Cloud AI UX

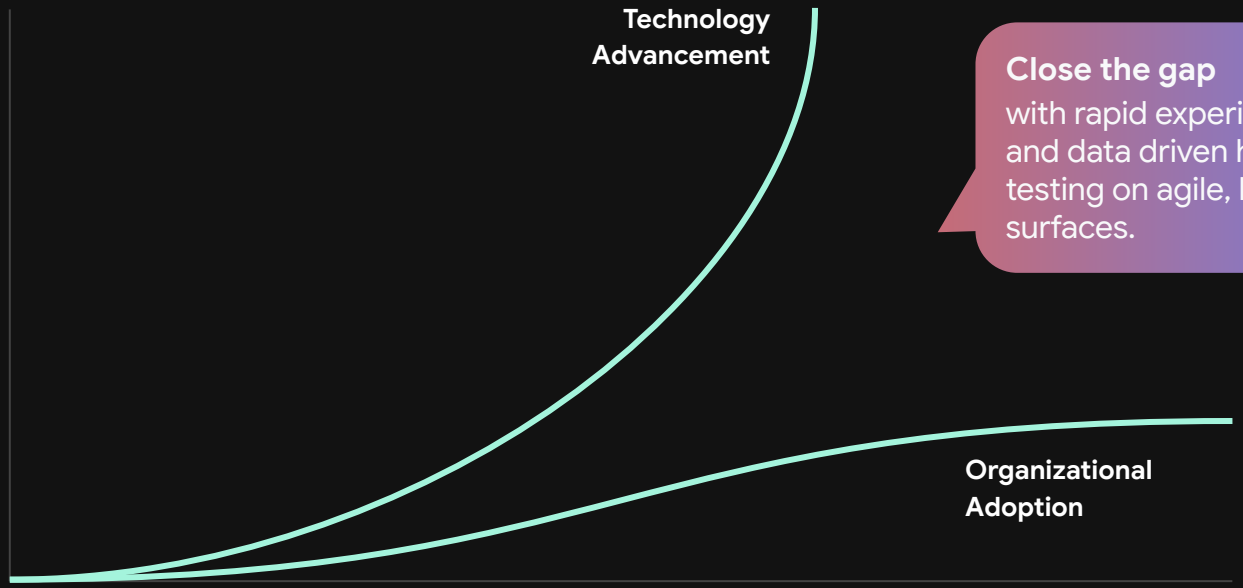
UX foundation, vision, strategy, patterns, and methodology for the neoAI/AI-Native era

Living document
2025

Cloud AI

Proprietary + Confidential

Our target is to close the gap



Close the gap
with rapid experimentation and data driven hypothesis testing on agile, lightweight surfaces.

Technology changes exponentially; organizations change logarithmically.

GenAI will accelerate 10x

Ensure users are running into the right direction.

Cloud AI

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

Cloud AI

Goal-oriented

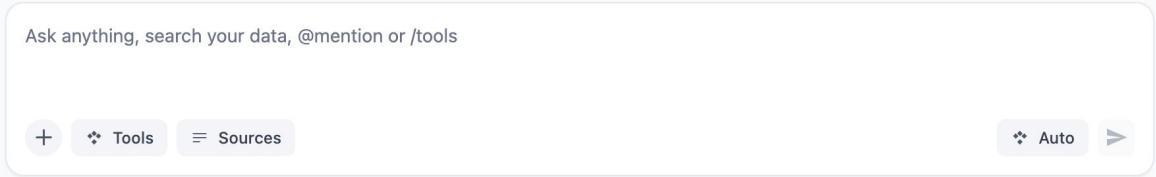
User / system invoked

Goal-oriented

Contextualized intent

Understand user intent within their specific environment and needs.

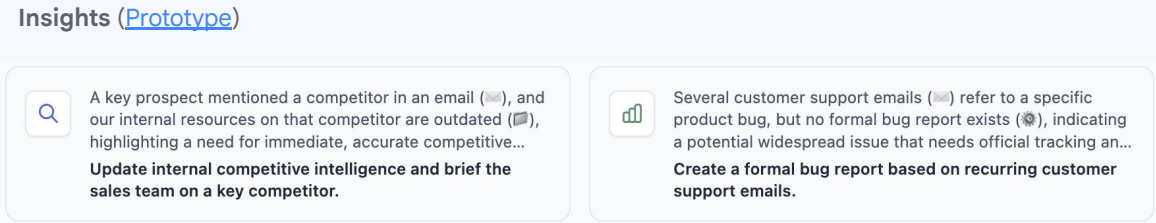
User invoked →



Proactive & personalized

Offer anticipatory guidance and tailored experiences to drive progress.

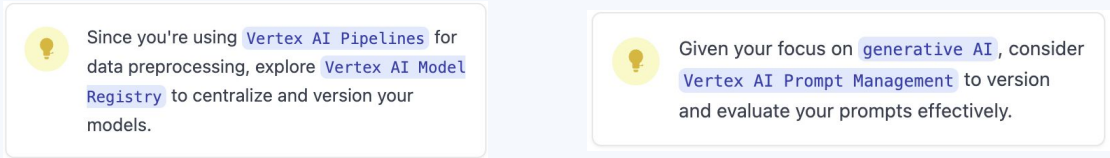
System invoked →



Product-led experience

Design the platform to intuitively guide users towards success.

