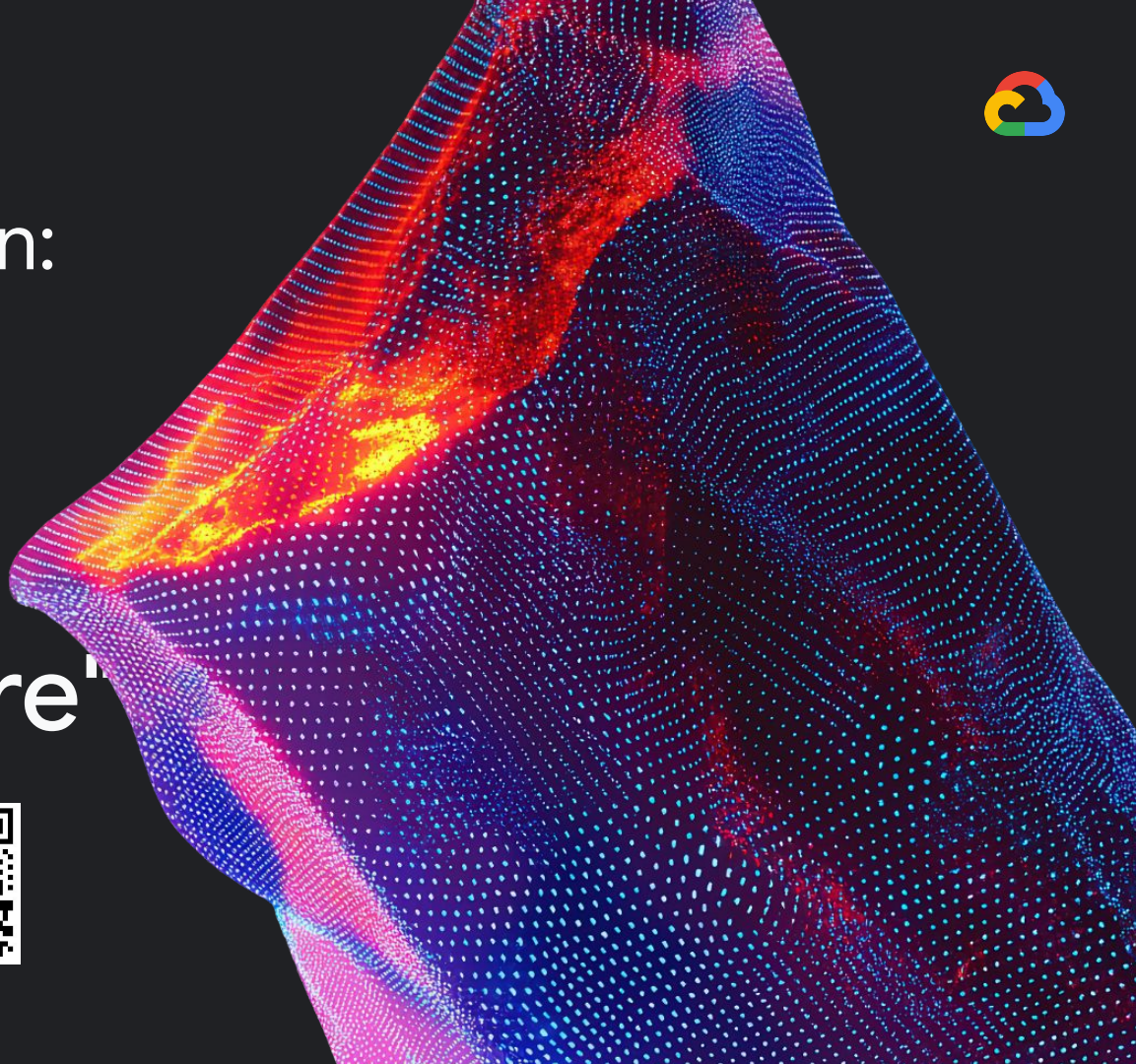


Beyond Automation: Agentic AI & the Rise of "Service-as- a-Software"

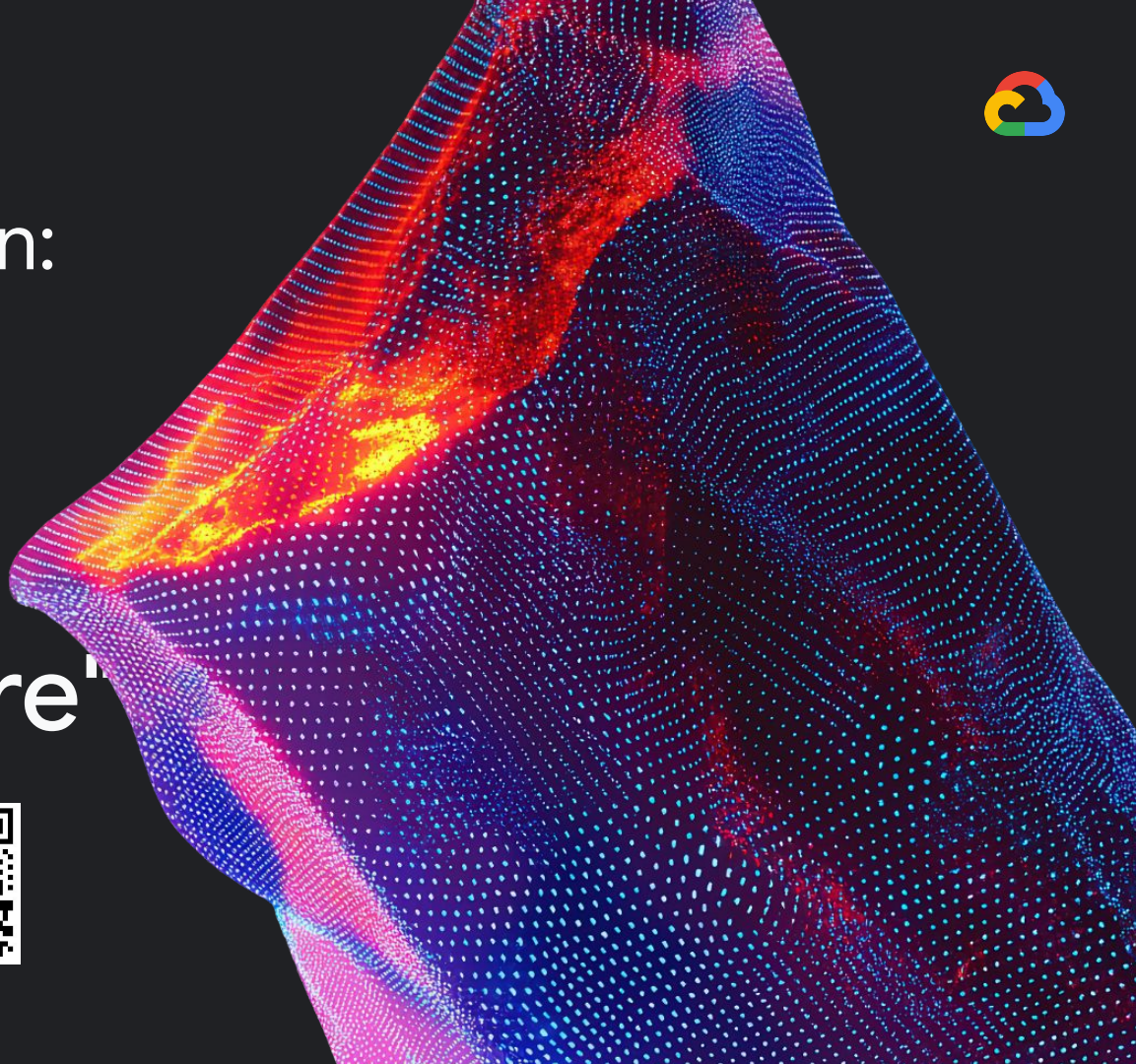
Gerhard Keller
Head of Manufacturing
Google Cloud DACH





