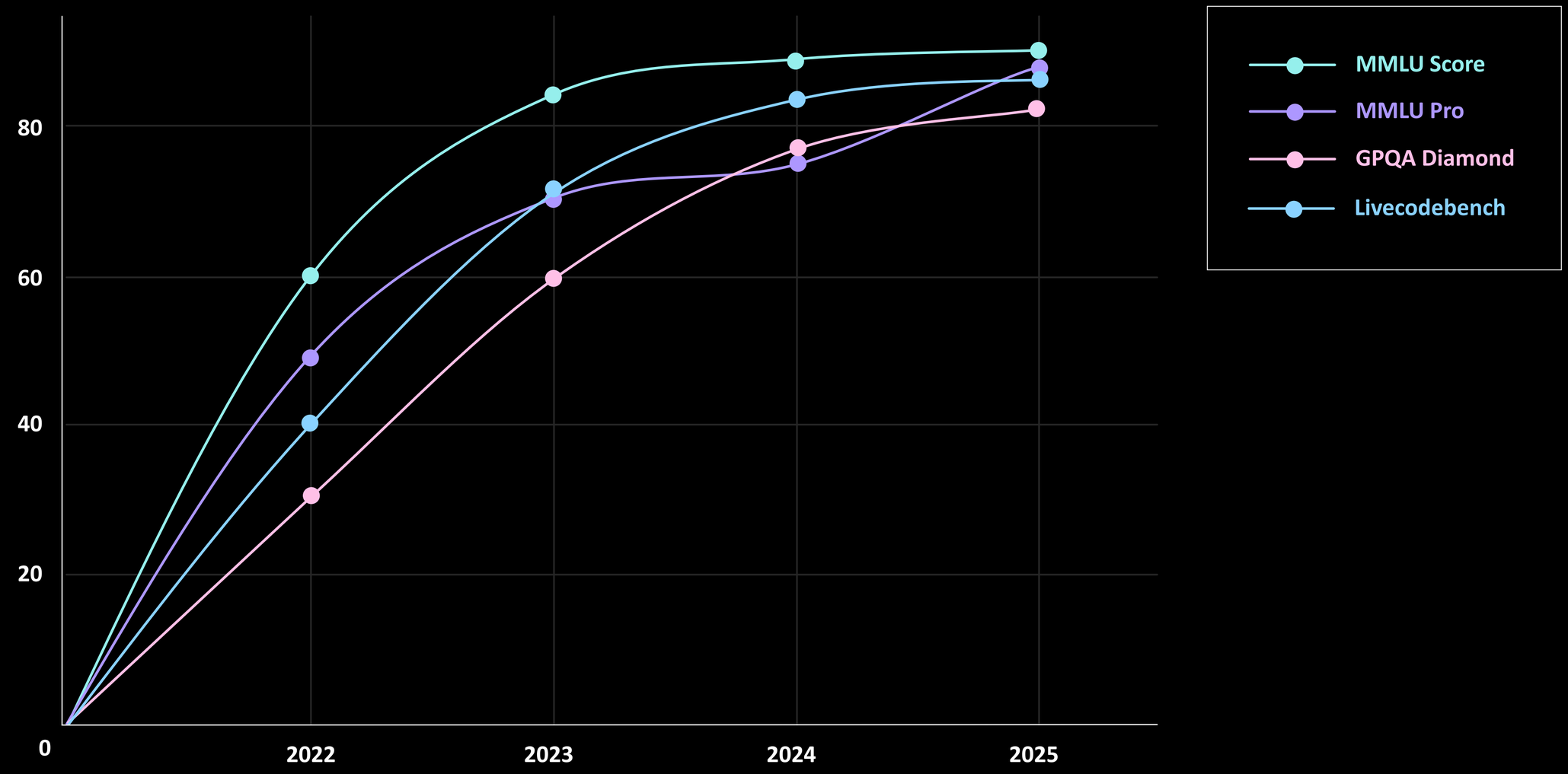
Cyber Stocks Slide as Anthropic Unveils Claude Security Tool

