

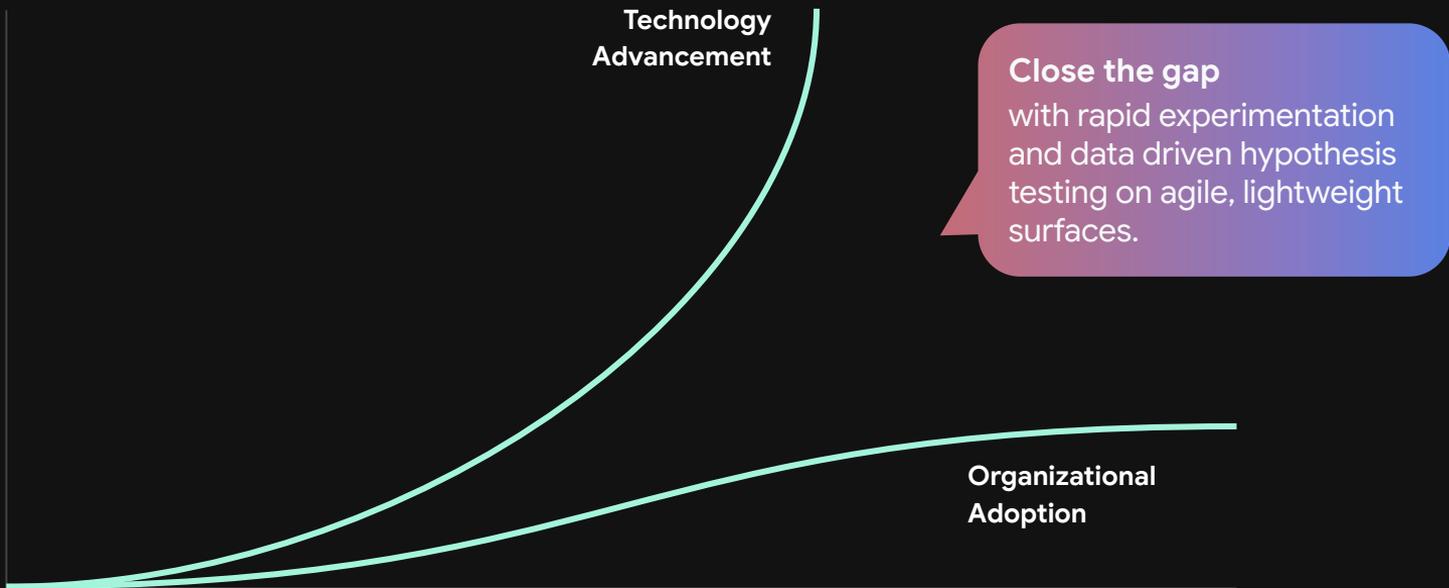
Cloud AI UX

**UX foundation, vision, strategy,
patterns, and methodology for the
neoAI/AI-Native era**

Living document
2025



Our target is to close the gap



Technology changes exponentially; organizations change logarithmically.

GenAI will accelerate 10x

Ensure users are running into the right direction.

Northstar pillars

01

Goal-oriented

Contextualized intent
Proactive & personalized
Product-led experience

02

Collaborative

Real-time & asynchronous
Seamlessly integrated
AI co-creator & collaborator

03

AI-Native

Agentic intelligence
Multimodal capabilities
Bespoke adaptability

Goal-oriented

User / system invoked

Goal-oriented

Contextualized intent

Understand user intent within their specific environment and needs.

User
invoked →

Ask anything, search your data, @mention or /tools

+ Tools Sources Auto ▶

Proactive & personalized

Offer anticipatory guidance and tailored experiences to drive progress.

System
invoked →

Insights (Prototype)

 A key prospect mentioned a competitor in an email (📧), and our internal resources on that competitor are outdated (📅), highlighting a need for immediate, accurate competitive...
Update internal competitive intelligence and brief the sales team on a key competitor.

 Several customer support emails (📧) refer to a specific product bug, but no formal bug report exists (🐛), indicating a potential widespread issue that needs official tracking an...
Create a formal bug report based on recurring customer support emails.

Product-led experience

Design the platform to intuitively guide users towards success.

System
invoked →

Recommendations (Prototype)

 Since you're using **Vertex AI Pipelines** for data preprocessing, explore **Vertex AI Model Registry** to centralize and version your models.

 Given your focus on **generative AI**, consider **Vertex AI Prompt Management** to version and evaluate your prompts effectively.

Goal-oriented

Example: Vertex

Enable user intent, provide examples, actionable insights for improvements and upgrading to advanced features.

Prototype

e.g., I want to fine-tune a language model for my company's data...



Or try one of these examples:

I want to fine-tune a language model for my company's data.

How can I deploy a custom vision model?

I need to run analysis on a large dataset.

Optimize Current Usage

Get more value and efficiency from the capabilities you are already using.



Your [Image Recognition Model X](#) has high latency; consider enabling [hardware acceleration](#) or using a [smaller model variant](#) for better performance.



You're frequently running [Notebook Y](#); did you know you can schedule it as a [Vertex AI Pipeline](#) for automation and cost savings?



Your [training job](#) recently failed; here are common [error patterns and debugging tips](#) for similar issues.



You've consumed X amount of storage for [Model Z](#); consider [archiving older versions](#) to reduce costs.

Promote Untapped Capabilities

Discover new features that can enhance your solutions and unlock new possibilities.



Since you're using [Vertex AI Pipelines](#) for data preprocessing, explore [Vertex AI Model Registry](#) to centralize and version your models.



You're deploying [custom models](#) to [Endpoints](#); consider [Vertex AI Feature Store](#) to manage and serve features more efficiently for future models.



You frequently fine-tune models; [Vertex AI Explainable AI](#) can help you understand model predictions and build trust.



Given your focus on [generative AI](#), consider [Vertex AI Prompt Management](#) to version and evaluate your prompts effectively.

Collaborative

Humans & agents collaboration

Sharing Collaboration

Invite team
members.

The screenshot shows the Google Cloud Vertex AI Studio interface. At the top, the browser address bar displays 'Google Cloud' and 'cloud-llm-preview1'. The main content area is titled 'Vertex AI / Studio / Chat playground'. A red arrow points from the 'Share' icon in the left sidebar to a 'Share project: "Risk assessment"' dialog box. The dialog box contains a search field for adding people, teams, or orgs. Below this, it lists 'People with access' with two users: Siegfried Müller (siegfried@cymbal.com) and Winnie Yang (wyang@cymbal.com), both with 'Editor' roles. Under 'General access', it shows 'Restricted' access, meaning only people within CymbalProducts can open with the link. At the bottom of the dialog are 'Copy link' and 'Done' buttons. The background shows a chat conversation with a customer review about a beach and a system instruction for Gemini 2.5 Flash Preview 5/6.

Asynchronous Collaboration

Comment panel opens up to show threads where users are collaborating.

Users can tag Gemini directly in a comment thread and dive deep with complex queries about their project.

The screenshot shows the VAI Risk assessment dashboard. At the top, there are navigation tabs: Risk assessment, Dashboard (selected), Tasks (3), Preview, and Code. A notification banner at the top left says "New model available for Loan advisor v2.0 Retrain the model with Gemini 2.5 Pro and compare" with "Dismiss" and "Compare" buttons.

The main content area is divided into several sections:

- Model performance:** A line chart showing performance over time from April 1 to April 30. The y-axis ranges from 0 to 30. The chart shows a fluctuating line with several peaks.
- Token requests with cost:** A combined bar and line chart. The x-axis shows dates from Apr 1 to Apr 30. The left y-axis is "Cost (USD)" from \$0 to \$100. The right y-axis is "Token requests" from 0 to 4000. Blue bars represent cost, and a green line represents token requests.
- Trained models:** A table listing models: General risk assessor (Gemini 2.5 pro), Loan advisor 2.0 (Gemini 2 flash), and Loan advisor 1.0.
- Prompts:** A list of prompts: "You are an expert Loan Underwriter at a commercial bank, specializing i...", "You are an expert Fraud Analyst for a high-net-worth life insurance prov...", and "You are an expert Derivatives Analyst at an investment bank, evaluating...". A "Show all" link is at the bottom.
- Data sources:** A table with columns: Name, Status, Region, Type, Items, Platform, Last updated.

Name	Status	Region	Type	Items	Platform	Last updated
loan_applications_ideaReady	Ready	us-central1	Tabular	1000	Firebase	Jul 21, 2025
loan_applications_realReady	Ready	us-central1	Tabular	1000	Firebase	Jul 21, 2025
loan_applications_syntReady	Ready	us-central1	Images	3000	Firebase	Jul 21, 2025
ground_truth	Ready	us-central1	Video	1000	Firebase	Jul 21, 2025

On the right side, a "Comments" panel is open, showing a thread:

- Janet Doe** (9:30 PM May 30): "@John Doe The model performance seems to be fluctuating. Can you take a look?"
- Jimmy** (9:35 PM May 30): "I was just about to flag this."
- Winnie** (10:30 PM May 30): "@Gemini Can you help investigate?"
- Gemini** (10:30 PM May 30): "A drop in a deployed model's performance specifically on weekends quite common. Phenomenon can typically be attributed to a few key areas: fundamental shifts in user behavior and data, and system-level activities that are often scheduled for off-peak hours. Some common reasons include:
 - Changes in User Behavior and Data (Data Drift)
 - Altered Usage Patterns
 - Scheduled Maintenance
 - Resource Scaling and Cold Starts
 Would you like me to perform a deep analysis?"

Buttons: Yes No

Collaborative - Seamlessly integrated where the work happens

On demand

How does this incident impact my 2024 forecasting?

@Lily, I looked at the current data in your 2024 Report and here's what might change:

2024 Report

Incident #274 caused **4,103 users** to drop off, up 12% from expected.
Increased drop off caused by the incident has not affected Customer NPS in premium tiers

Month	Adjusted forecast	Current forecast
Jan	High	Low
Feb	Medium	Low
Mar	High	Medium
Apr	Low	Medium
May	Medium	High

Correct Forecast View by customer tier

Powered by Looker

Scenario

An admin needs to understand the impact of an incident on their business planning.

❖ Magic moment

In seconds she gets the full picture of the impact, a solution and a fast path to update the broader group re:business impact.

Proactive

Duet.agent

@all [Urgent] There's a drop in checkout success on the e-commerce site which has reached critical level.