Beyond Automation: Agentic AI & the Rise of "Service-as- a-Software"

Gerhard Keller
Head of Manufacturing
Google Cloud DACH



How strong is your AI literacy?

B2B is less about movies.

B2B is about data driven
processes.



From Software-as-a-Service To Service-as-a-Software

What are AI agents?

Agents are systems that combine the **intelligence of** advanced **AI** models with **access to tools** so they can **take actions** on your behalf, under your control.

An agentic system is a digital assembly line

a **human-guided, multi-step workflow** that **orchestrates multiple agents** to run a business process end to end.



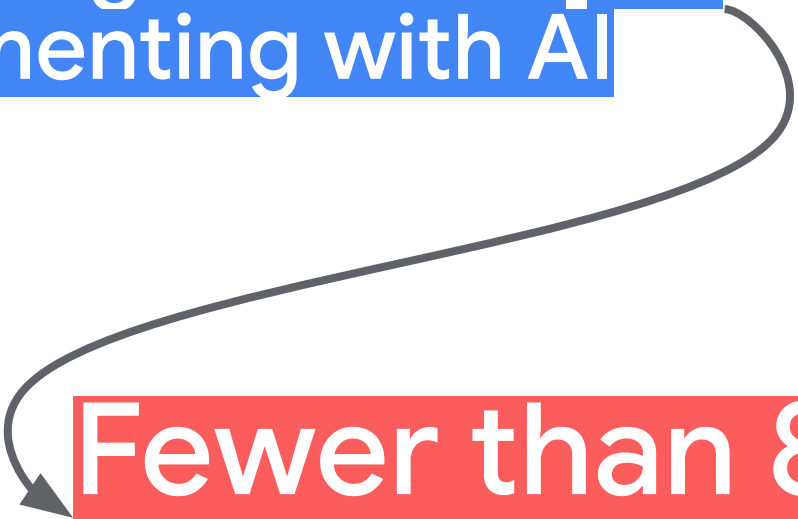
Every employee becomes
a human supervisor
to orchestrate a team of
specialized AI agents

Moving from **instruction-based**
to **intent-based** computing

State a desired outcome,
and the computer determines
how to deliver it.

- | Delegate mundane or repetitive tasks
- | Set goals
- | Outline strategy
- | Verify quality

62% of organizations are
experimenting with AI



Fewer than 8% made
it past the pilot

Successful AI adopters redesigned their business processes to be **AI first**

There are two kinds of AI agents



Individual Productivity



Enterprise Efficiency

Let's talk about Individual Productivity first



Individual Productivity



Enterprise Efficiency

Individual Productivity agents are built and controlled by your users, because **they know best** what they need

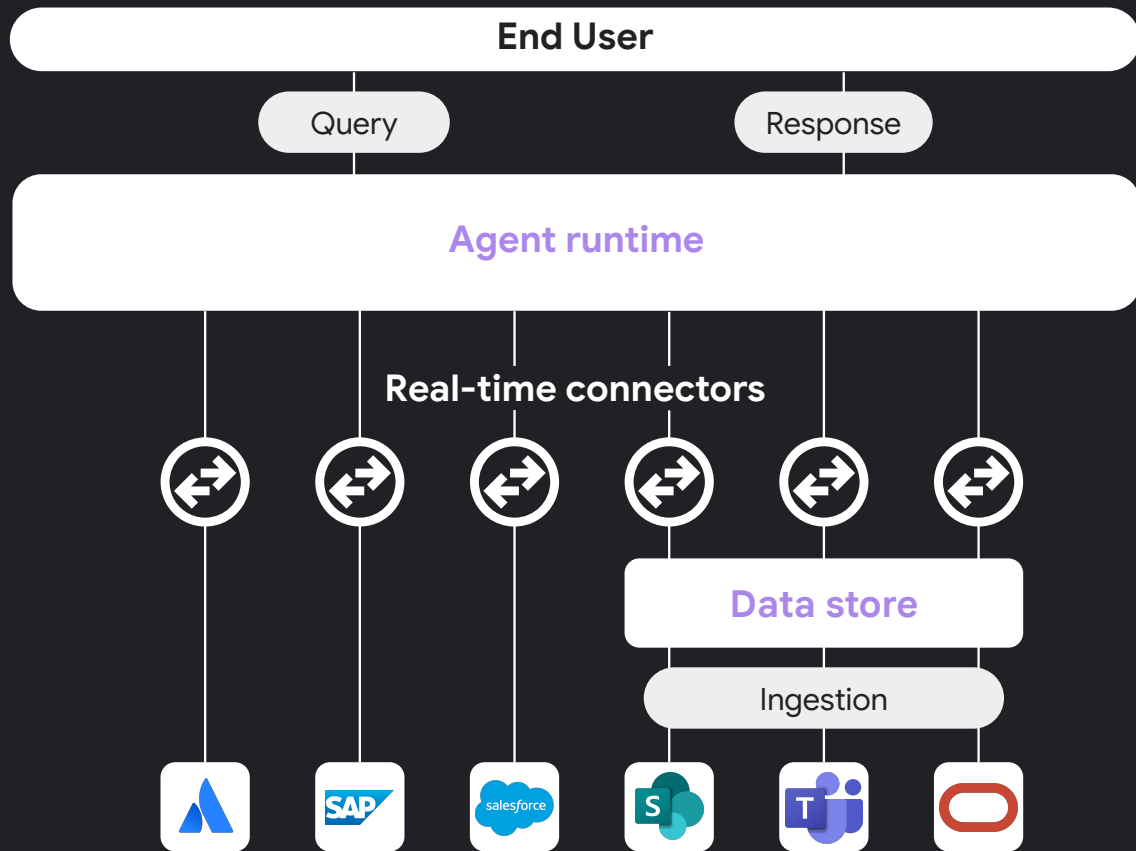
Conduct market analysis to find potential targets