System invoked →

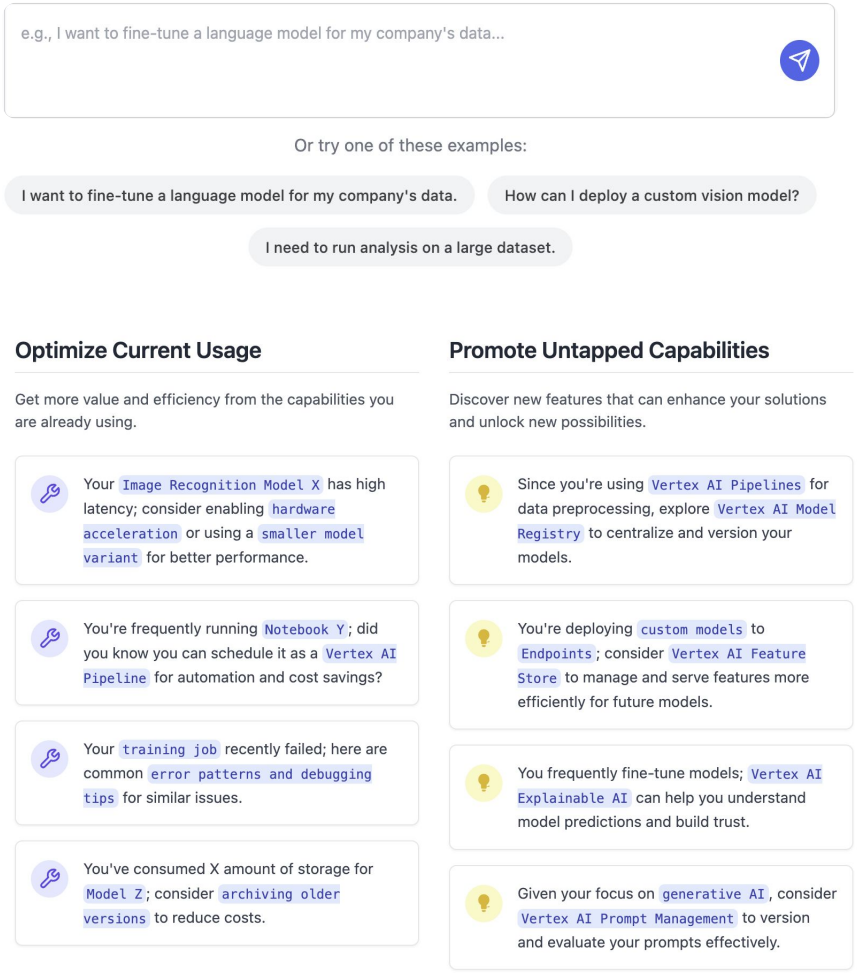


Goal-oriented

Example: Vertex

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

Prototype

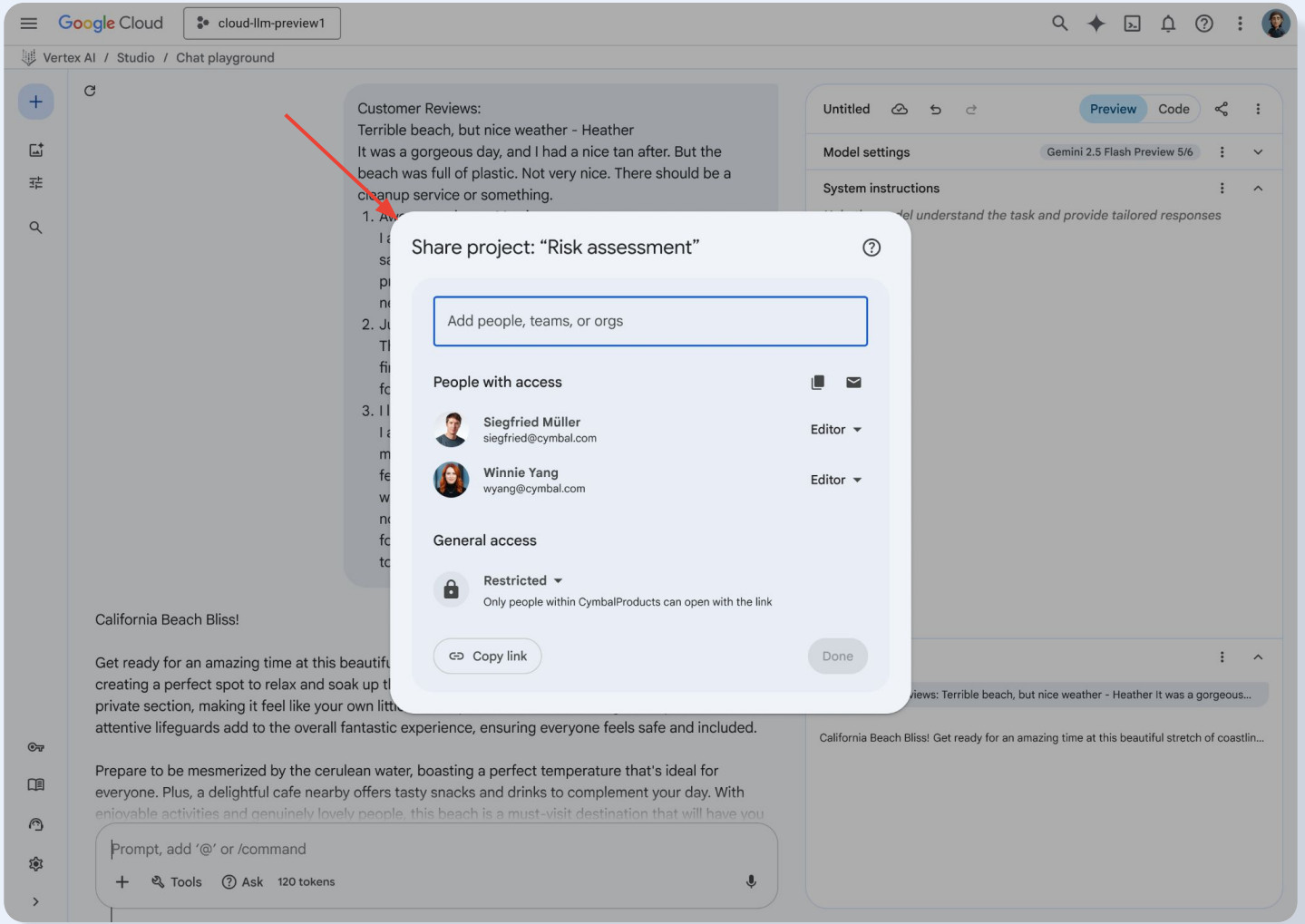


Collaborative

Humans & agents collaboration

Sharing
Collaboration

Invite team members.