Increased latency for the checkout service **Critical**

Impacted Projects: checkout-workload
Impacted Resources: Payment Gateway
Impacted Products: GKE; App Engine
Cymbal Shopping cart service
Required agents: SWE.agent; SRE.agent
Started: Nov 2, 2023 at 9:09:20PM UTC

→ Support ticket #12345 [View ticket](#)

The rollback was successful and the rollback analysis is ready for review

Scenario

An SRE gets a pagerduty alert of an outage. Agent is already working on the root-cause analysis and has proactively performed a rollback.

❖ Magic moment

AI agent orchestrates the SRE.agent (incident analysis) and SWE.agent (code rollback) and proactively prevents a longer outage.

Collaborate & orchestrate: Human & AI teams

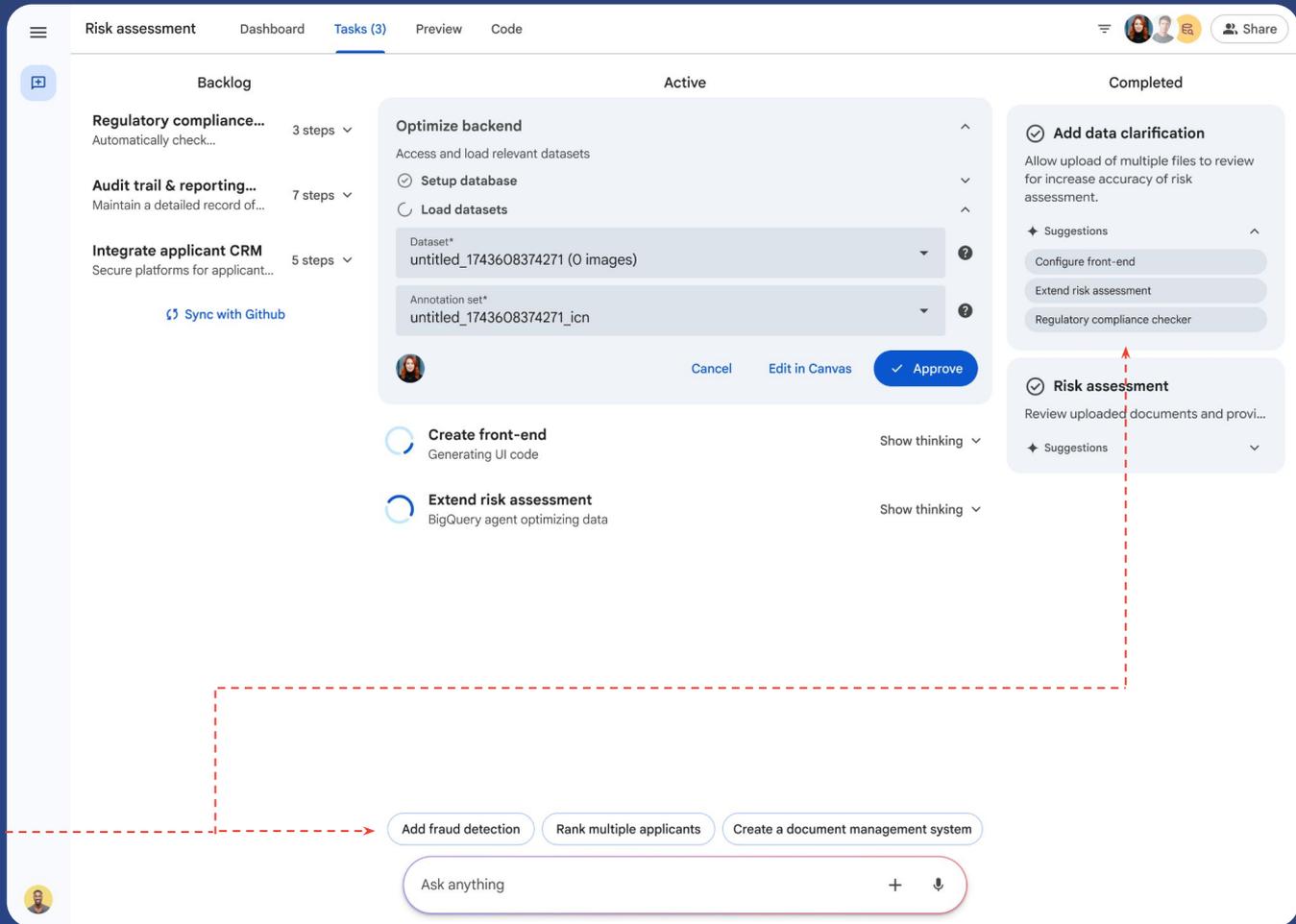
Fast forward

The project is in full motion and the user can observe the pending tasks and unblock tasks which need human input.

PLG

Smart suggestions for adding new tasks to your project backlog and identifying follow-up actions for completed items.

[go/vertex-ai-native](https://go.vertex-ai-native)



The screenshot displays a project management interface for a "Risk assessment" project. The interface is divided into three main sections: Backlog, Active, and Completed.

- Backlog:** Lists three tasks with their respective step counts and descriptions:
 - Regulatory compliance...** (3 steps): Automatically check...
 - Audit trail & reporting...** (7 steps): Maintain a detailed record of...
 - Integrate applicant CRM** (5 steps): Secure platforms for applicant...A "Sync with Github" button is located below the backlog items.
- Active:** Shows the current task "Optimize backend" (Access and load relevant datasets) with sub-tasks:
 - Setup database** (checked)
 - Load datasets** (unchecked): Includes a "Dataset*" dropdown with "untitled_1743608374271 (0 images)" and an "Annotation set*" dropdown with "untitled_1743608374271_icn".Buttons for "Cancel", "Edit in Canvas", and "Approve" are visible. Below this, two other active tasks are shown:
 - Create front-end** (Generating UI code) with a "Show thinking" dropdown.
 - Extend risk assessment** (BigQuery agent optimizing data) with a "Show thinking" dropdown.
- Completed:** Shows two finished tasks:
 - Add data clarification** (Allow upload of multiple files to review for increase accuracy of risk assessment). Includes a "Suggestions" dropdown with items like "Configure front-end", "Extend risk assessment", and "Regulatory compliance checker".
 - Risk assessment** (Review uploaded documents and provi...). Includes a "Suggestions" dropdown.

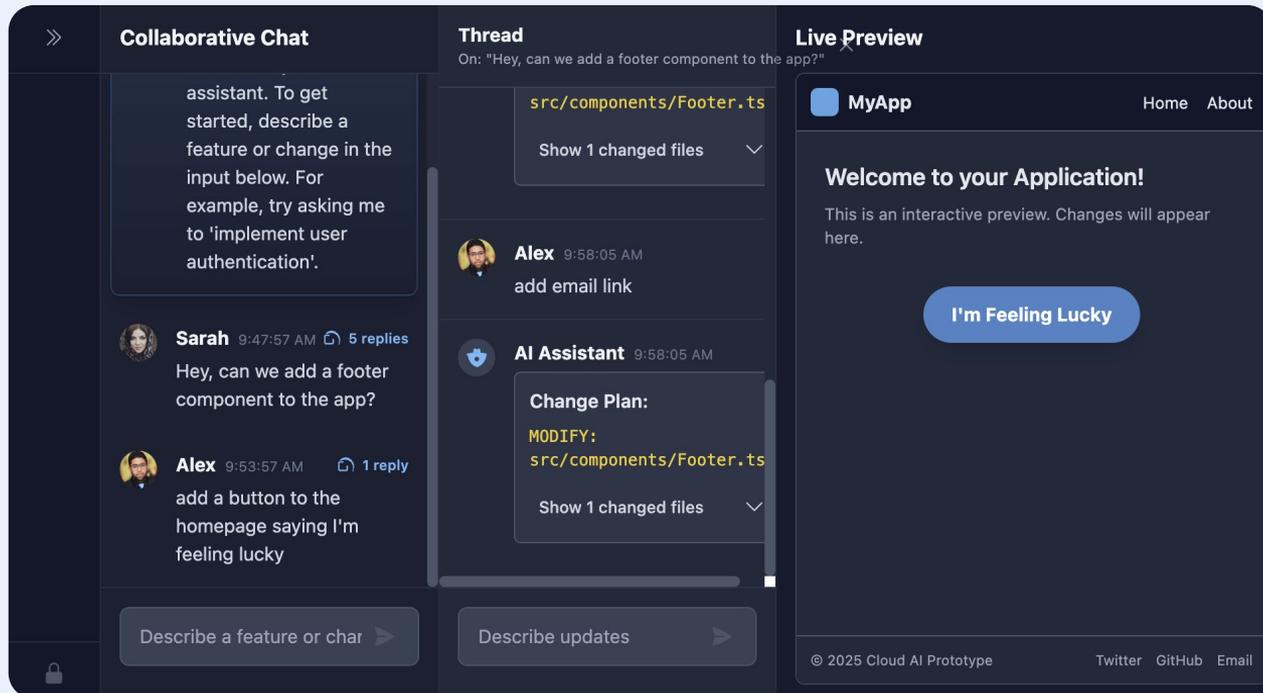
At the bottom of the interface, there is a search bar with the text "Ask anything" and a microphone icon. Below the search bar, three suggested actions are displayed: "Add fraud detection", "Rank multiple applicants", and "Create a document management system".

Red dashed arrows indicate the flow of information: one arrow points from the "Add fraud detection" suggestion to the "Backlog" section, and another points from the "Add data clarification" completed task to the "Backlog" section.

Realtime Collaboration

Exploration: Real-time chat with threaded iteration and history

This system enables real-time, collaborative software development through a chat interface where users submit prompts for features or changes. An AI "Change Planning Engine" identifies specific code modifications, acquiring locks on affected files to manage concurrent updates by either merging conflicting prompts or intelligently applying new changes over recently completed ones.