Inform customers of a price change

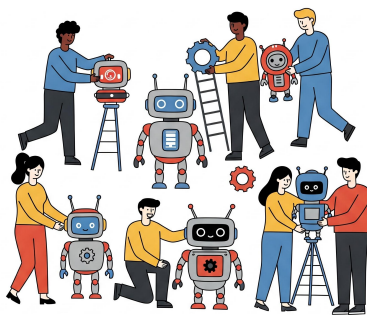
Draft an e-mail to respond to recent outages

Create unit tests for this module

The biggest challenge
Exposing your
enterprise data so it's
useful for agents



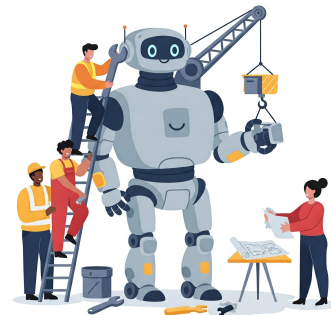
For Individual Productivity, we recommend a **pragmatic approach**: Experimentation, with guardrails



Anyone can build
agents
(within guardrails)

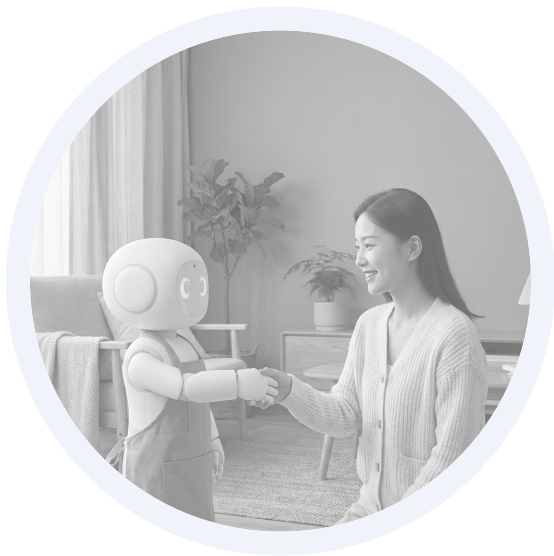


All agents fall under
central monitoring/
governance



Popular agents are
improved by a central
team

Now let's look at Enterprise Efficiency



Individual Productivity



Enterprise Efficiency

Enterprise Efficiency agents shine where you have repeatable, standardized processes



Coding assistants

Your engineers need to convert legacy code, write unit tests or perform user interface testing that are not differentiating



Invoice processing

You receive 1,000s of invoices per month; AI agents extract relevant information from invoices and enter into your accounting system



Asset creation

You need to create 1,000s of assets for your website each month, tailored to your customers' changing preference

Good processes for AI agents are similar to good processes for human users

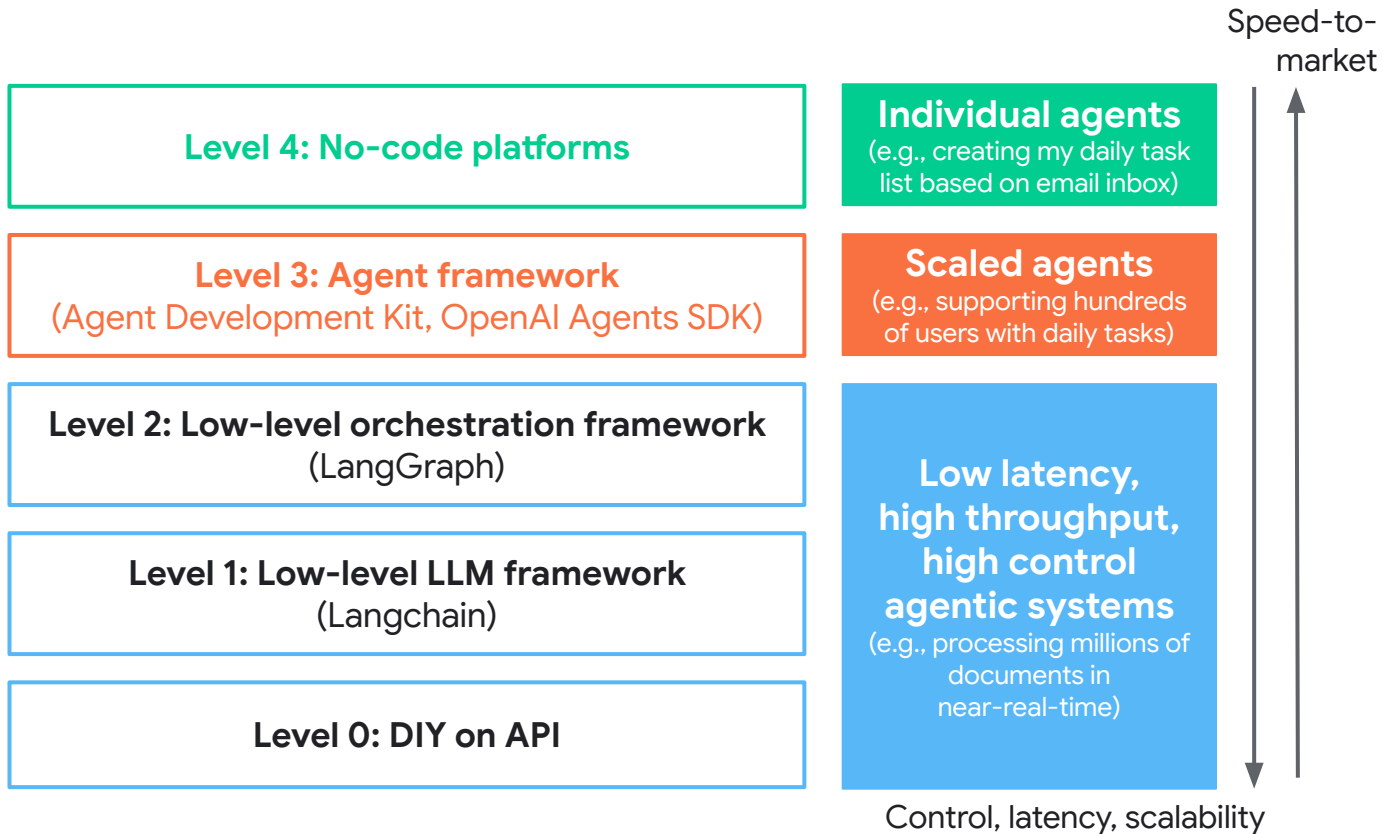
1 Process documentation (Bonus: It's machine-readable!)