By Ryan Vlastelica

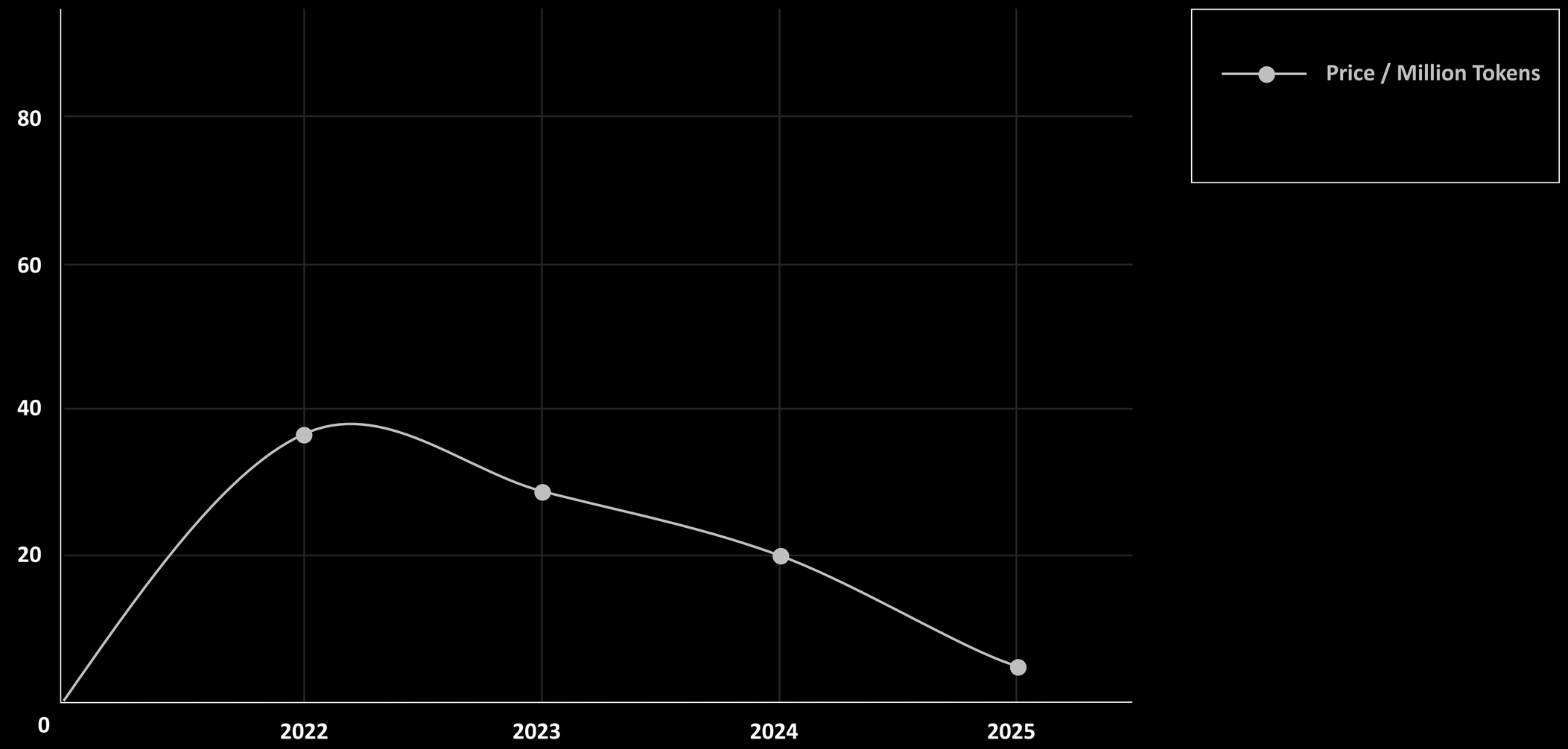
February 21, 2026 at 12:13 AM GMT+5:30
Updated on February 21, 2026 at 2:52 AM GMT+5:30

Save | Translate | Listen 2:29

Reasoning – A Programmable Resource

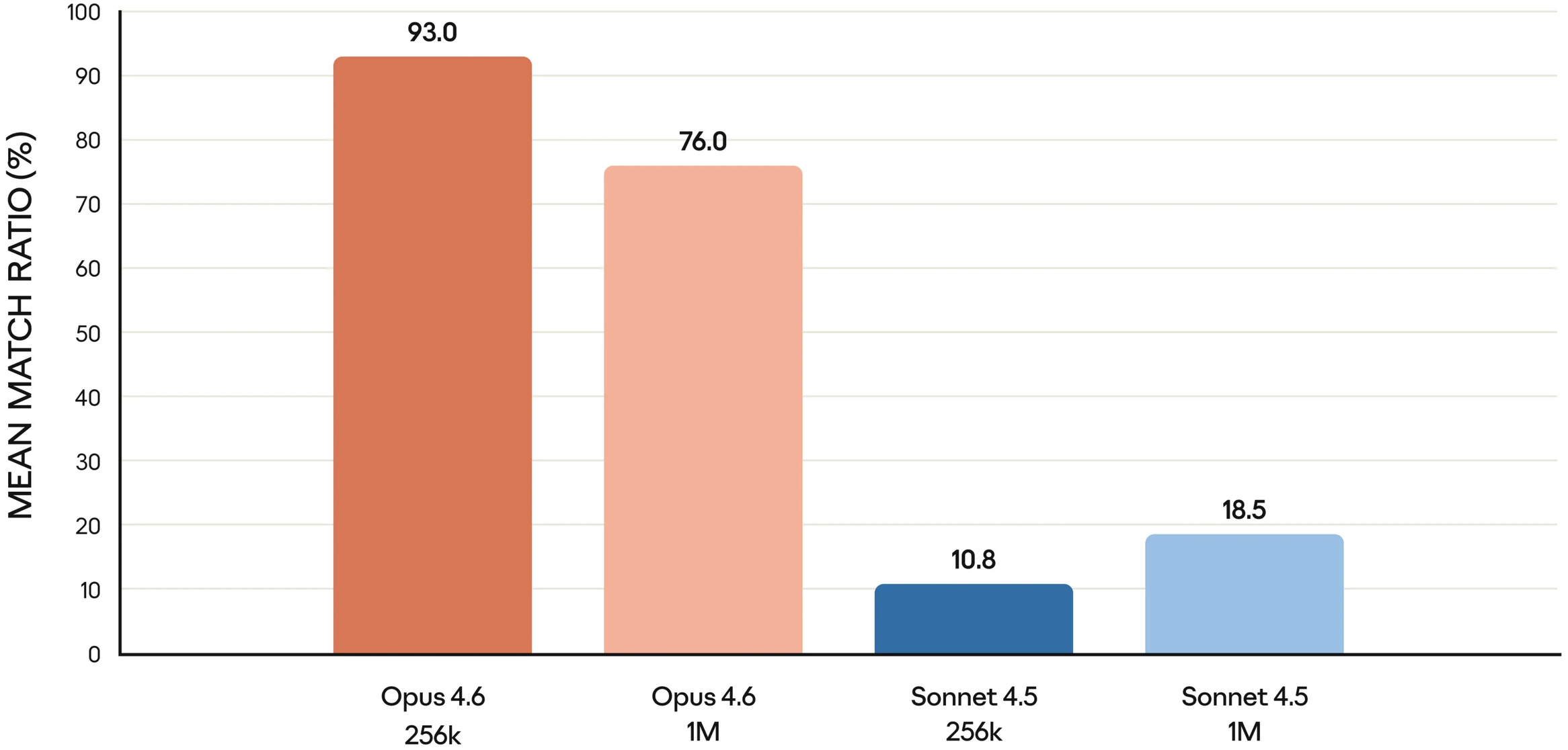


Reasoning – A Programmable Resource



Long-context retrieval

MRCR v2 (8-needle)