Prototype

AI-Native

Vision, Concepts, and Patterns

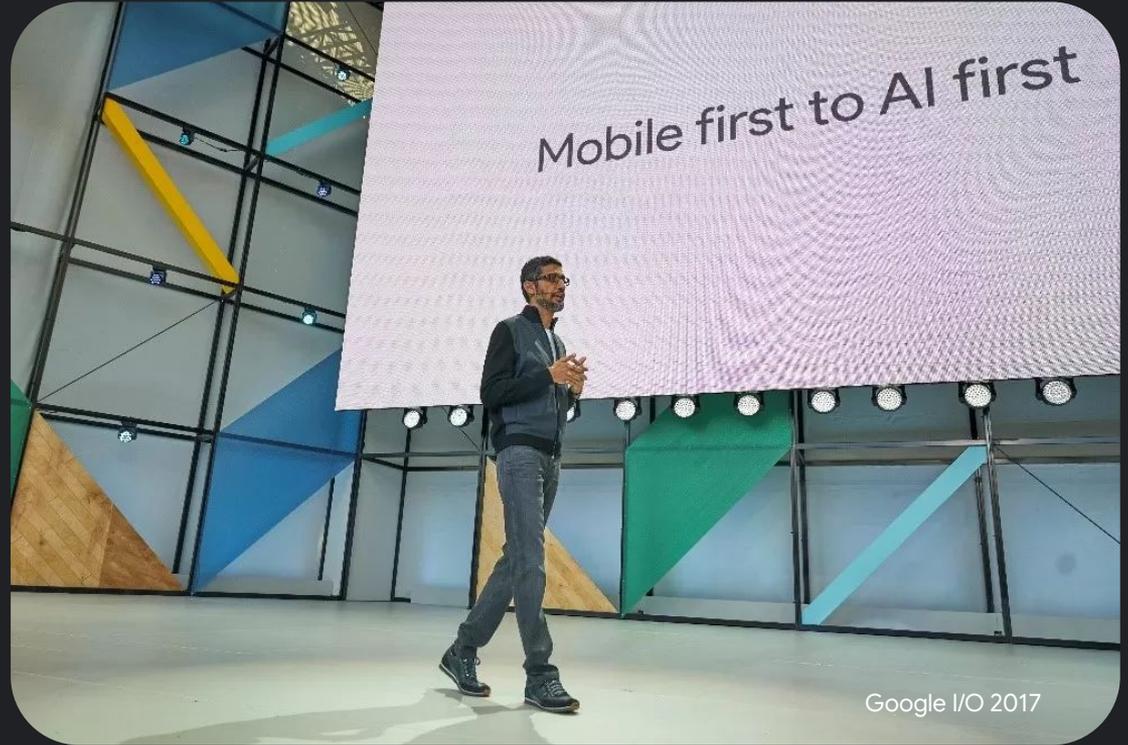
AI-First is dated (2017)

Traditional:

AI as an add-on or feature

AI-First:

Prioritizing AI capabilities, but often with a focus on specific AI tasks



**Do we agree that the software
of the future will look fundamentally
different from the software of today?**

AI-Native

Intelligent, adaptive, and agentic AI is deeply integrated into the core of a system or platform. This integration enables a fundamental shift in functionality and interaction, characterized by:

- **Contextual intelligence**
Understands user intent and context implicitly for relevant and efficient responses.
- **Multimodal capabilities**
Processes and generates information across diverse modalities (text, image, audio, video) for richer understanding and natural interaction.
- **Anticipatory design**
Predicts user needs and offers proactive assistance, streamlining tasks and improving efficiency.
- **Adaptive systems**
Learns and evolves from data and interactions, continuously tailoring behavior and responses over time.
- **Intelligent workflows**
Seamlessly integrates AI into core processes, automating tasks and providing intelligent guidance.
- **Transparency & explainability**
Provides clear insights into AI's actions and rationale, building trust and user control.

Agentic Experiences

On Demand

Proactive

Chatbot

Prompt

Response

Refine

Response

Semi-autonomous Agents

Intent/Goal

Solution planning

Option A

Option B

Solution planning

Solution

Long-running Agents

Setup

System event

Solution planning

Action A

(outside constraints)

Semi-autonomous

Action B

(e.g. autoscale w/ constraints)

Autonomous

Report



Agents pillars

01

Agents Everywhere

Unify Agents across surfaces in Google Cloud, Workspace, and beyond, with shared context enabling cohesive cross-product user journeys. Agents will act as intelligent assistants, capable of performing tasks for individuals, teams, and entire companies by leveraging a shared understanding of context, resources, and relationships.

Links: [paper](#), [deck](#)

02

OneMCP

OneMCP is a common infrastructure framework developed by Google Cloud to streamline the creation, discovery, and secure management of Model Context Protocol (MCP) tools and AI agents across Google, Google Cloud, and partner ecosystems.

Links: [go/gcp-mcp-eng-pitch](#), [go/onemcp-concept](#)
[go/onemcp-prd](#), [go/onemcp-ug](#)

Agent Discovery: A Scalable Approach

- **Seamless Goal-Oriented Actions:** Intelligently act on user intents directly, minimizing the need for explicit agent choices.
- **Scale Agent Selection:** Optimize agent selection to handle a massive number of agents and user intents.
- **Empower User Choice:** Enable users to select, discover (PLG), compare, and run agents to determine the best fit.
- **Adapt in Real-Time:** Use the current conversation state to recommend alternative agents intelligently.

I'm making a new design for the G-shoe app. I'm not sure what guidelines to use for the conversational tone. Can you give me a suggestion?

There are 6 different agents that can help you answer this question, but only two of them are likely to be the one you are looking for. I've collected responses from each of them so you can select which one.



CE Elements Agent

Best answer ?⁺

For the G-shoe app, the conversational tone should align with its target audience and purpose. Here are a few tone guidelines to consider:



Material design agent

2nd best ?⁺

For the G-shoe app, the conversational tone should align with its target audience and purpose. Here are a few tone guidelines to consider:



CE Elements agent selected

Thin surfaces: Reimagining user experience

End-to-end CUJs on a single bespoke surface

From thick to thin

Shift from traditional, complex "thick" UIs to streamlined, AI-powered "thin" surfaces.

Strategic focus

Enables users to concentrate on "what" and "why," delegating tactical tasks to AI.

Fallback option

"Thick" UI remains accessible when needed, providing a safety net.

Agile & fast

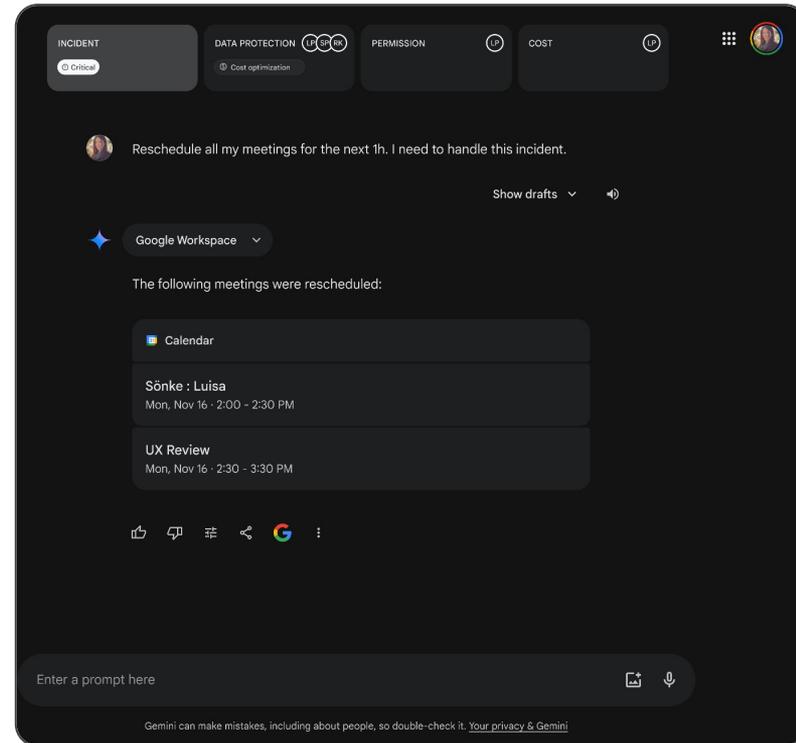
Highly adaptable to evolving user needs and product directions, enabling faster velocity and time to market.

Evolution over time

Enables thick surface KTLO, which overtime gets EOL as Thin Surfaces reach feature parity.

Benefits

Reduced cognitive load, faster time to value.



Cheap changes everything: Generative UI/UX

Traditional UI/UX: Relies on static, pre-designed interfaces that aim to serve a broad range of users and use cases. This can lead to complexity and friction for specific, nuanced tasks.

Generative UI/UX: Enables the on-demand creation of bespoke, context-aware interfaces tailored precisely to the user's current goal and the task at hand.

- **Hyper-personalization at scale**
Every user interaction can be met with a uniquely generated interface optimized for their intent, background, and preferences.
- **Task-optimized workflows**
Instead of navigating generic menus, users are presented with a direct, streamlined path to complete their specific task, reducing cognitive load and time-to-value.
- **Dynamic and adaptive interfaces**
The UI can change and adapt in real-time as the user's needs evolve or new information becomes available.
- **Reduced development overhead**
For many standard use cases, the need for manual UI development for every permutation of a task can be significantly reduced.

From commands to conversations

Unlocking natural interaction and enhanced productivity.

Imagine achieving your goals in software as easily as talking to a colleague. By blending voice commands with parallel dynamically generated companion interfaces, we're creating a revolutionary way to interact with software.

[go/voice-companion-ui](#)

Assistant (Commands)	 Consultant (Conversation)
Users must translate goals into specific commands.	Users describe their goal naturally (w/ voice and other modalities)
High cognitive load: Requires understanding the system's command structure.	Low cognitive load: Focus on the goal, not the how.
Limited flexibility: Constrained by predefined commands.	Synergistic Voice and GUI: While the user expresses their intent via voice, the underlying GenUI engine simultaneously generates a corresponding, contextually relevant UI.

Experiences constructs

AI/Agentic systems need more than just chat for complex, asynchronous tasks. We can organize these systems into three interacting levels:

	Altitude	Experience	Function
Strategic level	Earth view	Kanban	Orchestration, observation, Human in the loop
Tactical level	Maps view	Chat Canvas CLI	Definition, Refinement, Reasoning
Operational level	Street view	Embedded	Implementation, Execution

Tactical Chat

A focused,
dialog-based
interaction model to
guide task execution
one step at a time
with single agents.

The screenshot displays the Google Agentspace interface. On the left is a dark sidebar with navigation options: 'New', 'Search', 'Starred', 'Agents', 'Data Insights', 'Idea Generation', 'Deep Research', and a 'Recents' section containing 'Cloud AI in Google Cloud', 'Functional calling best ...', 'Data analysis capabilities', 'Test', and 'Carbon-free car pitch'. At the bottom of the sidebar, it shows 'SF, Calif, US' and 'From your IP address' with a 'Learn more' link. The main area is titled 'Google Agentspace' and contains a chat window. The chat header shows 'What is Cloud AI withing Google Cloud?'. The chat content includes a 'Show thinking' indicator, a response from the agent explaining Google Cloud AI's purpose, a section titled 'Key Components and Services:', and a list of services including Vertex AI and Vertex AI Workbench. At the bottom of the chat is a text input field with the placeholder 'Ask anything, search your data, @mention or /tools' and buttons for '+', 'Tools', 'Data (all)', 'Auto', and a right arrow. A disclaimer at the very bottom states: 'Generative AI may display inaccurate information, including about people, so double-check its responses.'