2 Clearly defined input and output criteria

3 Forms over free-hand

4 Controls implemented

You should start with **no-code** platforms to move fast, and go deeper if you need more controls

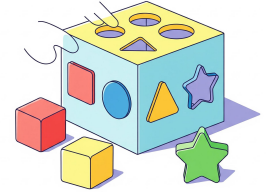




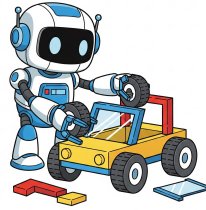
Understand the
business problem



Establish
testing framework



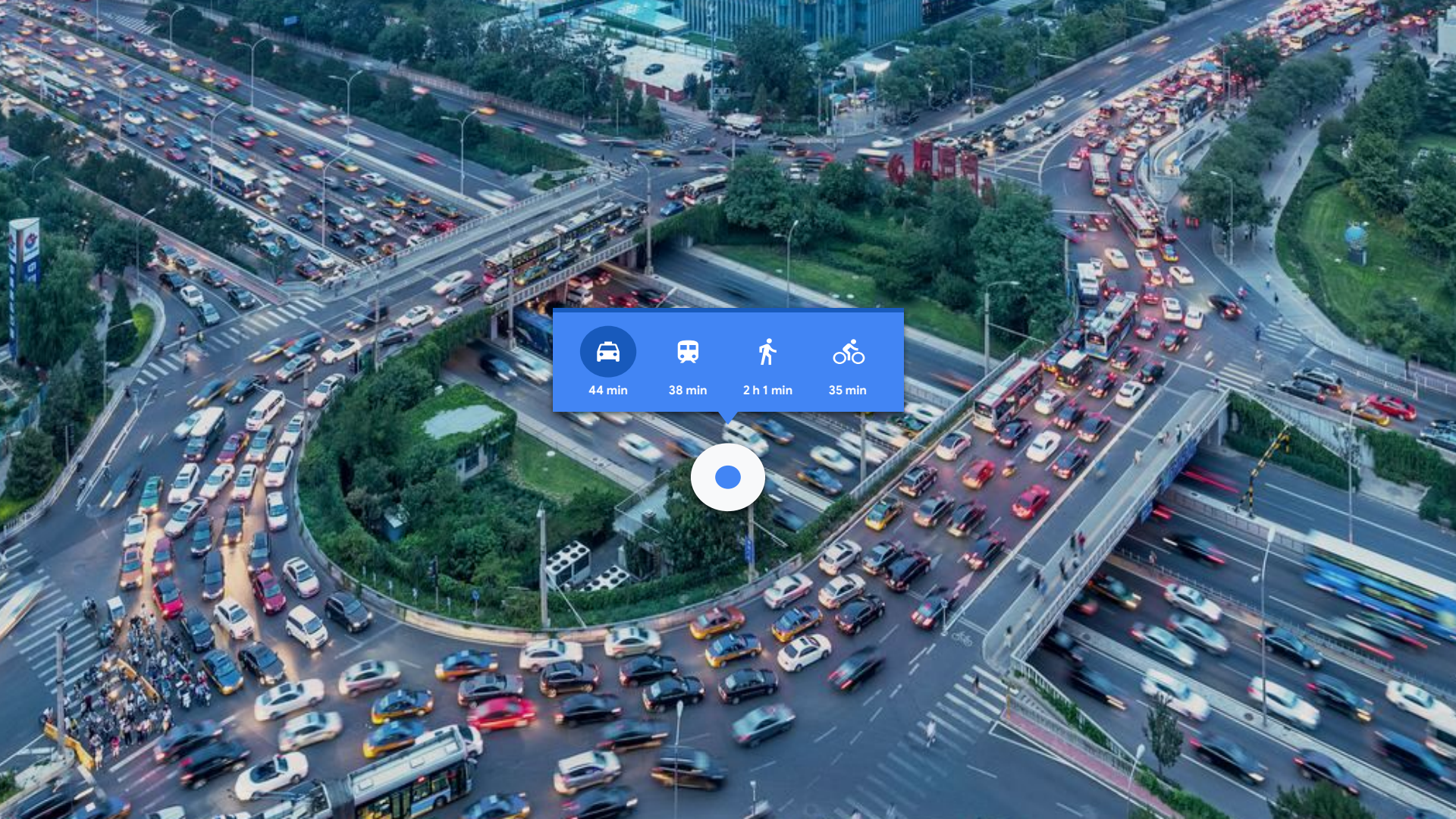
Implement
base application



Run experiments to
improve answer
quality



Path to production
Understand
business problem,
establish testing
framework, iterate,
iterate, iterate



44 min 38 min 2 h 1 min 35 min



Google
Local BETA

What

chinese

Where

11222

Google Search

[Driving Directions](#)

[Search Tips](#)

[Google Home](#)