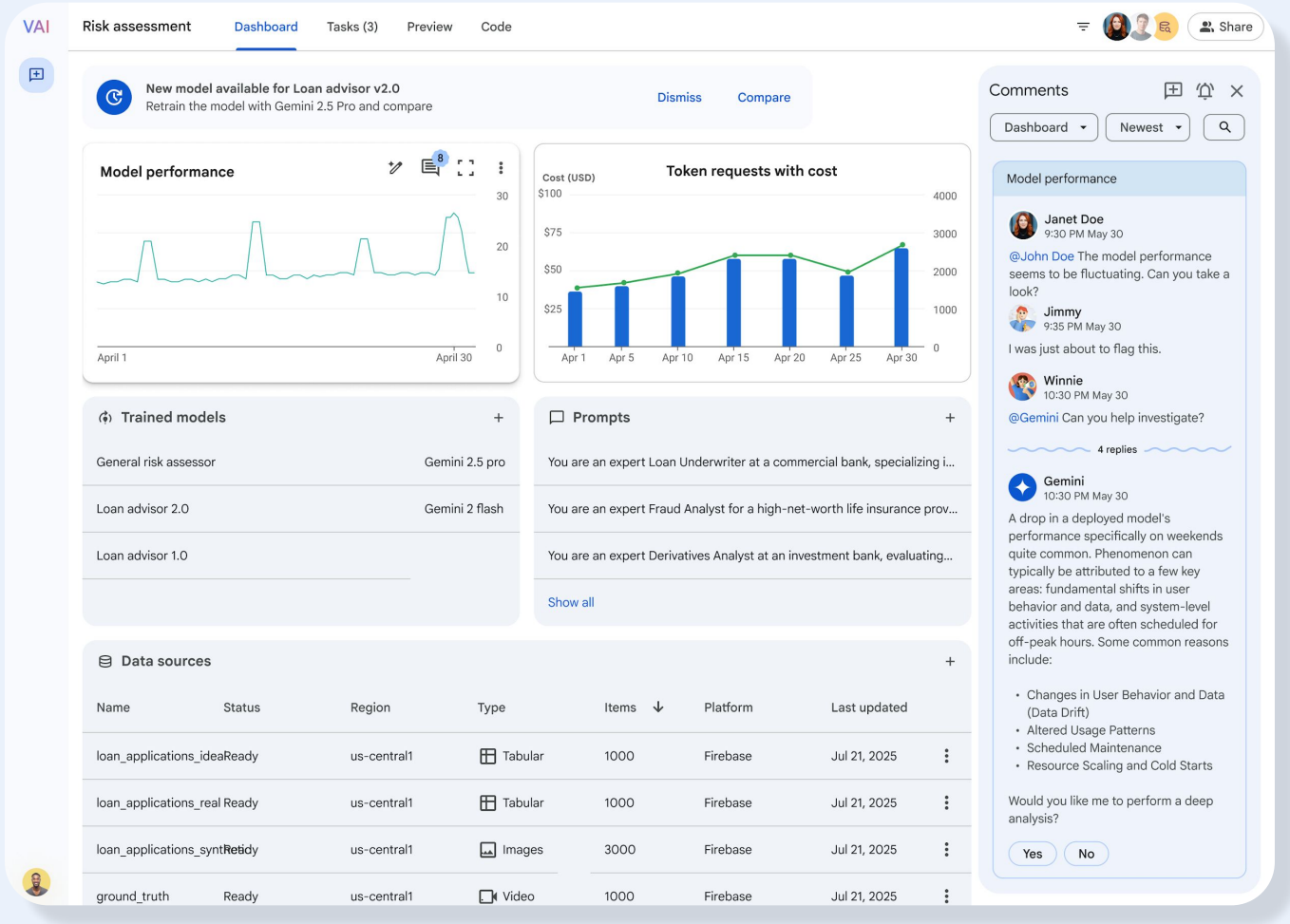


Cloud AI

Asynchronous
Collaboration

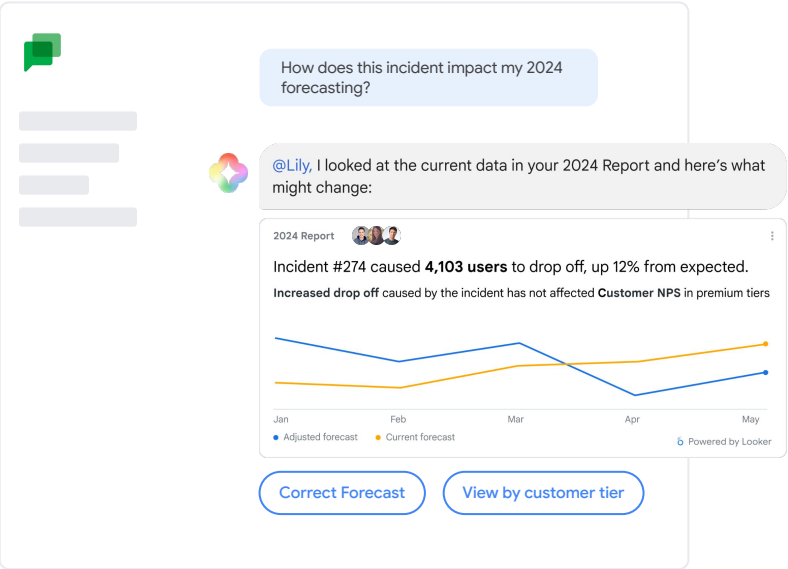
Comment panel opens up to show threads where users are collaborating.