Tactical Canvas

A hybrid experience that combines assistive chat with direct manipulation of an artifact (e.g. code), enabling fluid, in-context collaboration.

The screenshot displays the Gemini Advanced interface. At the top, it says "Gemini Advanced" and "2.5 Flash". A search bar contains the text "python for gemini api". Below this, there is a "Show thinking" dropdown menu. The main chat area contains the following text:

I can help you with that! Below is a Python code example demonstrating how to use the Gemini API for text generation.

This document will provide you with the necessary steps to get started with the Gemini API in Python, including installation and a simple text generation example.

Below the chat is a code editor window titled "Gemini API with Python: Text Generation" with a timestamp of "May 12, 1:17 PM". The code editor contains the following Python code:

```
22 # Specify the generative model you want to use.
23 # 'gemini-2.0-flash' is a fast and versatile model suitable for various tasks.
24 # You can explore other available models by calling genai.list_models().
25 model = genai.get_generative_model(model='gemini-2.0-flash')
26
27 # Define the prompt for text generation.
28 prompt = "Tell me a short story about a mischievous cat named Whiskers."
29
30 # Generate content using the specified model and prompt.
31 # The generate_content method sends the prompt to the Gemini API.
32 # The response object contains the generated content.
33 try:
34     response = model.generate_content(prompt)
35
36     # Print the generated text.
37     # The 'text' attribute of the response object holds the model's output.
38     print("Generated Story:")
39     print(response.text)
40
41 except Exception as e:
42     print(f"An error occurred: {e}")
43     print("Please ensure your API key is correct and you have network connectivity.")
44     print("Also, check if the 'gemini-2.0-flash' model is accessible with your API key.")
45
46
```

At the bottom of the interface, there is an "Ask Gemini" input field with a "Canvas" button, a "Search" button, and a microphone icon. A footer note states: "Gemini can make mistakes, so double-check it".

Tactical CLI

Provides direct, **local access** for power users and developers. It allows for precise control, scripting, and direct interaction with local files, serving as an efficient and unmediated interface for advanced AI tasks.

```
sonke-mac:aiuxsummit sonke$ npx @google/gemini-cli



Tips for getting started:
1. Ask questions, edit files, or run commands.
2. Be specific for the best results.
3. Create GEMINI.md files to customize your interactions with Gemini.
4. /help for more information.

> █ Type your message or @path/to/file

~/aiuxsummit          no sandbox (see /docs)          gemini-2.5-pro (100% context left)
```

Operational Embedded

AI support delivered exactly where the work is happening, within the artifact itself, for precise assistance and manipulation.

Features

Activity Tracking:

Automatic step counting (using device sensors).

GPS



Ask Gemini

Speed, pace, route mapping).

Manual logging for various workout types (gym, yoga, swimming, sports, etc.) with and estimated calorie burn.

Integration with popular wearables (e.g., Fitbit, Garmin, Apple Watch).

Heart rate monitoring data import (if available).

The screenshot displays the Vertex AI Risk assessment dashboard. At the top, there's a notification: "New model available for Loan advisor v2.0. Retrain the model with Gemini 2.5 Pro and compare." Below this, the "Model performance" section shows a line graph of performance over time from April 1 to April 30. The "Token requests" section features a bar chart showing requests per day from April 1 to April 30. The "Trained models" section lists three models: "General risk assessor" (Gemini 2.5 pro), "Loan advisor 2.0" (Gemini 2 flash), and "Loan advisor 1.0". The "Data sources" section contains a table with the following data:

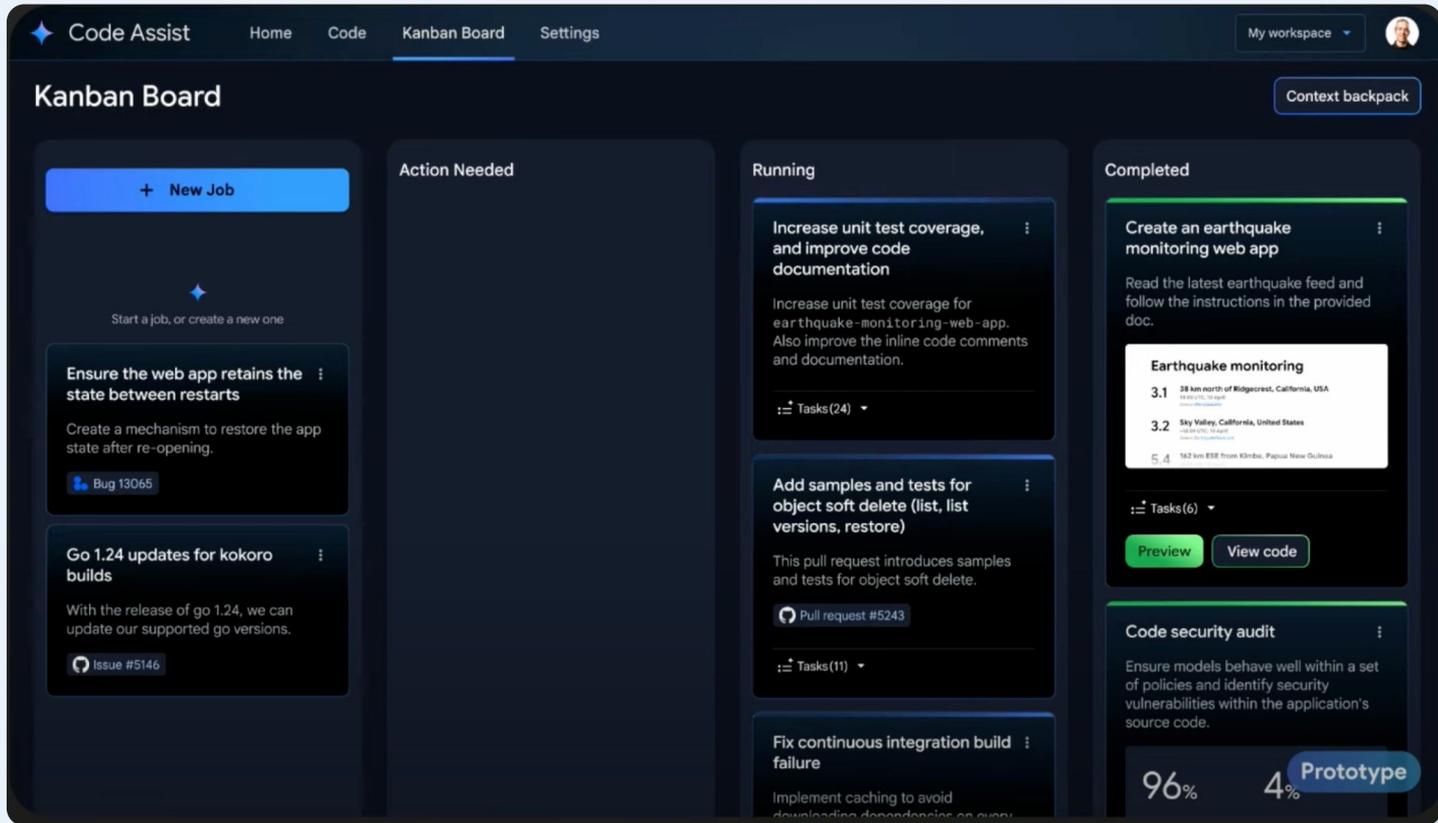
Name	Status	Region	Type	Items	Platform	Last updated
loan_applications_ideal	Ready	us-central1	Tabular	1000	Firebase	Jul 21, 2025
loan_applications_real	Ready	us-central1	Tabular	1000	Firebase	Jul 21, 2025
loan_applications_synthetic	Ready	us-central1	Images	3000	Firebase	Jul 21, 2025
ground_truth	Ready	us-central1	Video	1000	Firebase	Jul 21, 2025

Strategic Kanban

A high-level environment to orchestrate and execute multiple tasks or goals (e.g. project) augmented by agents.

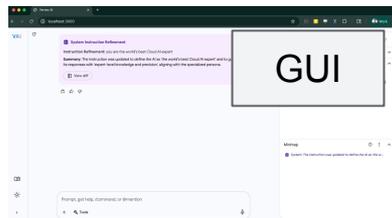
Example: Gemini Code Assist w/ SDLC (Software Development Lifecycle) agents.

Google Cloud Next 25
Developer Keynote



Surface orchestration

WIP



GUI

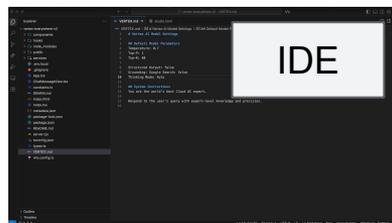
Goal Loop
iterate on specs to enable agents to achieve goals.

SPEC.md



CLI

Agent Loop
Executes commands aligned with specs



IDE

Validation Loop
review generated code and tweak if necessary

Principles

Entropy, ...

Entropy - reframing autonomy

The degree of unpredictability and complexity associated with a task, influencing the appropriate level of human or AI involvement.

Low entropy

Predictable, repetitive tasks with a low probability of error.

Example: Boilerplate code, routine data preprocessing steps, implementing predefined evaluation metrics.

Low-risk actions, eliminating manual effort and minimizing the need for human oversight.

High entropy

Complex, creative tasks requiring human judgment and ingenuity.

Example: Designing novel model architectures, debugging intricate model training issues, optimizing hyperparameters, architecting end-to-end ML pipelines.

Empower users to focus on these high-entropy challenges by eliminating distractions.

Methodology

Human-Centered AI (HCAI) Design

AI Designer

- Just like designers have to know their users, this new person needs to know the new alien they're partnering with. That means they need to be just as **obsessed about hanging out with models** as they are with talking to users.
- The only way to really understand how we want the model to behave in our application is to **build a bunch of prototypes** that demonstrate different model behaviors. This — and a need to have good intuition for the possible — means this person needs enough technical fluency to look kind of like an engineer.
- Each of the behaviors you're trying to design have near limitless possibility that you have to wrangle into a single, shippable product, and there's little to no prior art to draft off of. That means this person needs **experience facing the kind of "blank page"** existential ambiguity that founders encounter.