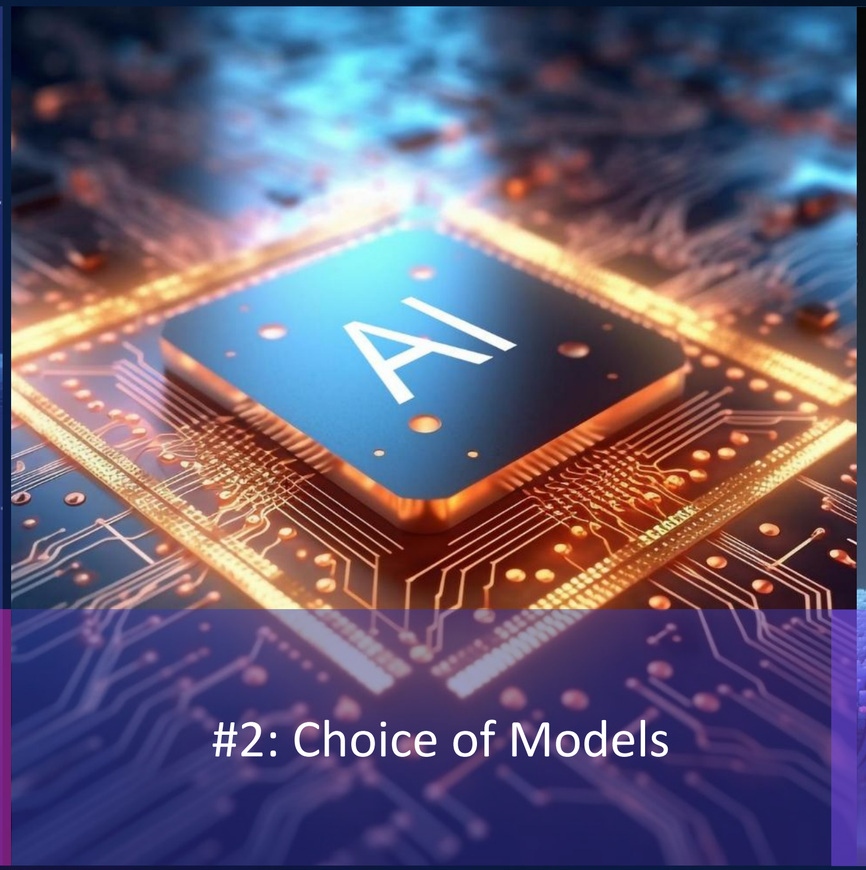



MCP: The USB-C Port
for AI !!

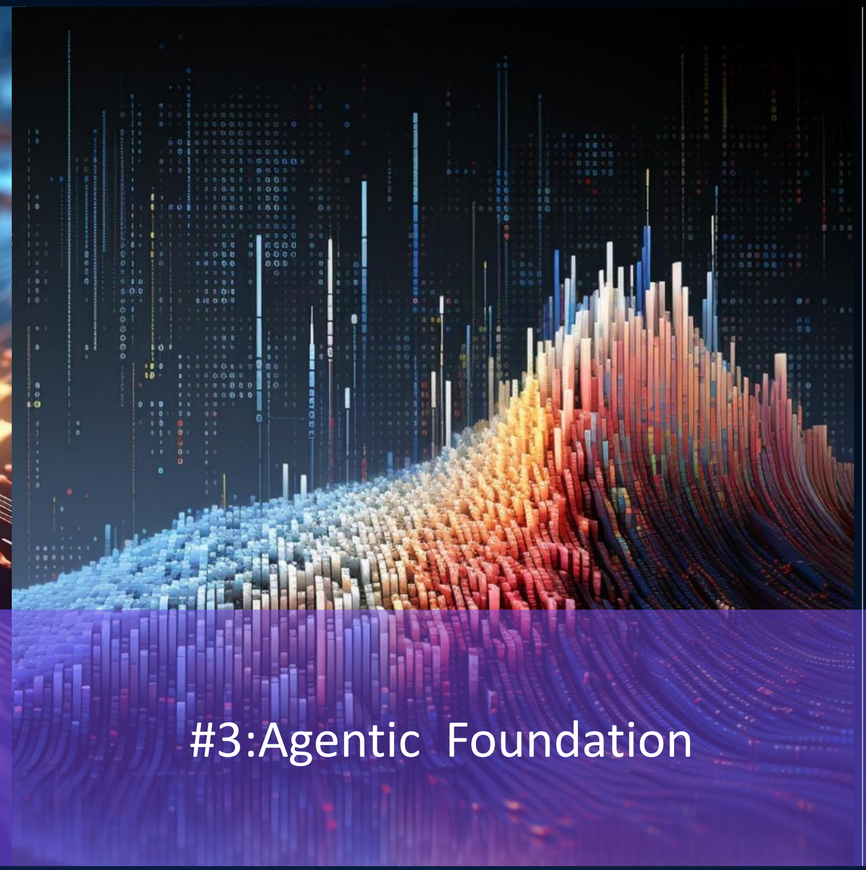
Lets see this in Action !

Success Factors

#1: Data Strategy



#2: Choice of Models



#3:Agentic Foundation

Your AI engine

Data: The critical fuel

Quality vs. Quantity

Use case-specific data needs

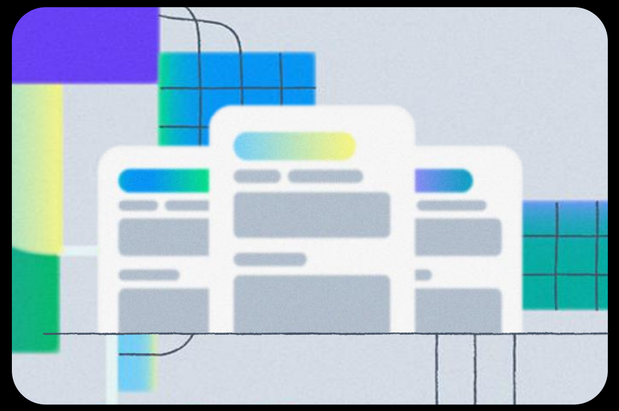
Data availability and accessibility

80% of companies are revising their data strategies to support AI initiatives, emphasizing data quality improvements