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



Collaborative - Seamlessly integrated where the work happens

On demand



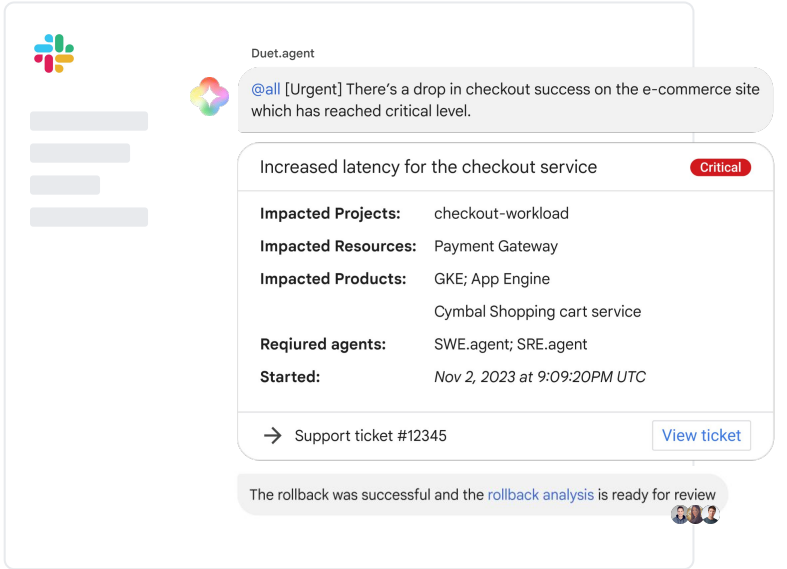
Scenario

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



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

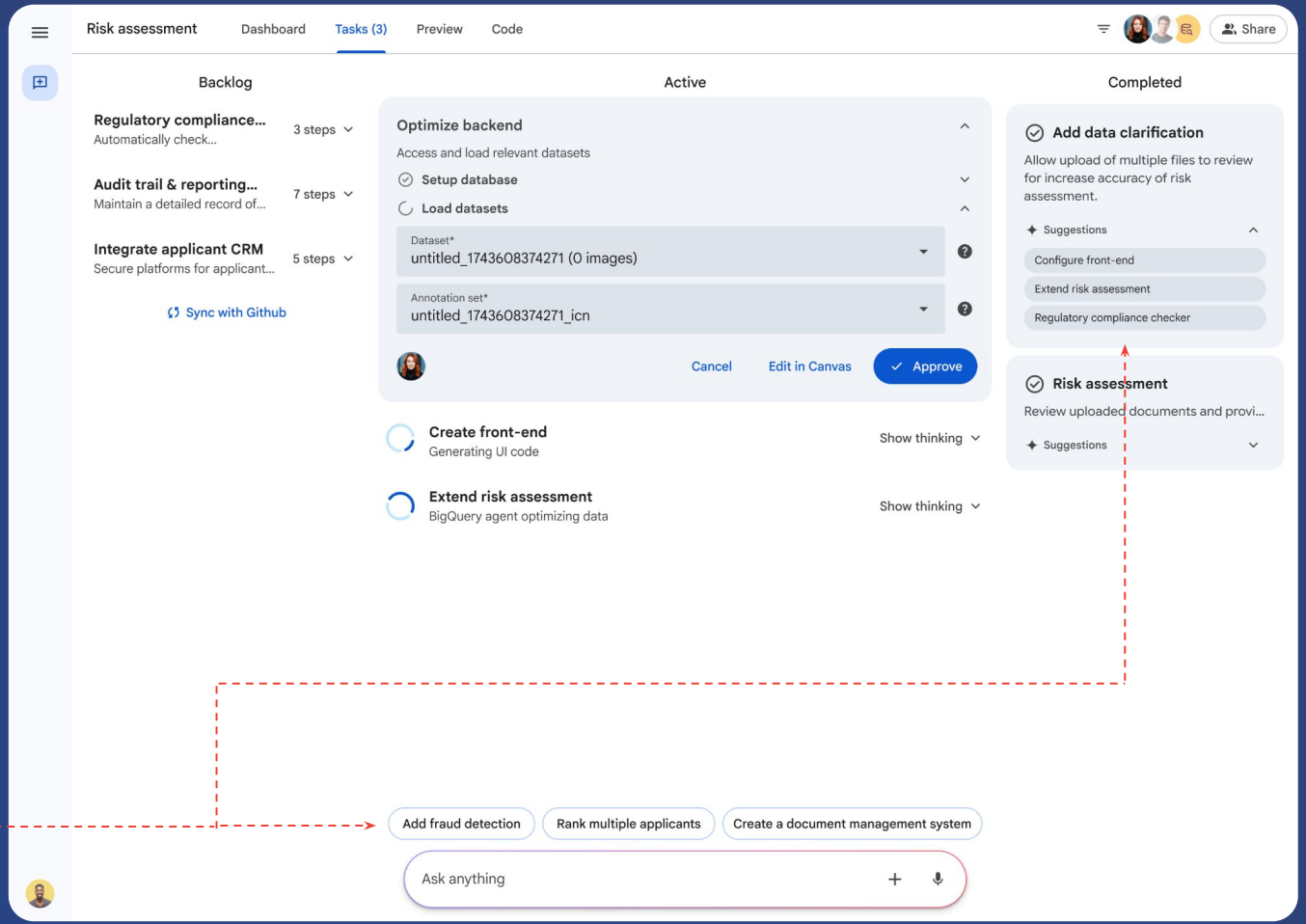
🕐 Fast forward

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

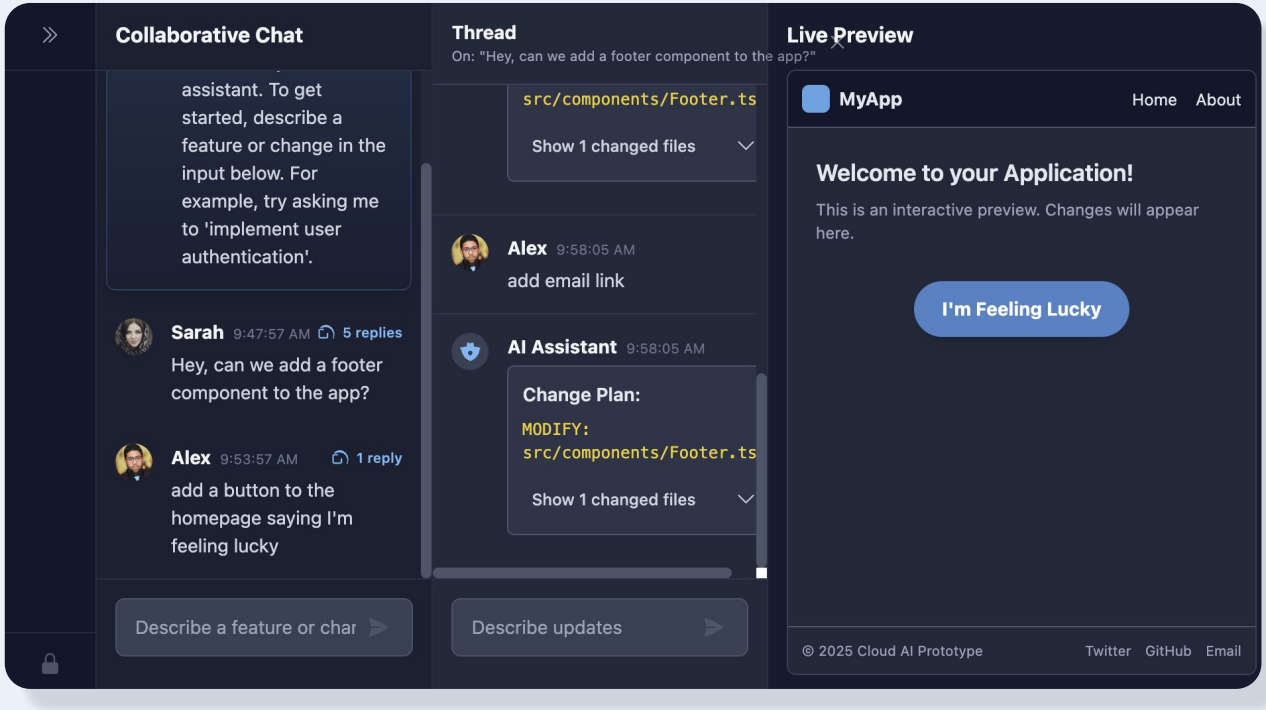


Exploraration

Realtime
Collaboration

Exploraration: 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

Cloud AI

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?