©2005 Google

Google
Local BETA

chinese near 11222



[Zoom In](#) | [Zoom Out](#)
[North](#)
[West](#) | [East](#)
[South](#)



Turn left

on Spring St



Then →



Google on Google AI

How Google uses our own AI internally



Increase productivity

- Increased efficiency
- Higher employee output
- Higher job satisfaction



Grow revenue

- Faster time to market
- Higher conversion rates
- Improved customer experience



Innovate faster

- Industry leadership
- Competitive differentiation





"30% OF ALL NEW CODE
AT GOOGLE IS NOW
AI-GENERATED"

SUNDAR PICHAI

ALPHABET EARNINGS CALL

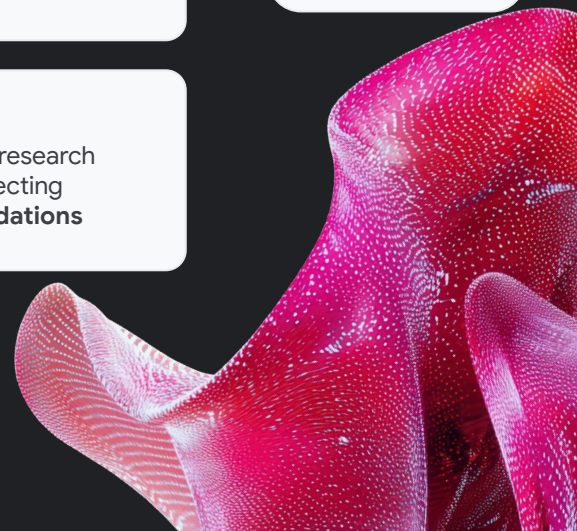
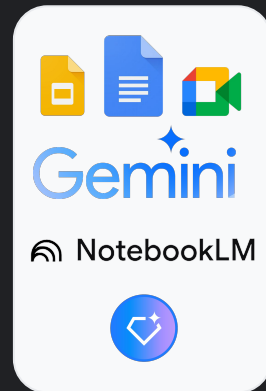
APRIL 2025



"75% OF ALL NEW CODE
AT GOOGLE IS NOW
AI-GENERATED"

SUNDAR PICHAI
GOOGLE CLOUD NEXT
APRIL 2026

How Google is using AI internally



Back office

Partner

Reduced approval time 25%, by evaluating partner SoWs against predefined criteria before submission to Google.

Support

Personalized learning content gives Support specialists a skill boost, leading to 20% faster mitigations.

Support

AI case summaries empower Support agents to deliver better customer outcomes, with a 10% reduction in 'time with support' for complex cases.



Back office

Finance

Doubled invoice validation team's capacity by converting 100% of expert execution to expert training.

Security

Faster threat detection & proactive defense, by augmenting security experts.

AI-assisted tasks



Alert review

15%

of security alerts are now automatically investigated, eliminating manual review

The AI Game changer out of Google Labs

NotebookLM



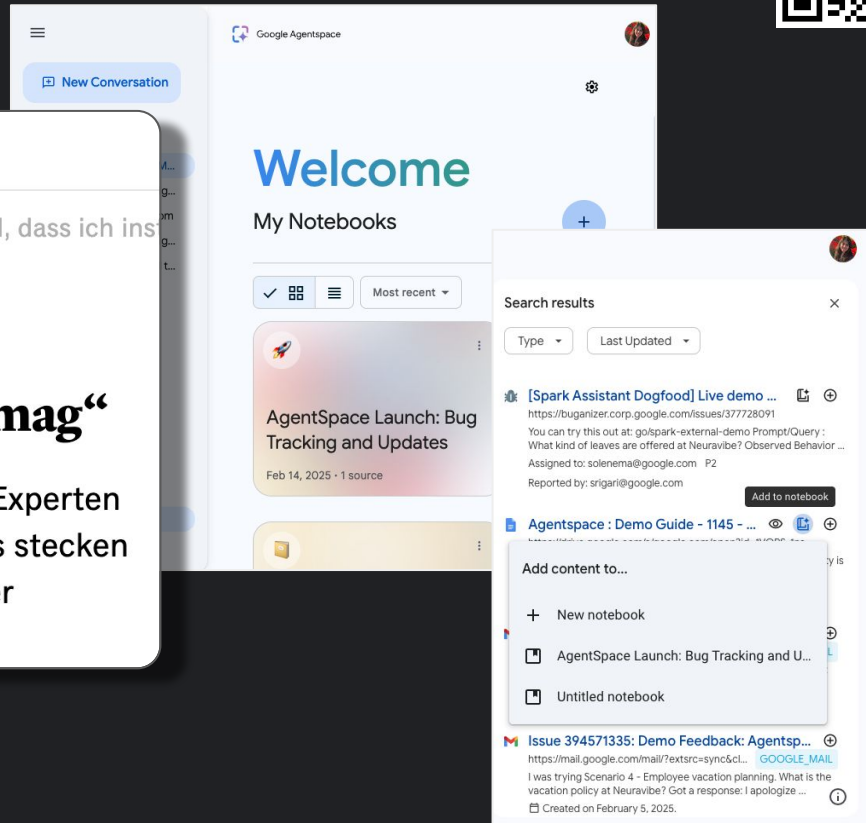
Handelsblatt

Technik > IT + Telekommunikation > Neues Google-Tool: „Das erste Mal, dass ich ins