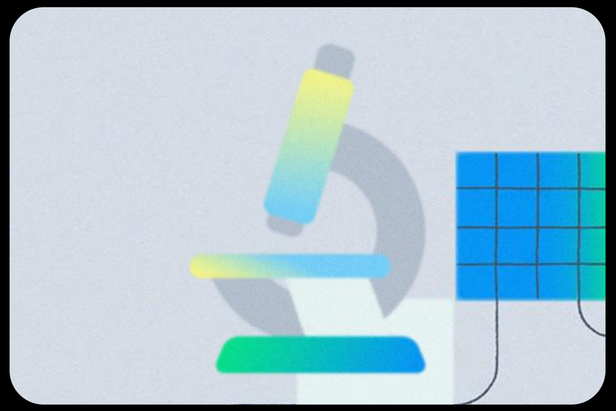
Source: Gartner

10

AI is only as smart as your data



YOUR EXISTING DATA



NEW DATA YOU CREATE

THIRD PARTY DATA

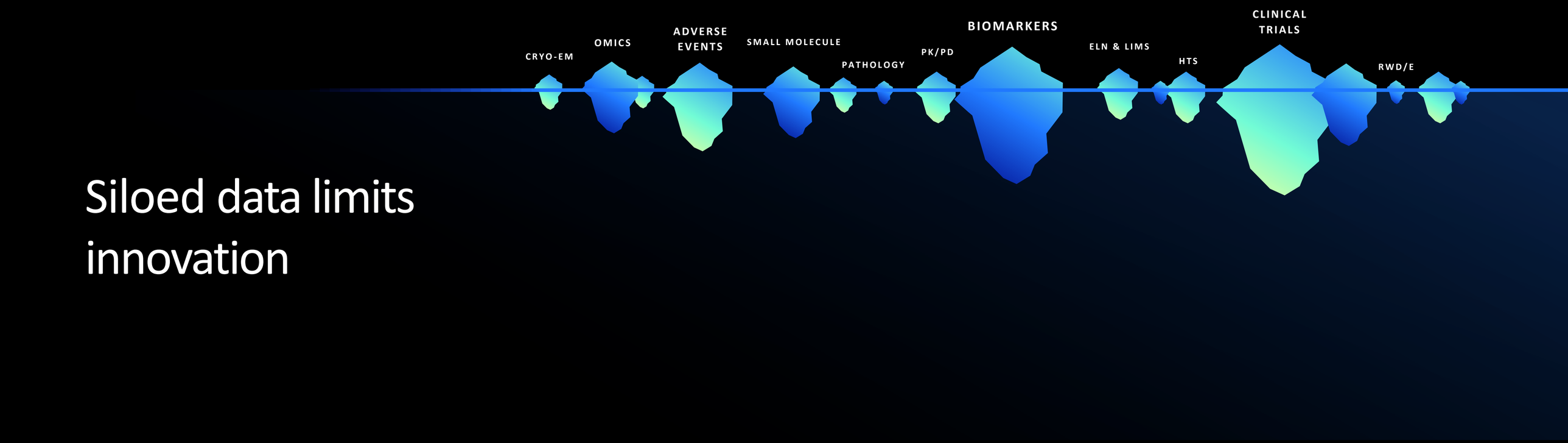
AI is only as smart as your data

YOUR DATA

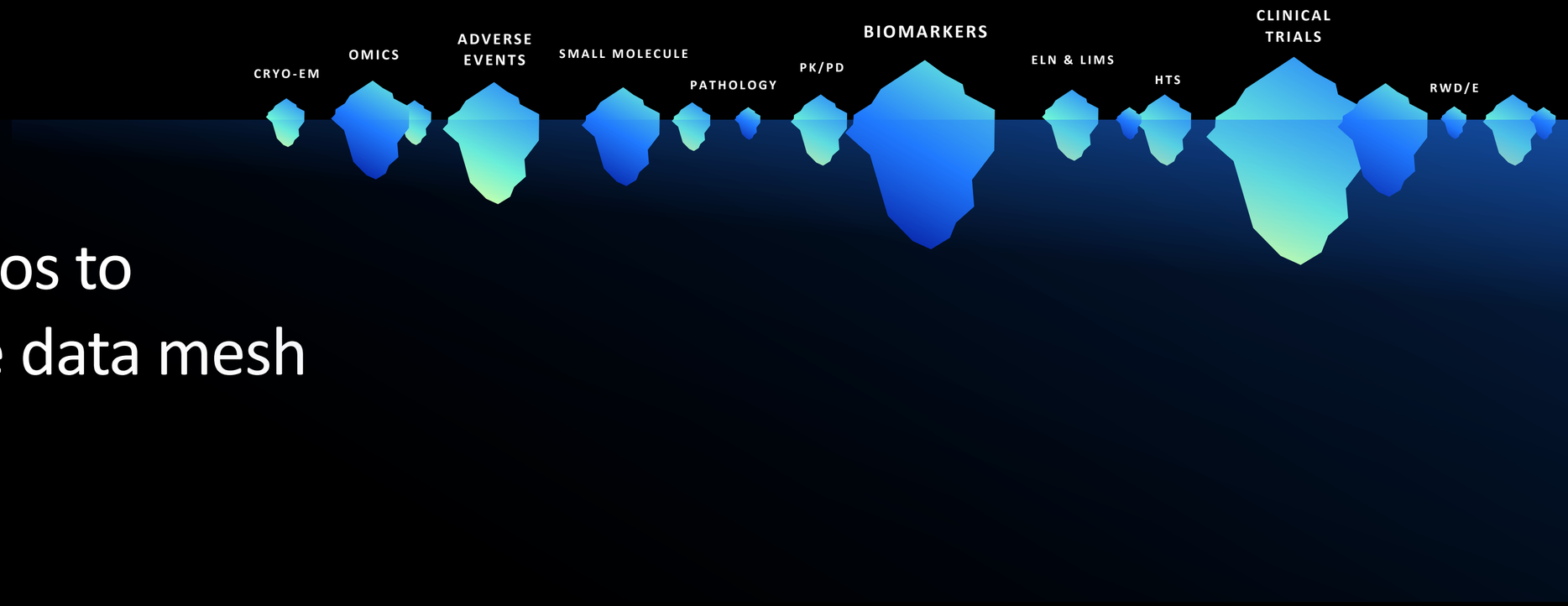
YOUR EXISTING DATA

NEW DATA YOU CREATE

THIRD PARTY DATA



Siloed data limits innovation



From data silos to an accessible data mesh

AI is only as smart as your data

1P DATA

DATA YOU CREATE

THIRD PARTY DATA

DISCOVERY

Drug discovery
(Target ID, MoA)

Early product
planning

CLINICAL

Clinical
trial design