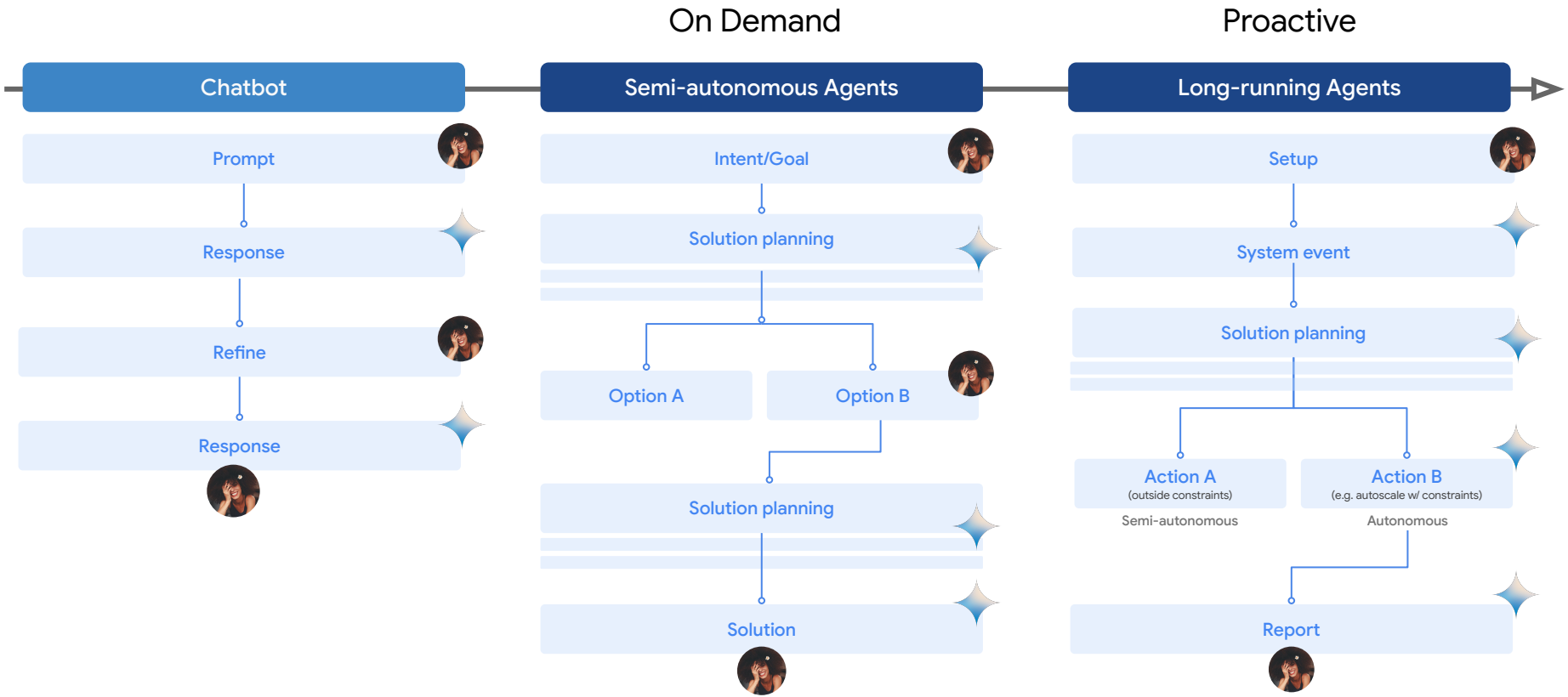
Cloud AI

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



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

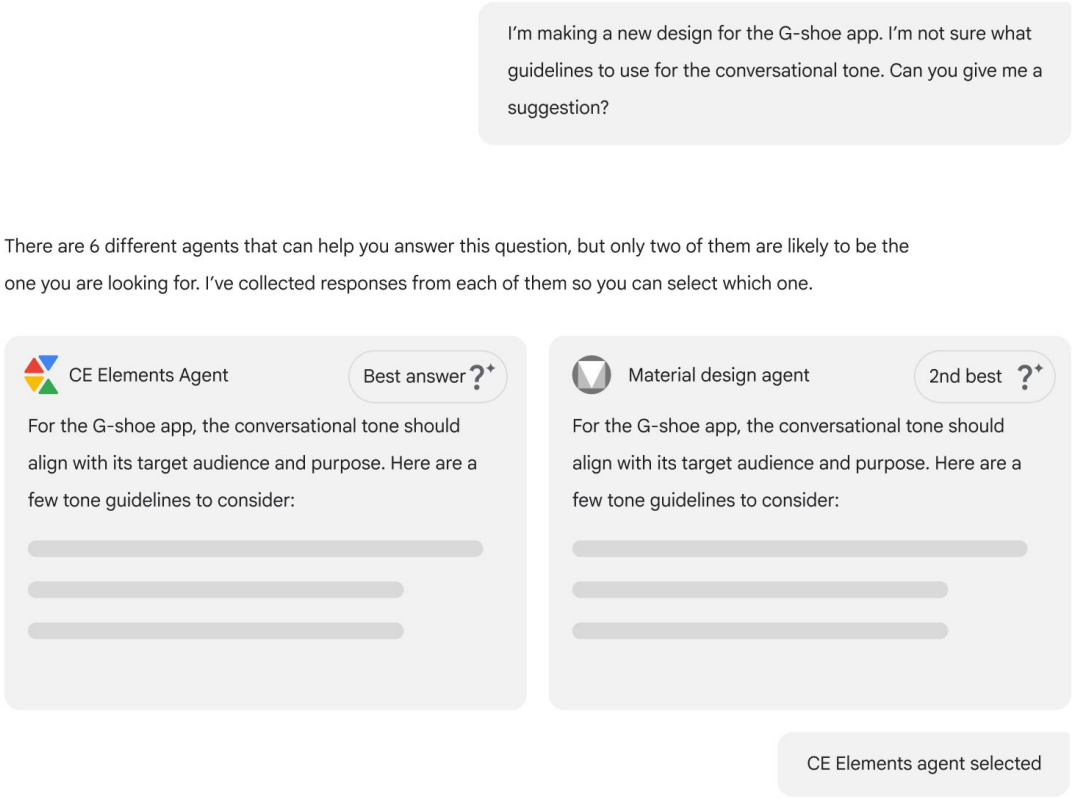
OneMCP is a common infrastructure framework developed by Google Cloud to streamline the creation, discovery, and secure management of Model Context Protocol (MCP) tools and AI agents across Google, Google Cloud, and partner ecosystems.