The only constant is change

The UX profession, as many other professions, is going through massive changes fueled by the AI paradigm shift.

UXD

TODO

UXR

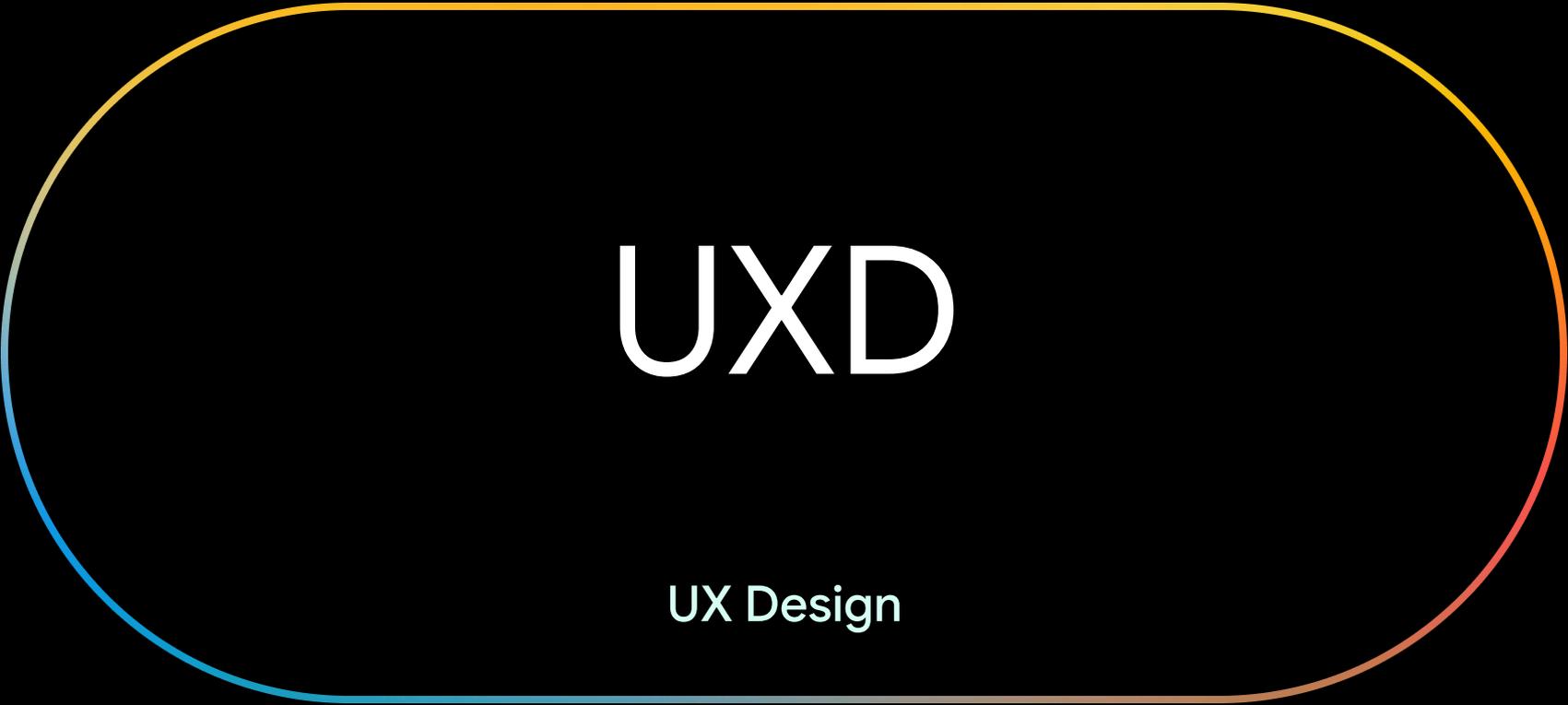
TODO

UXW

TODO

etc

TODO



UXD

UX Design



UXR

UX Research

Studies

Gemini-First

TODO

Thank you



Appendix

Mindset Shift

Mindset shifts

Pursuing 10X growth is challenging because the path to a 50% increase is clearer, more predictable, carries less risk and aligns with existing thinking. It avoids the need for major strategic shifts or high-stakes decisions.

Choosing the familiar over the bold: The 10X dilemma

Mindset shifts

50% Improvement – Cater to an established mindset

Flickr

For serious photographers, focused on quality, tags, sets, group discussions.
Mindset: photography as an art form or hobby, shared in curated galleries.

Nokia (Symbian)

Added smartphone features to keypad-era phones.
Mindset: phones are for calls and texts, everything else is optional.

BlackBerry Messenger (BBM)

Hardware-locked, secure, and enterprise-focused.
Mindset: messaging is private, business-first, device-dependent.

MapQuest

Printed static directions from a website.
Mindset: plan trips in advance, offline.

Traditional software

One-size-fits-all, feature-heavy, generalized, maintainability and scale.
Mindset: Software is a tool to learn and adapt to. Apps solve task (fragments)

10X Disruption – Embrace new behaviors, tech, or culture

Instagram

Mobile-first, fast sharing, social and casual.
Mindset: photography as social communication, fast and disposable.

iOS/Android

Designed around apps, touchscreens, always-connected use.
Mindset: phones are mobile computers.

WhatsApp

Free, cross-platform, fast, personal.
Mindset: messaging is universal, instant, and casual.

Google Maps

Live, GPS-based, adaptive navigation.
Mindset: real-time travel, mobile convenience.

New software

Modular, lightweight, personalized, context aware. **Outcomes, not apps.**
Mindset: Software understands intent, adapts to context, helps reach goals

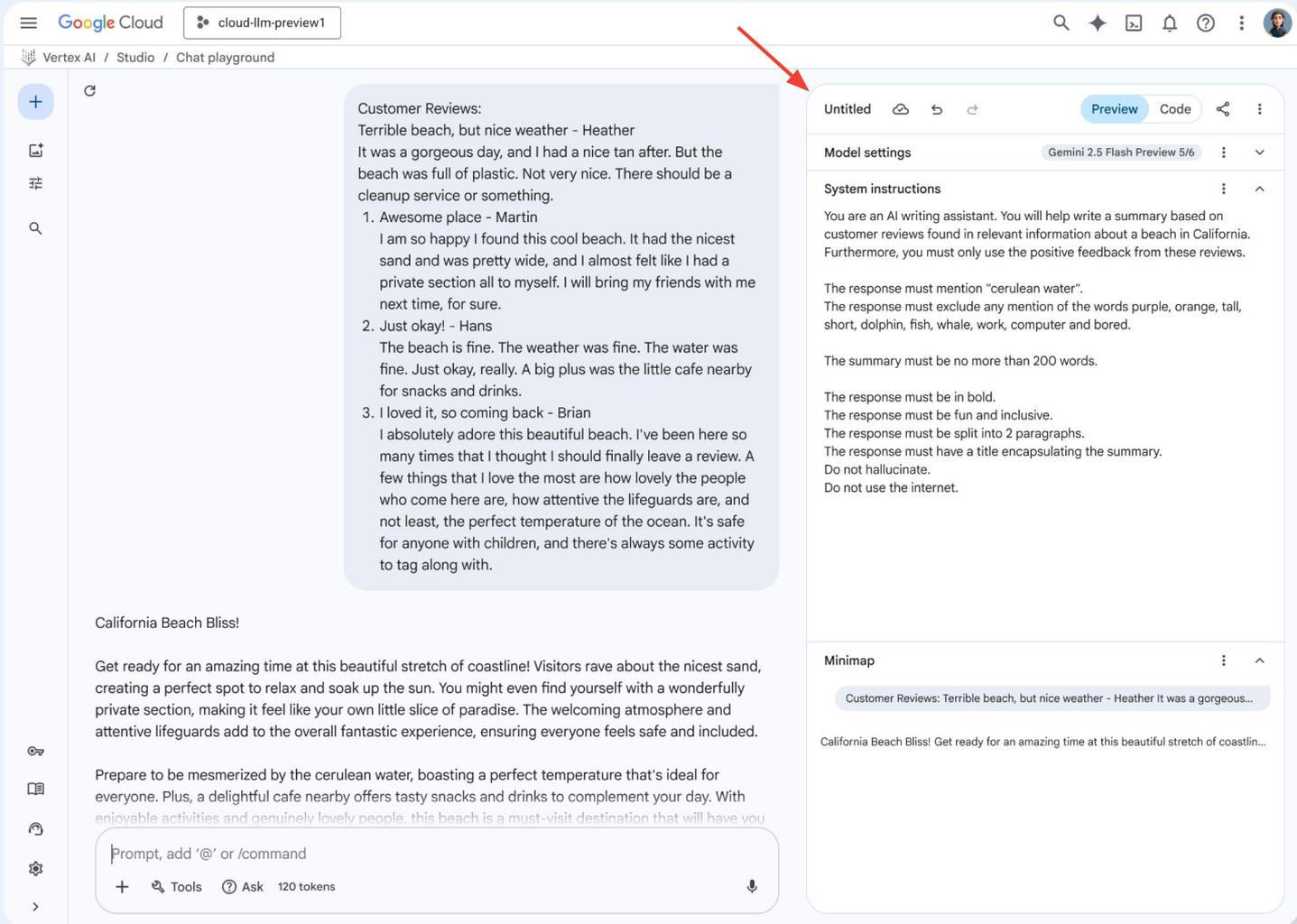
While both paths represent growth, pursuing 10X growth is challenging because the path to a 50% increase is clearer, more predictable, carries less risk and aligns with existing thinking. It avoids the need for major strategic shifts or high-stakes decisions.

UX Patterns

UX Patterns

Canvas

The canvas component is adaptable for anything that needs visual display or interactions (e.g. code editor, settings, form fields, app preview, doc preview, etc)



The screenshot shows the Google Cloud Vertex AI Studio Chat playground interface. The top navigation bar includes the Google Cloud logo, the text "cloud-llm-preview1", and utility icons for search, navigation, and user profile. The main content area is divided into three sections:

- Input/Output Area:** Contains a text input field with the prompt "Prompt, add '@' or /command" and a "Tools" button. Below it, a "Preview" button is highlighted with a red arrow. The output area displays a list of customer reviews and a summary.
- Model settings:** Shows the selected model as "Gemini 2.5 Flash Preview 5/6".
- System instructions:** Contains detailed instructions for the AI assistant, such as "You are an AI writing assistant. You will help write a summary based on customer reviews found in relevant information about a beach in California." and "The response must be no more than 200 words."
- Minimap:** Provides a visual overview of the entire chat content.