Synthetic control
arms

Clinical trial
recruitment

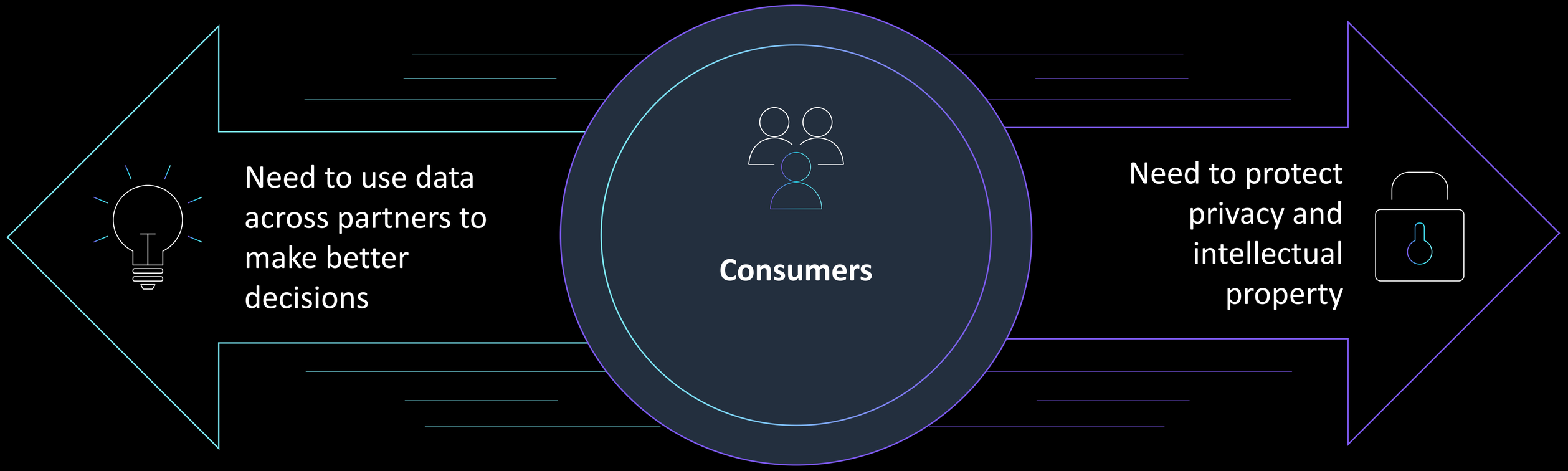
COMMERCIAL

HEOR:
Value & Economics

HEOR: Regulatory,
Safety & Lifecycle Mgmt.

Sales and
marketing strategy

REAL-WORLD DATA IS USED THROUGHOUT THE VALUE CHAIN


Need to use data across partners to make better decisions
Consumers
Need to protect privacy and intellectual property

Discover, evaluate & subscribe

datavant

DATA PRODUCERS

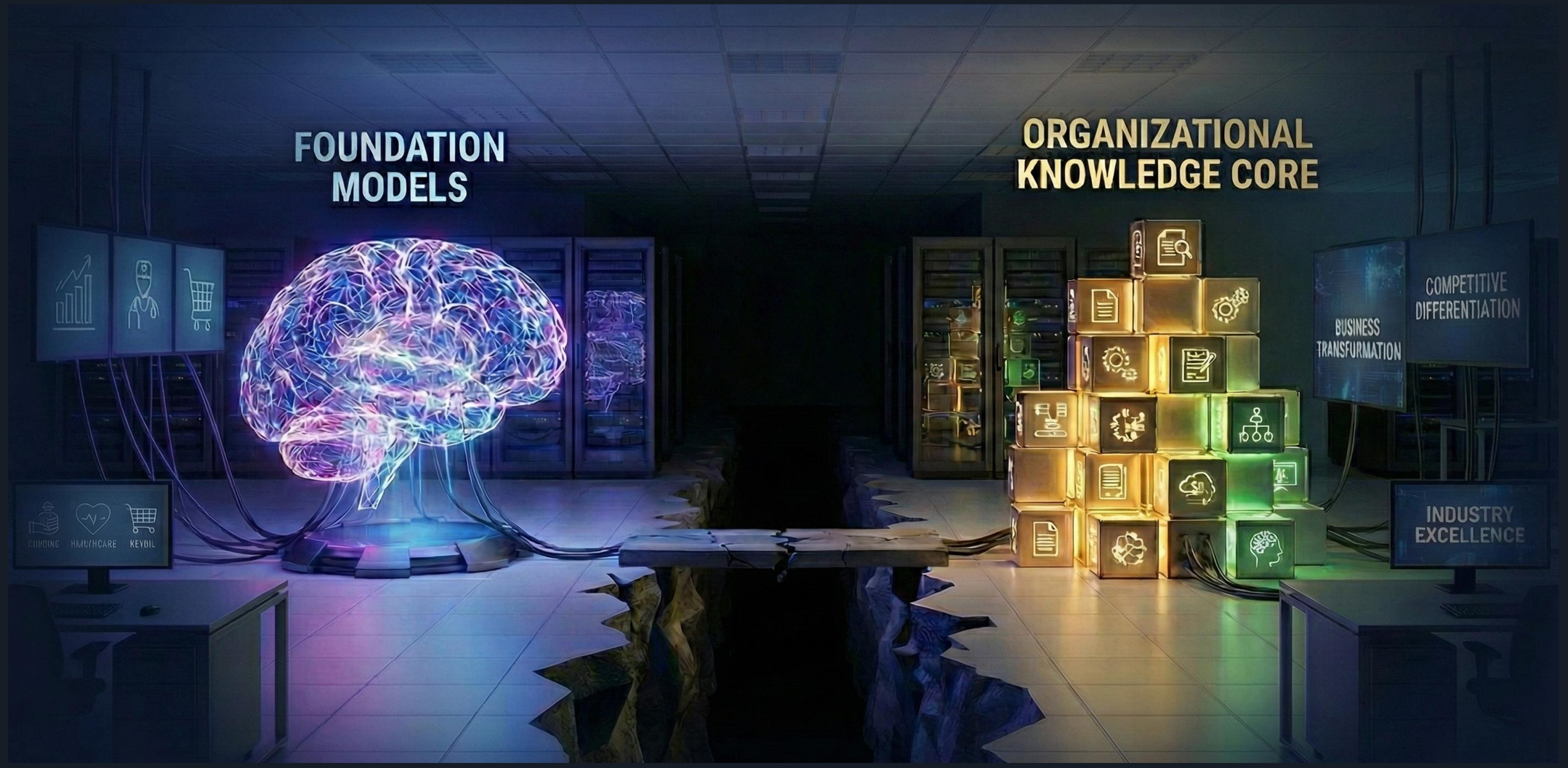
AnalyticsIQ · carelon · Diaceutics (Better Testing, Better Treatment) · prognos health · OMNY HEALTH · ONEMEDNET · Verana Health