Links: go/gcp-mcp-eng-pitch, go/onemcp-concept go/onemcp-prd, go/onemcp-ug

Cloud AI

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.



Cloud AI

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.

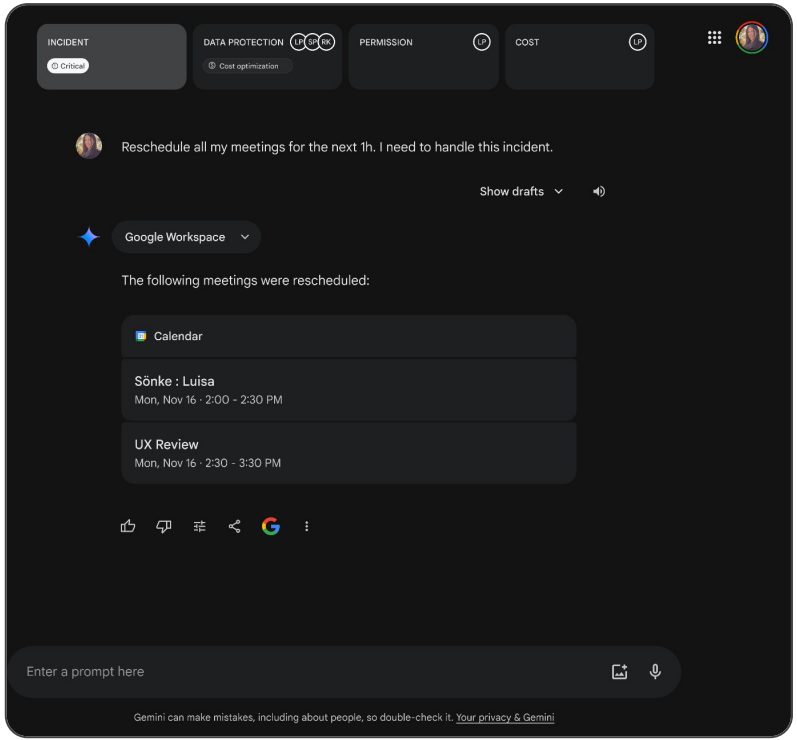
Agile & fast
Highly adaptable to evolving user needs and product directions, enabling faster velocity and time to market.