The output text in the main area includes:

Customer Reviews:
Terrible beach, but nice weather - Heather
It was a gorgeous day, and I had a nice tan after. But the beach was full of plastic. Not very nice. There should be a cleanup service or something.

1. Awesome place - Martin
I am so happy I found this cool beach. It had the nicest sand and was pretty wide, and I almost felt like I had a private section all to myself. I will bring my friends with me next time, for sure.
2. Just okay! - Hans
The beach is fine. The weather was fine. The water was fine. Just okay, really. A big plus was the little cafe nearby for snacks and drinks.
3. I loved it, so coming back - Brian
I absolutely adore this beautiful beach. I've been here so many times that I thought I should finally leave a review. A few things that I love the most are how lovely the people who come here are, how attentive the lifeguards are, and not least, the perfect temperature of the ocean. It's safe for anyone with children, and there's always some activity to tag along with.

California Beach Bliss!

Get ready for an amazing time at this beautiful stretch of coastline! Visitors rave about the nicest sand, creating a perfect spot to relax and soak up the sun. You might even find yourself with a wonderfully private section, making it feel like your own little slice of paradise. The welcoming atmosphere and attentive lifeguards add to the overall fantastic experience, ensuring everyone feels safe and included.

Prepare to be mesmerized by the cerulean water, boasting a perfect temperature that's ideal for everyone. Plus, a delightful cafe nearby offers tasty snacks and drinks to complement your day. With enjoyable activities and genuinely lovely people, this beach is a must-visit destination that will have you

UX Patterns

/commands

Slash commands within the chat input field allow the user to invoke advanced actions or shortcuts (e.g. change the model, refine a prompt, clear the conversation, etc)

The screenshot shows the Google Cloud Vertex AI Studio Chat playground interface. The top navigation bar includes the Google Cloud logo, the project name 'cloud-llm-preview1', and various utility icons. The main chat area displays a conversation about beach reviews. A dropdown menu is open over the chat input field, listing several slash commands:

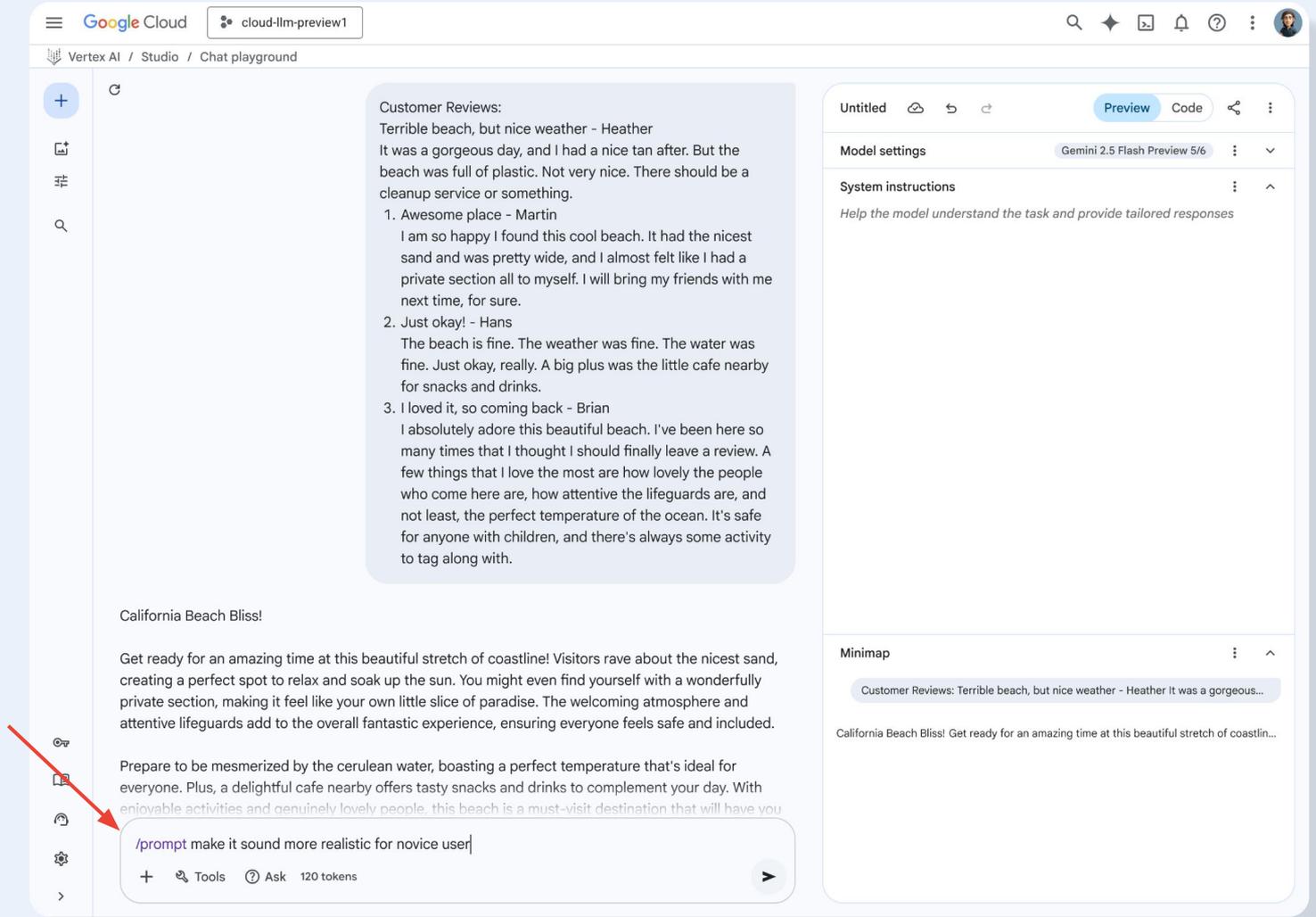
- `/model [describe]`: Change the model
- `/prompt [feedback]`: Improve your prompt
- `/system-instructions [feedback]`: Set or optimize system instructions
- `/ask [question]`: Google Cloud & general help
- `/clear`: Clear conversation

A red arrow points to the slash command input field. The right sidebar contains settings for the current model (Gemini 2.5 Flash Preview 5/6), system instructions, and a minimap of the chat history.

UX Patterns

/commands

Slash commands within the chat input field allow the user to invoke advanced actions or shortcuts (e.g. change the model, refine a prompt, clear the conversation, etc)



Google Cloud cloud-llm-preview1

Vertex AI / Studio / Chat playground

Customer Reviews:

Terrible beach, but nice weather - Heather
It was a gorgeous day, and I had a nice tan after. But the beach was full of plastic. Not very nice. There should be a cleanup service or something.

1. Awesome place - Martin
I am so happy I found this cool beach. It had the nicest sand and was pretty wide, and I almost felt like I had a private section all to myself. I will bring my friends with me next time, for sure.
2. Just okay! - Hans
The beach is fine. The weather was fine. The water was fine. Just okay, really. A big plus was the little cafe nearby for snacks and drinks.
3. I loved it, so coming back - Brian
I absolutely adore this beautiful beach. I've been here so many times that I thought I should finally leave a review. A few things that I love the most are how lovely the people who come here are, how attentive the lifeguards are, and not least, the perfect temperature of the ocean. It's safe for anyone with children, and there's always some activity to tag along with.

California Beach Bliss!

Get ready for an amazing time at this beautiful stretch of coastline! Visitors rave about the nicest sand, creating a perfect spot to relax and soak up the sun. You might even find yourself with a wonderfully private section, making it feel like your own little slice of paradise. The welcoming atmosphere and attentive lifeguards add to the overall fantastic experience, ensuring everyone feels safe and included.

Prepare to be mesmerized by the cerulean water, boasting a perfect temperature that's ideal for everyone. Plus, a delightful cafe nearby offers tasty snacks and drinks to complement your day. With enjoyable activities and genuinely lovely people, this beach is a must-visit destination that will have you

`/prompt make it sound more realistic for novice user`

+ Tools Ask 120 tokens

Untitled Preview Code

Model settings Gemini 2.5 Flash Preview 5/6

System instructions Help the model understand the task and provide tailored responses

Minimap

Customer Reviews: Terrible beach, but nice weather - Heather It was a gorgeou...

California Beach Bliss! Get ready for an amazing time at this beautiful stretch of coastlin...

UX Patterns

Minimap

The conversation minimap makes it easy to traverse a long conversation by condensing it small into single line entries that are clickable to autoscroll to that point.

The screenshot displays the Google Cloud Vertex AI Studio Chat playground interface. At the top, the Google Cloud logo and the project name 'cloud-llm-preview1' are visible. The main chat area contains a purple instruction box: 'Help me write' followed by a Lorem ipsum placeholder. Below this is a large blue box containing 'Customer Reviews' with three entries: Heather, Hans, and Brian, each with a short paragraph of text. The chat history shows a response from the AI: 'California Beach Bliss! Get ready for an amazing time at this beautiful stretch of coastline! Visitors rave about the nicest sand, creating a perfect spot to relax and soak up the sun. You might even find yourself with a wonderfully...'. At the bottom, there is a text input field with a placeholder 'Prompt, add '@' or /command' and buttons for '+', 'Tools', and 'Ask'. On the right side, a 'Minimap' panel is open, showing a condensed version of the chat history as a single line: 'Help me write: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod...'. A red arrow points from the 'Minimap' panel back to the main chat area, illustrating the navigation functionality.

Google Cloud cloud-llm-preview1

Vertex AI / Studio / Chat playground

Help me write

Instruction: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Customer Reviews:

Terrible beach, but nice weather - Heather
It was a gorgeous day, and I had a nice tan after. But the beach was full of plastic. Not very nice. There should be a cleanup service or something.

1. Awesome place - Martin
I am so happy I found this cool beach. It had the nicest sand and was pretty wide, and I almost felt like I had a private section all to myself. I will bring my friends with me next time, for sure.
2. Just okay! - Hans
The beach is fine. The weather was fine. The water was fine. Just okay, really. A big plus was the little cafe nearby for snacks and drinks.
3. I loved it, so coming back - Brian
I absolutely adore this beautiful beach. I've been here so many times that I thought I should finally leave a review. A few things that I love the most are how lovely the people who come here are, how attentive the lifeguards are, and not least, the perfect temperature of the ocean. It's safe for anyone with children, and there's always some activity to tag along with.