H+ Neues Google-Tool

„Das erste Mal, dass ich instinktiv KI mag“

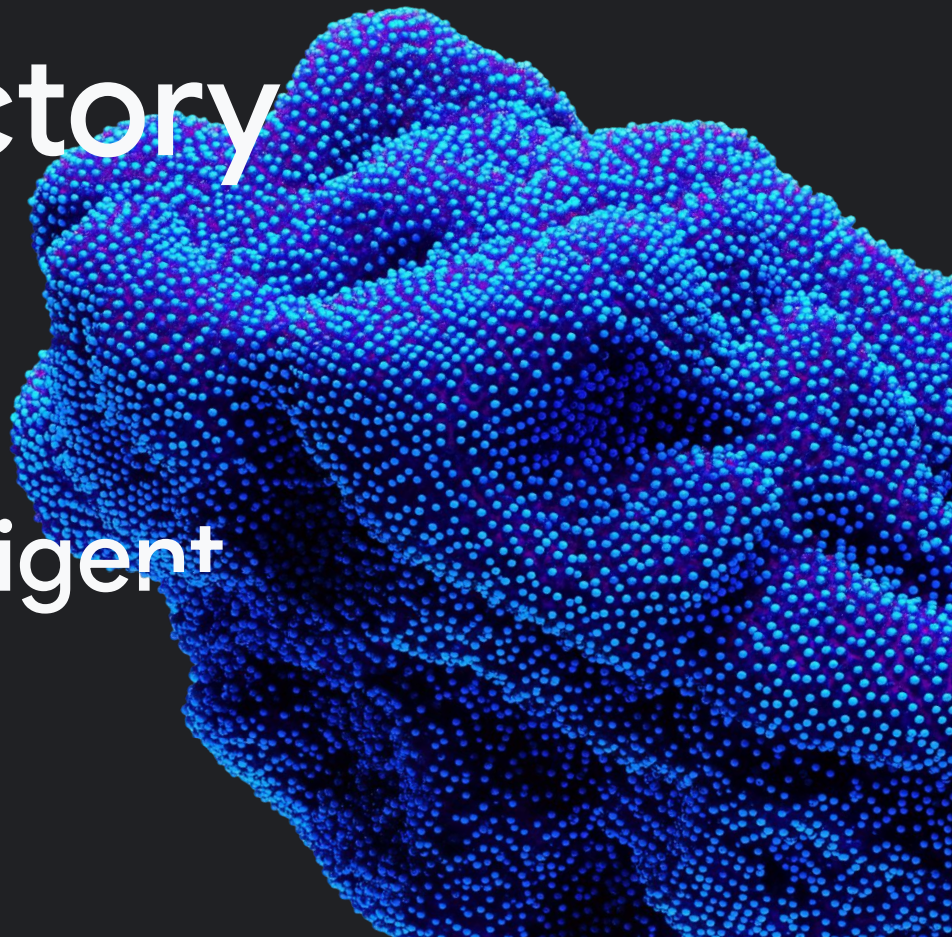
Die KI-Anwendung Notebook LM von Google begeistert Experten wie Anwender. Hinter dem neuen Ansatz des IT-Konzerns stecken eine junge Produktmanagerin und ein Zettel-sammelnder Schriftsteller.



Google Cloud

The Agent Factory

How Kärcher uses
APIs to create the
foundation for intelligent
automation



Decreasing “Time to Agent” and Scale with Confidence

95% of AI Pilots* Do Not Go Into Production.

Why

- Your data is stuck in SAP, Salesforce, and Oracle.
- Lack of governance against
 - a. Shadow AI via MCPs
 - b. Data leakage
 - c. ‘Prompt Injection’
- Do you have an ‘Emergency Stop Switch’ for your AI agents?



* "The GenAI Divide: State of AI in Business 2025", MIT NANDA Report

The Agent Factory at **KÄRCHER**

zoi



Implement a Single Namespace as a Gateway.



"We need to move away from thinking in individual tools and towards a consistent digital journey."

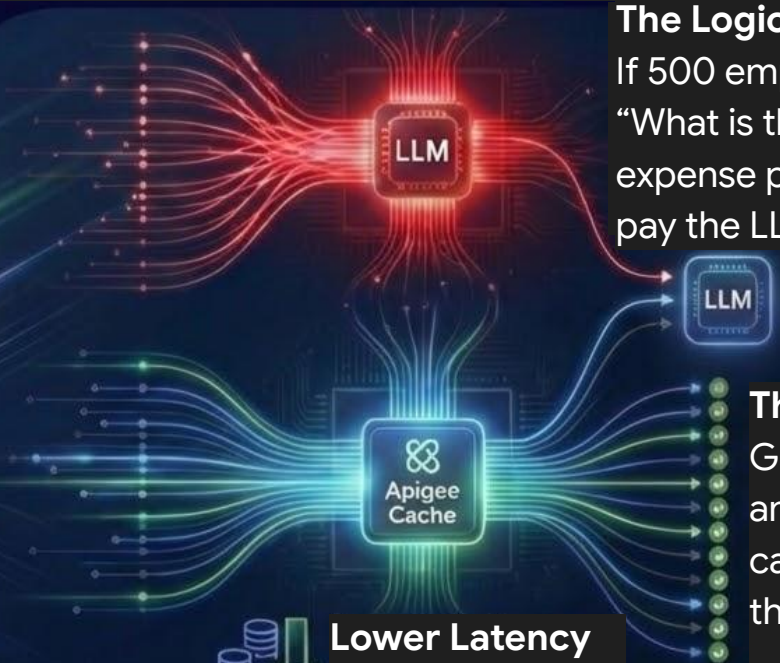
– Leonhard Kerscher, CIO Kärcher

FinOps for the CFO

Semantic Caching & Smart Routing

46%

Reduction
in AI Bill



The Logic:

If 500 employees ask “What is the travel expense policy?”, you pay the LLM 500x.

The Solution:

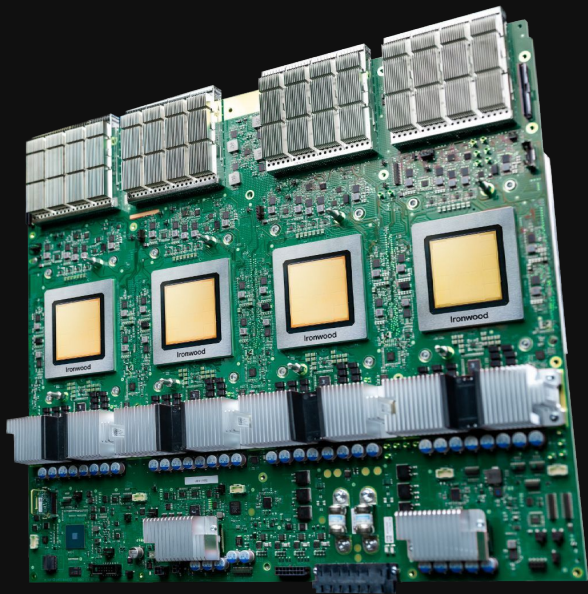
Gateway answers from cache for free the 2nd time.