Unlocking Real-World Data and Evidence

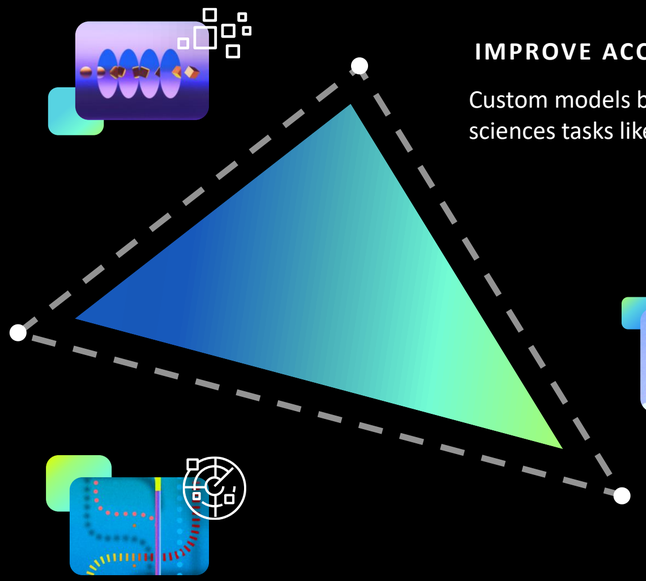
Boehringer Ingelheim · datavant · Lilly · carelon Behavioral Health

DATA CONSUMERS

The challenge



Training foundation models with your data



IMPROVE ACCURACY

Custom models boost accuracy in life sciences tasks like drug discovery

PROPRIETARY DATA UTILIZATION

Life sciences companies use proprietary data for competitive advantage

REGULATORY COMPLIANCE

Custom models ensure compliance with strict regulations in industries like pharmaceuticals and healthcare

Small Language Models are the Future of Agentic AI

Peter Belcak[1] **Greg Heinrich**[1] **Shizhe Diao**[1] **Yonggan Fu**[1] **Xin Dong**[1]
Saurav Muralidharan[1] **Yingyan Celine Lin**[1,2] **Pavlo Molchanov**[1]
[1]NVIDIA Research [2]Georgia Institute of Technology
agents@nvidia.com

Abstract

Large language models (LLMs) are often praised for exhibiting near-human performance on a wide range of tasks and valued for their ability to hold a general conversation. The rise of agentic AI systems is, however, ushering in a mass of applications in which language models perform a small number of specialized tasks repetitively and with little variation.

Here we lay out the position that small language models (SLMs) are *sufficiently powerful, inherently more suitable, and necessarily more economical for many invocations in agentic systems, and are therefore the future of agentic AI*. Our argumentation is grounded in the current level of capabilities exhibited by SLMs, the common architectures of agentic systems, and the economy of LM deployment. We further argue that in situations where general-purpose conversational abilities are essential, heterogeneous agentic systems (i.e., agents invoking multiple different models) are the natural choice. We discuss the potential barriers for the adoption of SLMs in agentic systems and outline a general LLM-to-SLM agent conversion algorithm.

Our position, formulated as a value statement, highlights the significance of the operational and economic impact even a partial shift from LLMs to SLMs is to have on the AI agent industry. We aim to stimulate the discussion on the effective use of AI resources and hope to advance the efforts to lower the costs of AI of the present day. Calling for both contributions to and critique of our position, we commit to publishing all such correspondence at `research.nvidia.com/labs/lpr/slm-agents`.

1 Introduction

The deployment of agentic artificial intelligence is on a meteoric rise. Recent surveys show that more than a half of large IT enterprises are actively using AI agents, with 21% having adopted just within the last year [12]. Aside from the users, markets also see substantial economic value in AI agents: As of late 2024, the agentic AI sector had seen more than USD 2bn in startup funding, was valued at USD 5.2bn, and was expected to grow to nearly USD 200bn by 2034 [42, 47]. Put plainly, there is a growing expectation that AI agents will play a substantial role in the modern economy.