California Beach Bliss!

Get ready for an amazing time at this beautiful stretch of coastline! Visitors rave about the nicest sand, creating a perfect spot to relax and soak up the sun. You might even find yourself with a wonderfully

Prompt, add '@' or /command

Tools Ask

Untitled Preview Code

Model settings Gemini 2.5 Flash Preview 5/6

System instructions

You are an AI writing assistant. You will help write a summary based on customer reviews found in relevant information about a beach in California. Furthermore, you must only use the positive feedback from these reviews.

The response must mention "cerulean water".
The response must exclude any mention of the words purple, orange, tall, short, dolphin, fish, whale, work, computer and bored.

The summary must be no more than 200 words.

The response must be in bold.
The response must be fun and inclusive.
The response must be split into 2 paragraphs.
The response must have a title encapsulating the summary.
Do not hallucinate.
Do not use the internet.

Minimap

Help me write: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod...

Customer Reviews: Terrible beach, but nice weather - Heather It was a gorgeous...

California Beach Bliss! Get ready for an amazing time at this beautiful stretch of coastlin...

UX Patterns

In-app help

Users can ask general or documentation related questions when switching to Ask mode or typing “/ask”. Answers appear visually different in the conversation view.

The screenshot displays the Google Cloud Vertex AI Studio Chat playground interface. The top navigation bar includes the Google Cloud logo, a project ID 'cloud-llm-preview1', and various utility icons. The main content area shows a conversation with a question and a detailed answer. A red arrow points to the 'Ask anything' header in the conversation view. The right sidebar shows model settings and system instructions.

Google Cloud cloud-llm-preview1

Vertex AI / Studio / Chat playground

Ask anything

Question: how can i lorem ipsum?

Answer: To ground your data in Google Cloud's Vertex AI, you'll need to connect your generative AI model to verifiable information sources. This process helps reduce hallucinations, anchors responses to your data, and provides auditability. Here's a breakdown of what you need:

- 1. Choose a Grounding Method:**
 - **Google Search:** Use publicly available web data to provide a wide range of knowledge to your model. If using in production, you'll need a Google Search entry point.
 - **Vertex AI Search Data Store:** Use your own data, such as website content, unstructured documents, or structured data, stored within Vertex AI Search.
 - **Elasticsearch:** Utilize existing Elasticsearch indexes for retrieval-augmented generation (RAG).
- 2. Data Preparation (if using Vertex AI Search or Elasticsearch):**
 - **Create a Data Store:** If using Vertex AI Search, create a data store and add your data (website data or documents).
- 3. Enable the API and Permissions:**
 - Make sure the AI Applications API is enabled.
 - In the Google Cloud console, go to the IAM page, and search for the `discoveryengine.servingConfigs.search` permission. This is required for the grounding service to work.
- 4. Create a Search Application (if using Vertex AI Search):**
- 5. Configure Grounding in Vertex AI Studio:**
 - In Vertex AI Studio, go to the Freeform page.
 - Enable grounding and select your data source (Vertex AI Search or Elasticsearch).
 - Specify the data store ID (for Vertex AI Search) or Elasticsearch endpoint and index details.

By following these steps, you can effectively ground your data and improve the accuracy of your generative AI applications.

Sources:

- [Web Grounding for Enterprise](#) | [Generative AI on Vertex AI](#) | [Google Cloud](#)
- [Gen AI](#) | [Generative AI](#) | [Google Cloud](#)
- [Grounding with Elasticsearch](#) | [Generative AI on Vertex AI](#) | [Google Cloud](#)

Untitled **Preview** Code

Model settings Gemini 2.5 Flash Preview 5/6

System instructions

Help the model understand the task and provide tailored responses

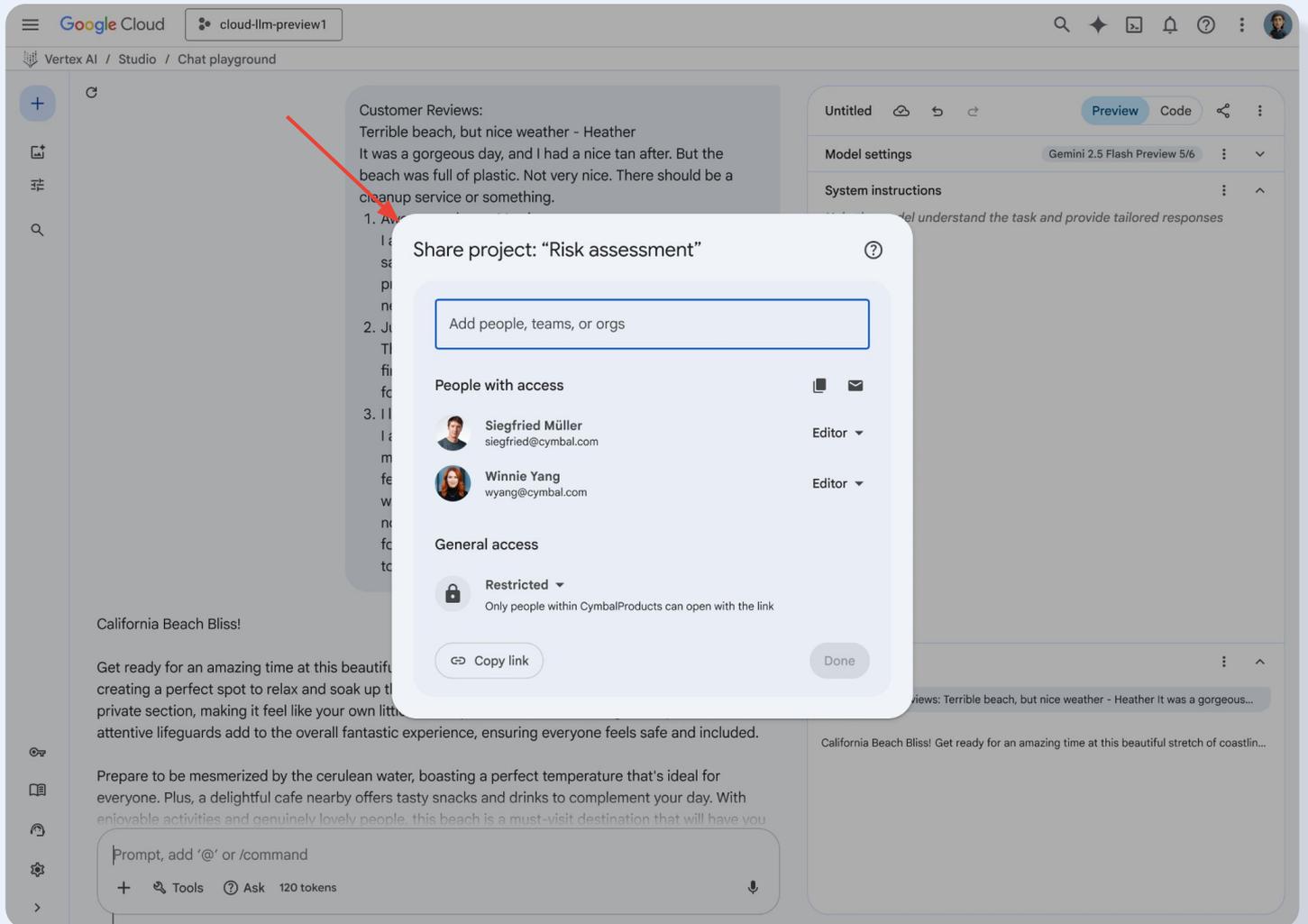
Minimap

Ask anything about Vertex AI

Tools Ask 120 tokens

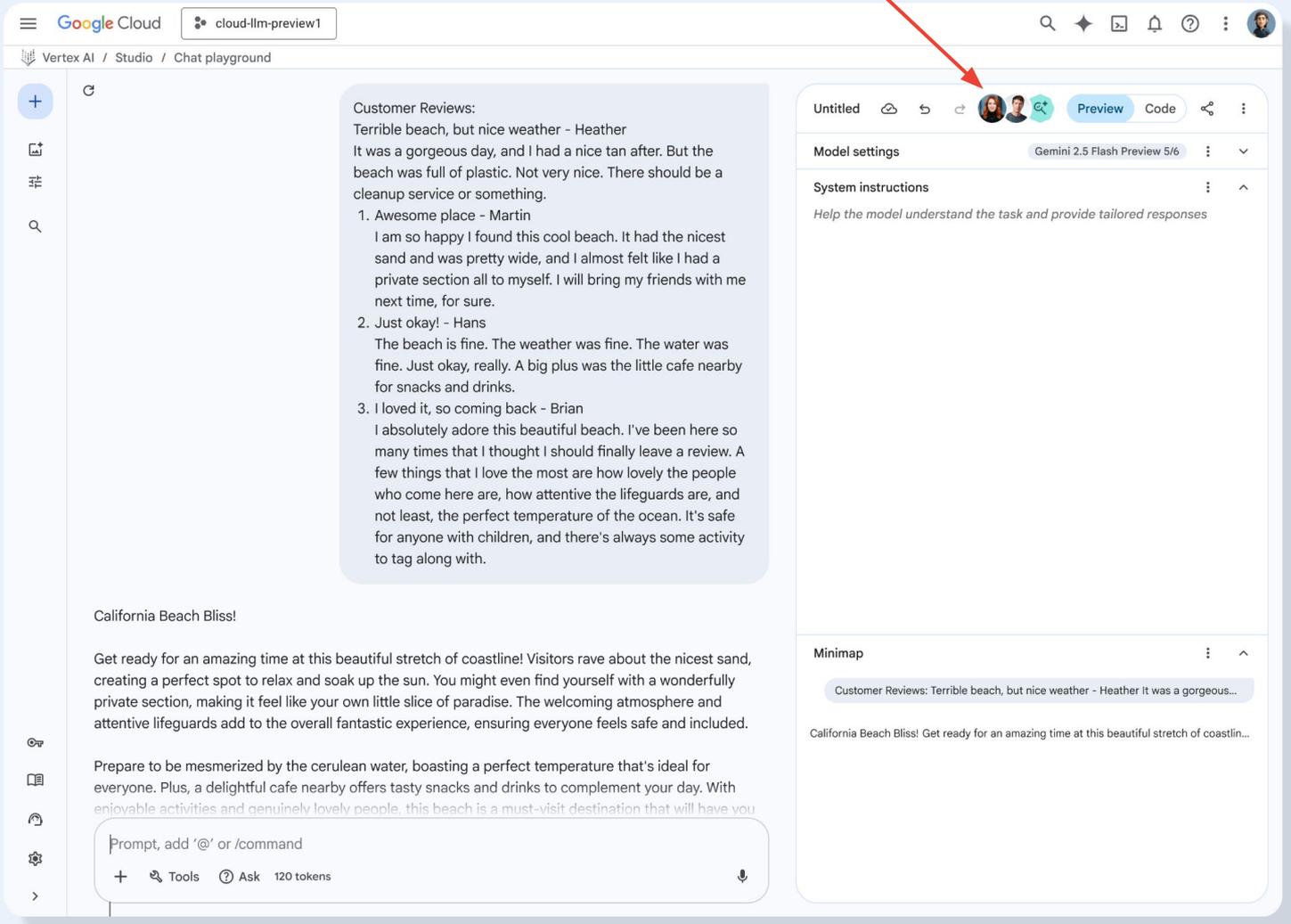
UX Patterns Collaboration

Sharing with
different access
levels.