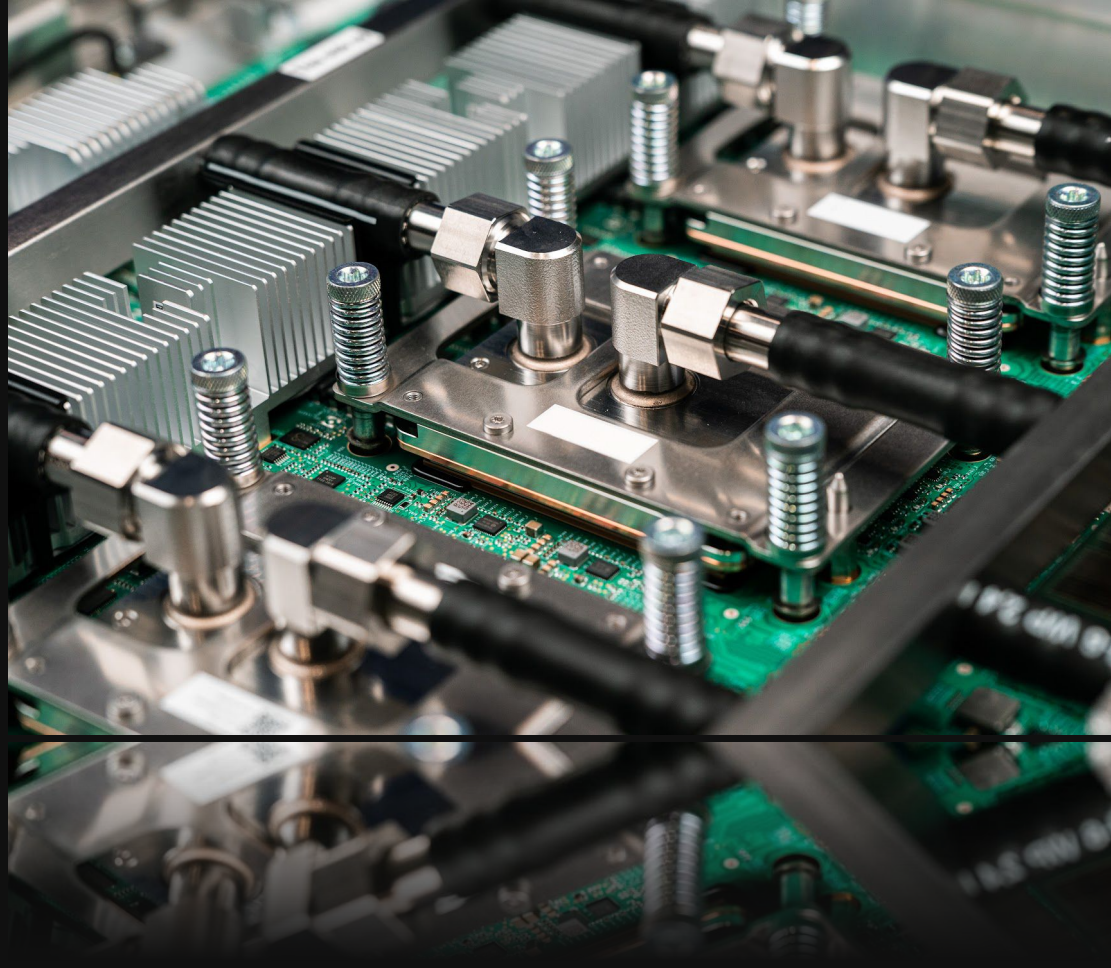
Lower Latency
Lower Costs

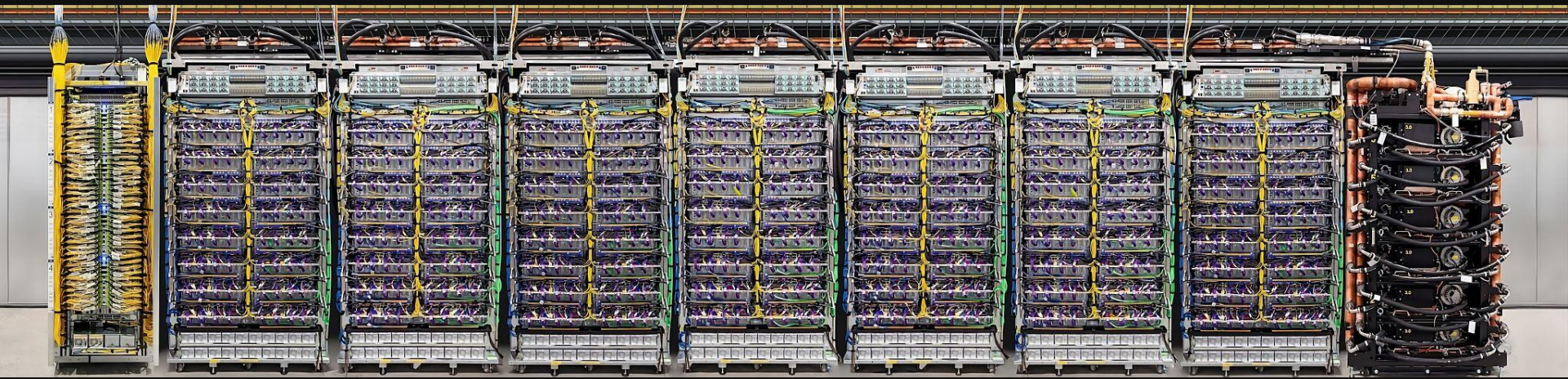
Building the AI Infrastructure of the future





Ironwood TPU V8
12,600 TFLOPs





4 TPUs per Tray
16 Trays per Rack
Superpod scales to 144 Racks
up to 9,216 chips per pod, 116 Exaflops

$$289,347 \times 656,651$$



8 BILLION PEOPLE

1 MIN PER CALCULATION

24 HOURS STRAIGHT

FOR 3 YEARS

= 1 SECOND



Thales Announces Strategic Partnership with Google Cloud to Launch a New Sovereign Cloud in Germany



“Germany represents a critical market for sovereign technologies, and this partnership is a direct response to [...] organizations wanting access to Google Cloud’s technology under full German control.

By launching this locally operated infrastructure, we are delivering a solution that guarantees sensitive workloads remain protected from any extraterritorial reach while meeting the unique security and compliance requirements of our customers,”

Hallo Deutschland!