The core components powering most modern AI agents are (very) large language models [48, 44]. It is the LLMs that provide the foundational intelligence that enables agents to make strategic decisions about when and how to use available tools, control the flow of operations needed to complete tasks, and, if necessary, to break down complex tasks into manageable subtasks and to perform reasoning for action planning and problem-solving [48, 14]. A typical AI agent then simply communicates with a chosen LLM API endpoint by making requests to centralized cloud infrastructure that hosts these models [48].

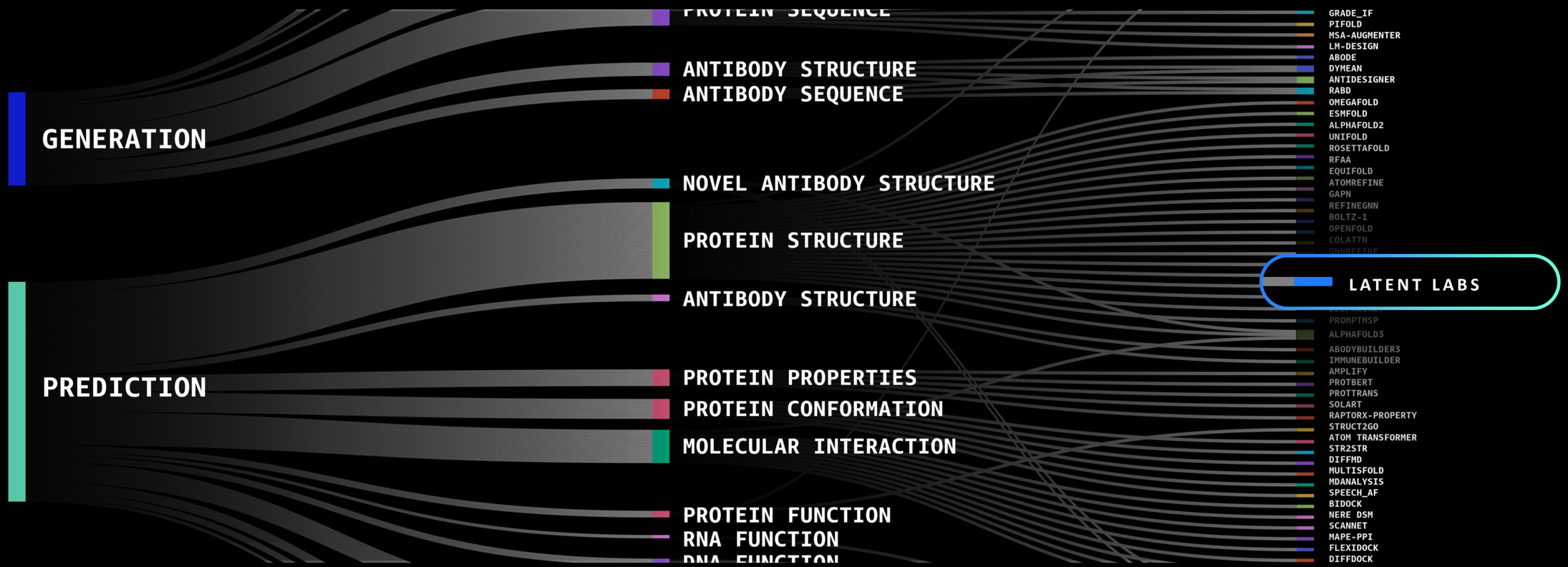
Preprint. Under review.