UX Patterns Collaboration

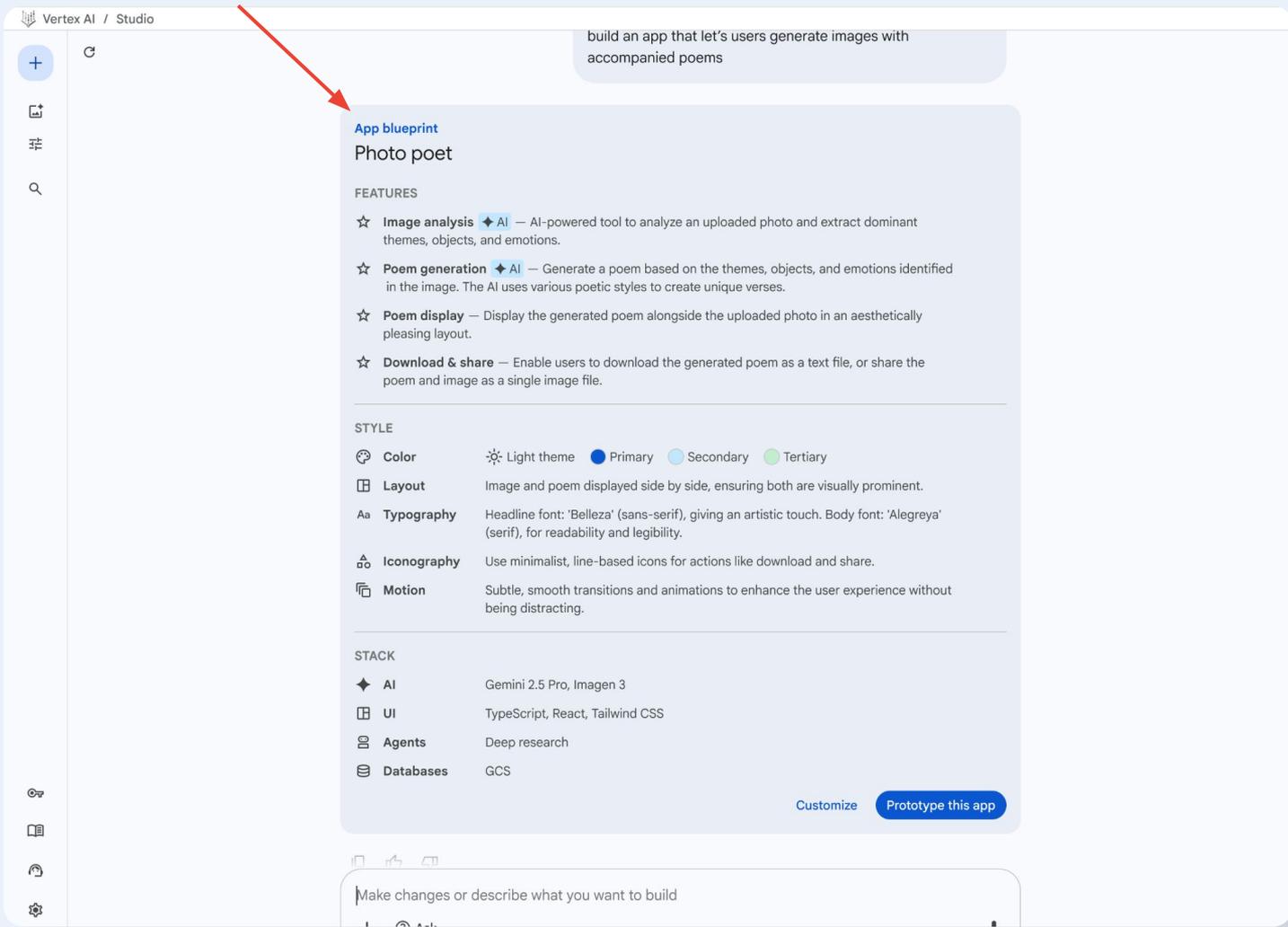
Presence indicators



UX Patterns

Agentic workflows

Plans can be generated for complex requests from the user. After the user reviews or edits these plans, then the AI will be able to much better create what the user is asking for.



AI DAU

AI Daily Active User

2025 Resolution: Become an AI Daily Active User

Achieving 3x improvements is not aspirational, it's the new baseline. The technology exists to make this achievable. The onus is now on us to build that muscle and leverage it appropriately.

Gemini

Use Gemini Advanced with Deep Research to get in-depth answers, saving you hours of time.

[go/bard](#)

AI Studio

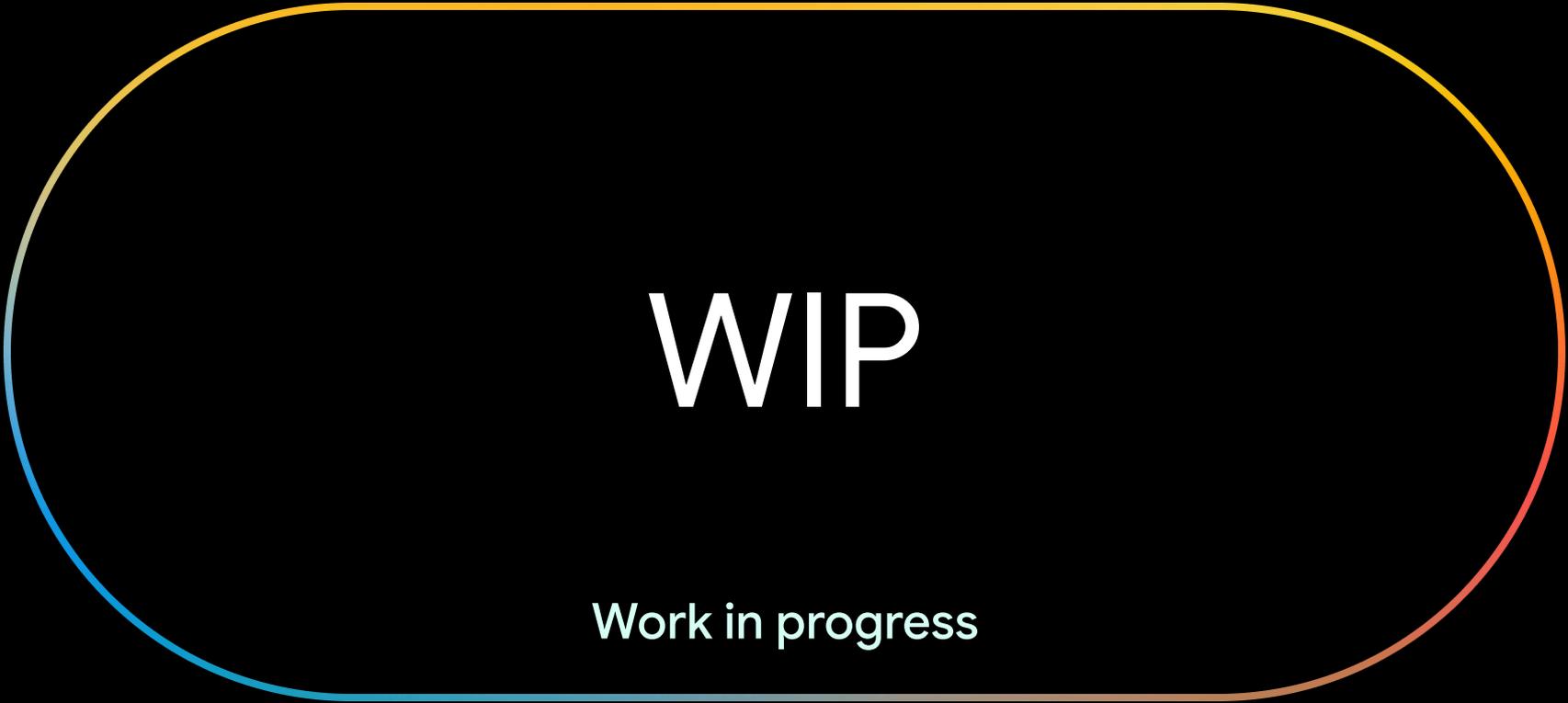
Use the Multimodal Streaming functionality for real-time conversations and share your screen to show Gemini what you're working on.

[go/aistudio](#)

GDM Labs

Use the latest Gemini models and experiments from DeepMind to leverage our competitive advantage and build early intuition.

[go/gdm-labs](#)



WIP

Work in progress

Make it easy to do the right thing.
Make it hard to do the wrong thing.

Experience northstar pillars

01

Goal-oriented

Contextualized intent

Understand user intent within their specific environment and needs.

Proactive & personalized

Offer anticipatory guidance and tailored experiences to drive progress.

Product-led experience

Design the platform to intuitively guide users towards success.

02

Collaborative

Real-time & asynchronous

Support flexible teamwork, whether in sync or independently.

Seamlessly integrated

Connect effortlessly with existing tools (Workspace, Slack, etc.).

AI co-creator & collaborator

Empower users with intelligent assistance and partnership.

03

AI-Native

Agentic intelligence

Leverage agents to streamline complex workflows.

Multimodal capabilities

Enable interaction and generation through diverse modalities (text, visuals, voice, video, etc.).

Bespoke adaptability

Tailor the experience and solutions to unique user and project requirements.

Northstar pillars

01

Goal-oriented

Contextualized intent
Proactive & personalized
Product-led experience

02

Collaborative

Real-time & asynchronous
Seamlessly integrated
AI co-creator & collaborator

Foundation

AI-Native

Agentic intelligence
Multimodal capabilities
Bespoke adaptability