THALES Google Cloud

**Christoph Ruffner,
CEO and Country Director of
Thales in Germany.**

Google Research : Project Suncatcher



<https://arxiv.org/abs/2306.01875>

Towards a future space-based, highly scalable AI infrastructure system design

Blaise Agüero y Arcas¹, Travis Bole¹, Maria Biggs¹, Jessica V. Bloom¹, Thomas Fischbacher¹, Kunturam Gromov¹, Urs Käster¹, Rishabh Iyer¹, and James Manyia¹
Equal contribution, Google

If AI is a foundational general-purpose technology, we should anticipate that demand for AI compute — and energy — will continue to grow. The Sun is by far the largest energy source in our solar system, and thus it warrants consideration how future AI infrastructure could most efficiently tap into that power. This work explores a scalable compute system for machine learning in space, using fleets of satellites equipped with solar arrays, inter-satellite links using free-space optics, and Google tensor processing unit (TPU) accelerator chips. To facilitate high-bandwidth, low-latency inter-satellite communication, the satellites would be flown in close proximity. We illustrate the basic approach to formation flight via a 61-satellite cluster of 1 km radius, and describe an approach for using high-precision ML-based models to control large-scale constellations. Trillion TPUs are radiation tested. They survive a total ionizing dose equivalent to a 5 year mission life without permanent failures, and are characterized for bit-flip errors. Launch costs are a critical part of overall system cost; a learning curve analysis suggests launch to low-Earth orbit (LEO) may reach ~\$200/kg by the mid-2030s.

1. Introduction

1.1. Motivation for ML in space

The development of the Transformer model (1) and the subsequent rise of generative, multimodal AI have led to an explosion of demand for compute capacity. Although dramatic gains have been made in efficiency (e.g. Gemini query energy consumption was reduced 32x over a one year period (2)), AI-based products and services have grown even faster, leading to a rapid increase in data center energy demand. To that end, Google is invested in advancing new forms of power generation (e.g. (3–5)). However, given AI appears to be a foundational general-purpose technology (6), which its characteristics are then assessed.

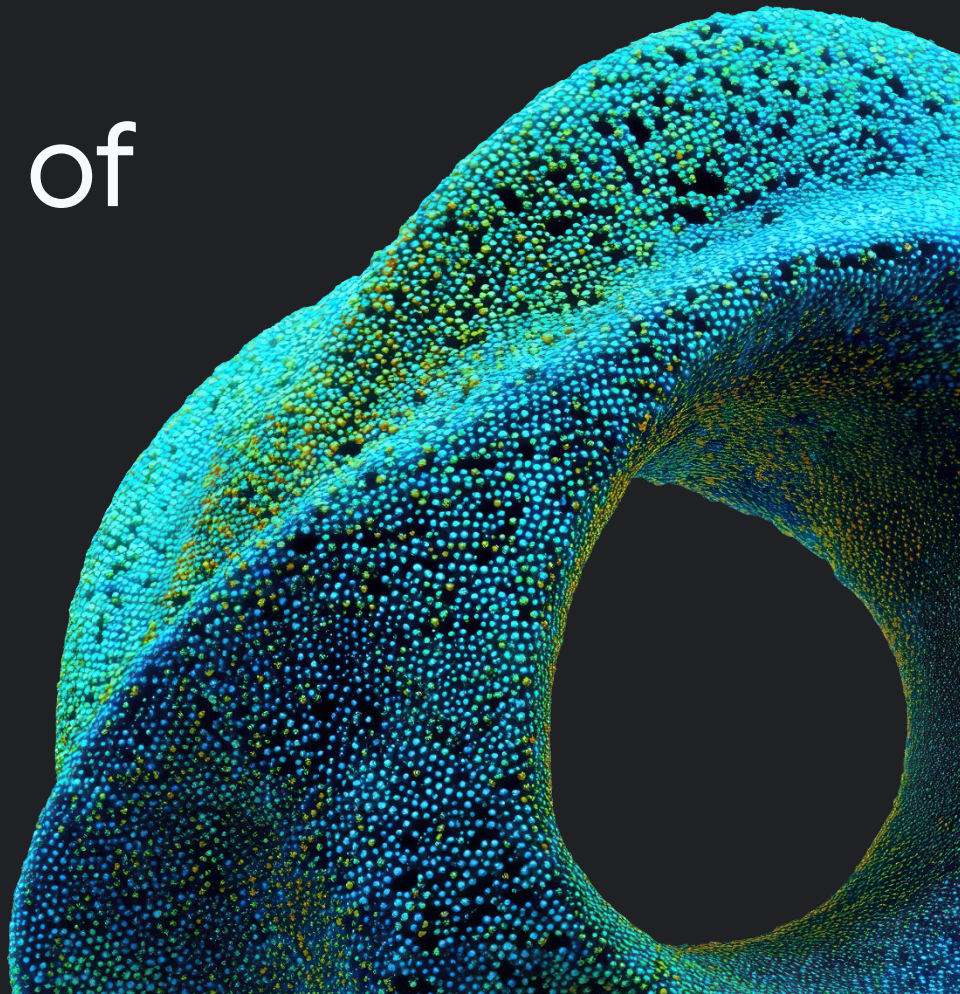
of power in our solar system: with an output of 3.86×10^{26} W, the Sun emits more than 100 trillion times humanity's total electricity production. At some point in the future, the best way to power AI will likely thus be to more directly tap into that enormous source of energy. Space-based solar power has long been proposed — first in a 1941 *Ainova* short story, *Reason*, then more seriously (10, 11). It has many of the attractive qualities of terrestrial solar power, with the additional advantages that solar panels in certain orbits are exposed to nearly continuous sunshine, and receive up to 8x more solar energy per year than a panel located on Earth at mid-latitude (12). However, getting the generated power back to Earth has been a major challenge for such proposals.

24h free power
free cooling
zero water



Powering the era of
physical agents

Gemini
Robotics



Powering the era of physical agents with Gemini Robotics 1.5

Enabling robots to perceive, plan, think, use tools and act



CES 2026: Boston Dynamics partnering with Google DeepMind

to integrate the Gemini AI "brain" into their Atlas hardware



Designed to give humanoid robots
genuine decision-making power and
autonomy



production-ready robots intended for
industrial use, e.g. automobile factory
floors





Thank you

Gerhard Keller
Head of Manufacturing
Google Cloud DACH