Nimbus Therapeutics

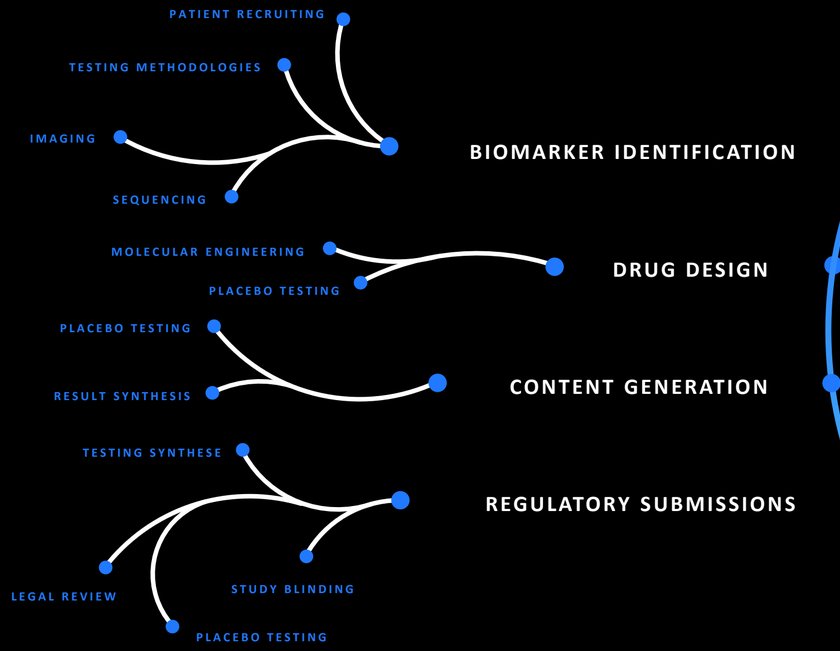
Drug Discovery Assistant

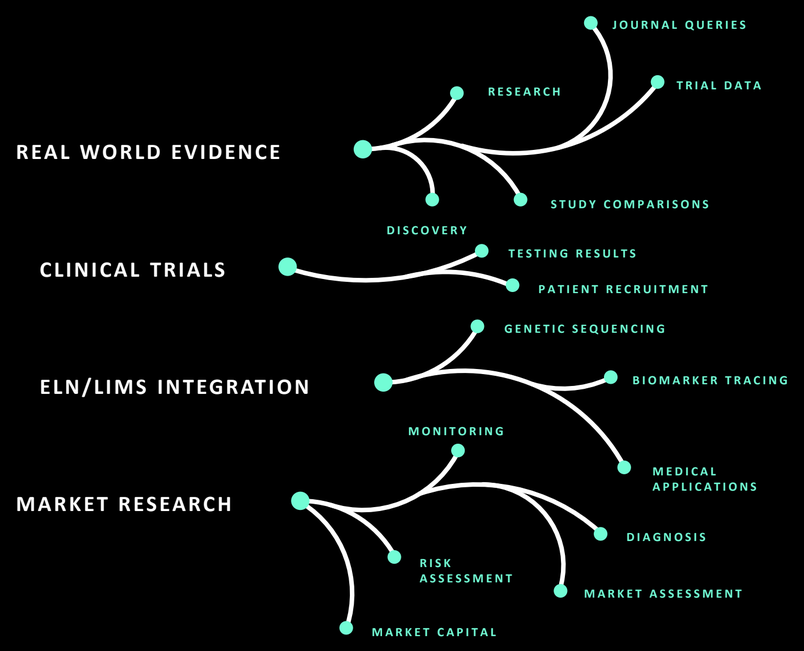
Use Case	Solution	Result
LLM for Pharmaceutical R&D and Drug Discovery	Supervised Fine-Tuning and Reinforcement Learning	Converged several task-specialist GNN models into one model Outperformed LLMs such as Sonnet 4

> *"We're using Nova Forge to build a unified drug discovery assistant that can predict molecular properties, reason through chemistry problems, and generate novel drug candidates. By exploring thousands of candidates computationally before testing in the lab, where each experiment costs in the thousands, we can bring better medicines to patients faster while reducing costs. Through supervised fine-tuning and reinforcement fine-tuning with Nova 2 Lite, we have already outperformed existing large language models such as Sonnet 4 by 20-50% on property prediction tasks; exceeded or matched the performance of several specialized GNN models on the same tasks, and we are now moving into molecular generation."*
>
> *- Leela Dodda, Director of Computational Chemistry*

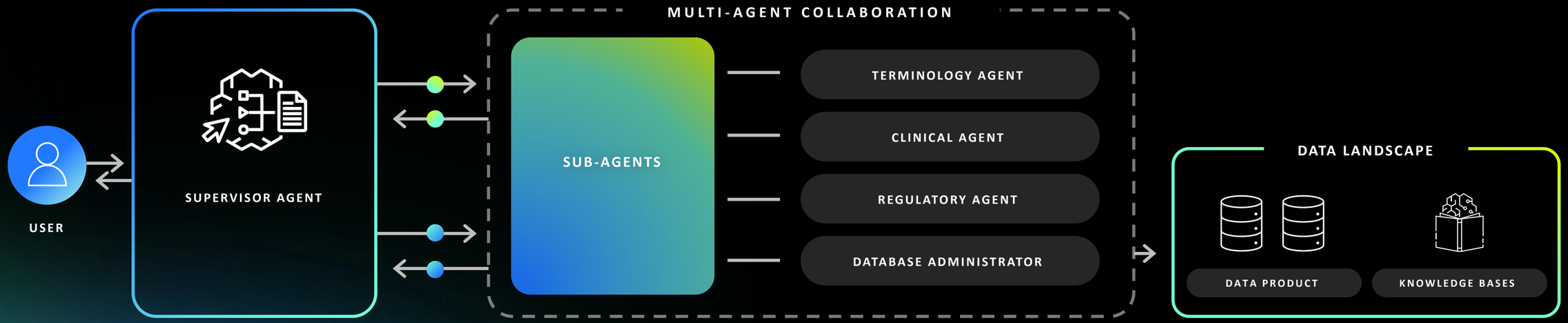

GENERATION
PREDICTION
PROTEIN SEQUENCE
ANTIBODY STRUCTURE
ANTIBODY SEQUENCE
NOVEL ANTIBODY STRUCTURE
PROTEIN STRUCTURE
ANTIBODY STRUCTURE
PROTEIN PROPERTIES
PROTEIN CONFORMATION
MOLECULAR INTERACTION
PROTEIN FUNCTION
RNA FUNCTION
GRADE_IF
PIFOLD
MSA-AUGMENTER
LM-DESIGN
ABODE
DYMEAN
ANTIDESIGNER
RABD
OMEGAFOLD
ESMFOLD
ALPHAFOLD2
UNIFOLD
ROSETTAFOLD
RFAA
EQUIFOLD
ATOMREFINE
GAPN
REFINEGNN
BOLTZ-1
OPENFOLD
COLATTN
GNNREFINE
LATENT LABS
PROMPTMSP
ALPHAFOLD3
ABODYBUILDER3
IMMUNEBUILDER
AMPLIFY
PROTBERT
PROTTRANS
SOLART
RAPTORX-PROPERTY
STRUCT2GO
ATOM TRANSFORMER
STR2STR
DIFFMD
MULTISFOLD
MDANALYSIS
SPEECH_AF
BIDOCK
NERE DSM
SCANNET
MAPE-PPI
FLEXIDOCK
DIFFDOCK

Proliferation of agents





Multi-agent systems enable deeper expertise and scalability



Lets see this in Action !

What you need for getting agents into production

Open Source Life Sciences Agent Toolkit on AWS

- 20+ starter agents
- Built-in multi-agent orchestration
- Customizable to meet the needs of your organization

BIOMARKER SUPERVISOR AGENT

- Statistician
- Evidence research
- Biological pathways
- Wiley 3P

RESEARCH

Target identification and research

CLINICAL STUDY SUPERVISOR

- Clinical protocol generator
- Clinical trials search

CLINICAL DEVELOPMENT

Clinical trial protocol analysis and design

CONTENT SUPERVISOR AGENT

- Competitive intelligence
- Content generator
- Compliance reviewer
- MLP

COMMERCIAL

Competitive intelligence and content generation

AND MORE TO COME


Reinforcement Learning at Scale in the next frontier…

Physical AI : Meet “Vulcan” - A robot with a sense of touch

THANK YOU