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

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

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

Benefits
Reduced cognitive load, faster time to value.



Design Exploration go/gcp-gemini-home

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

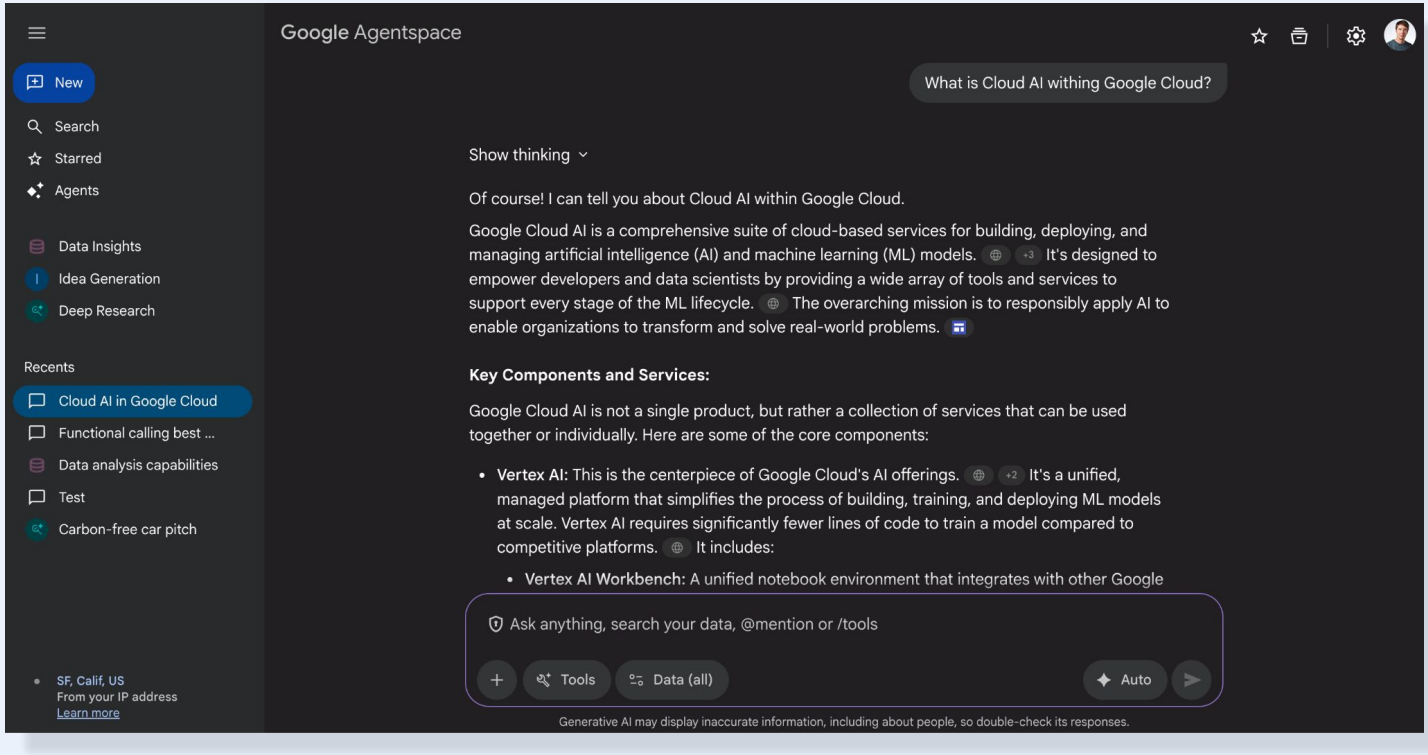
AI/Agentic systems need more than just chat for complex, asynchronous tasks. We can organize these systems into three interacting levels:

	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

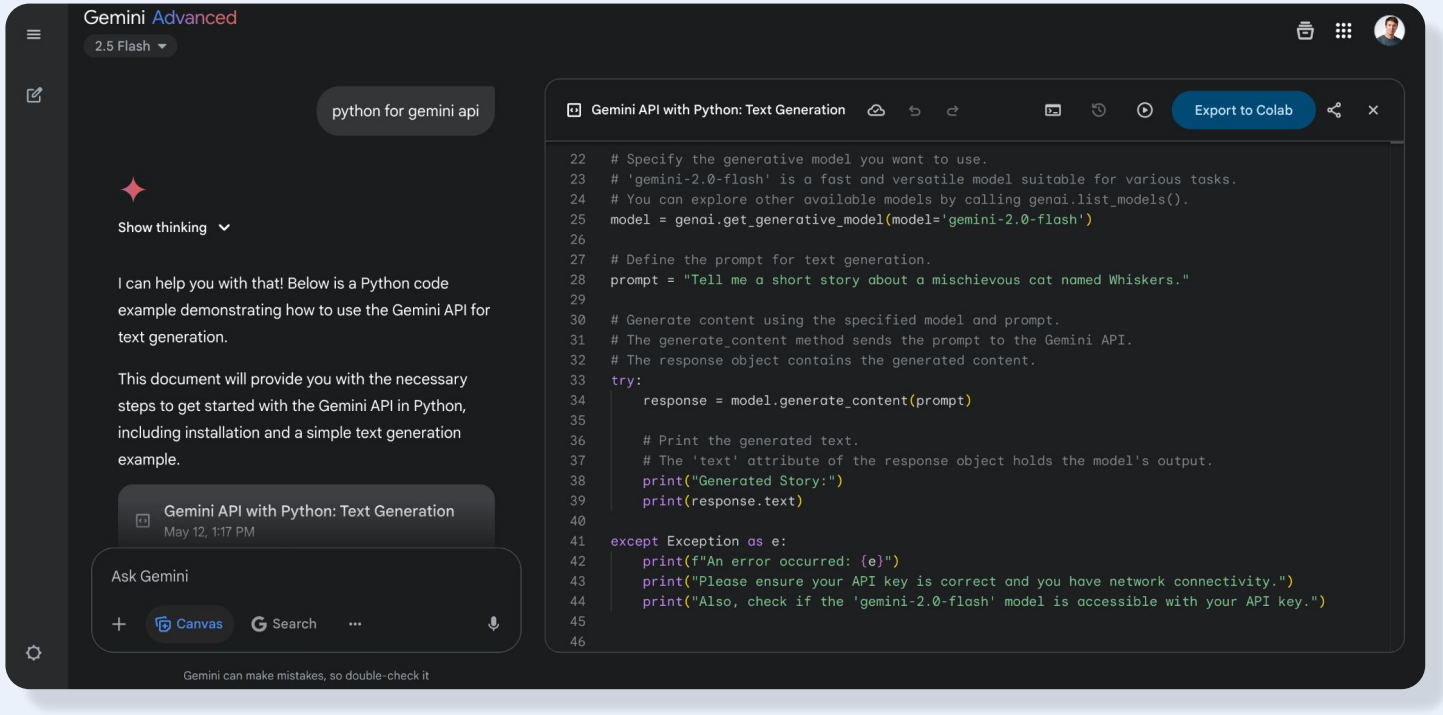
A focused, dialog-based interaction model to guide task execution one step at a time with single agents.



Tactical

Canvas

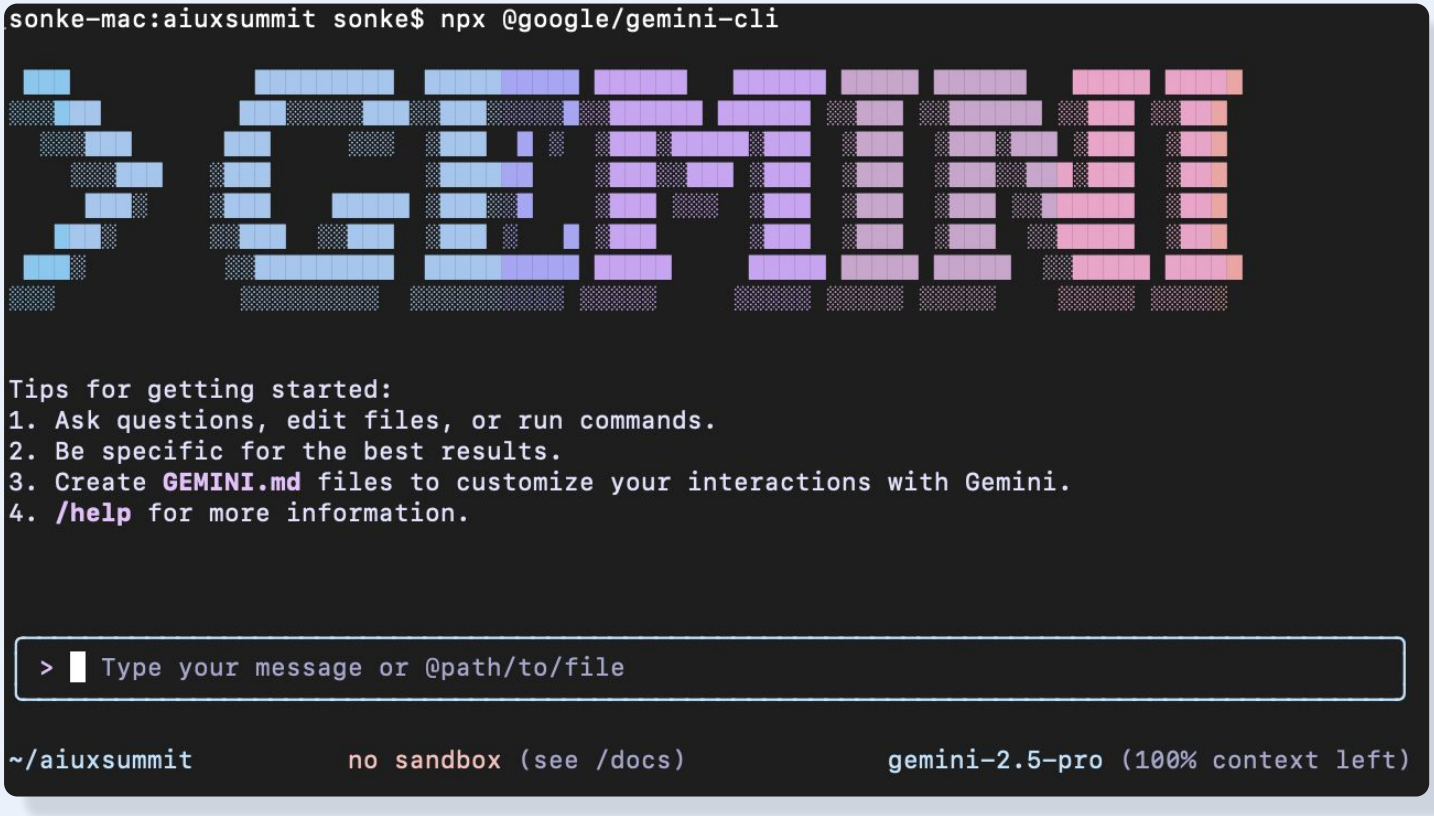
A hybrid experience that combines assistive chat with direct manipulation of an artifact (e.g. code), enabling fluid, in-context collaboration.



Tactical

CLI

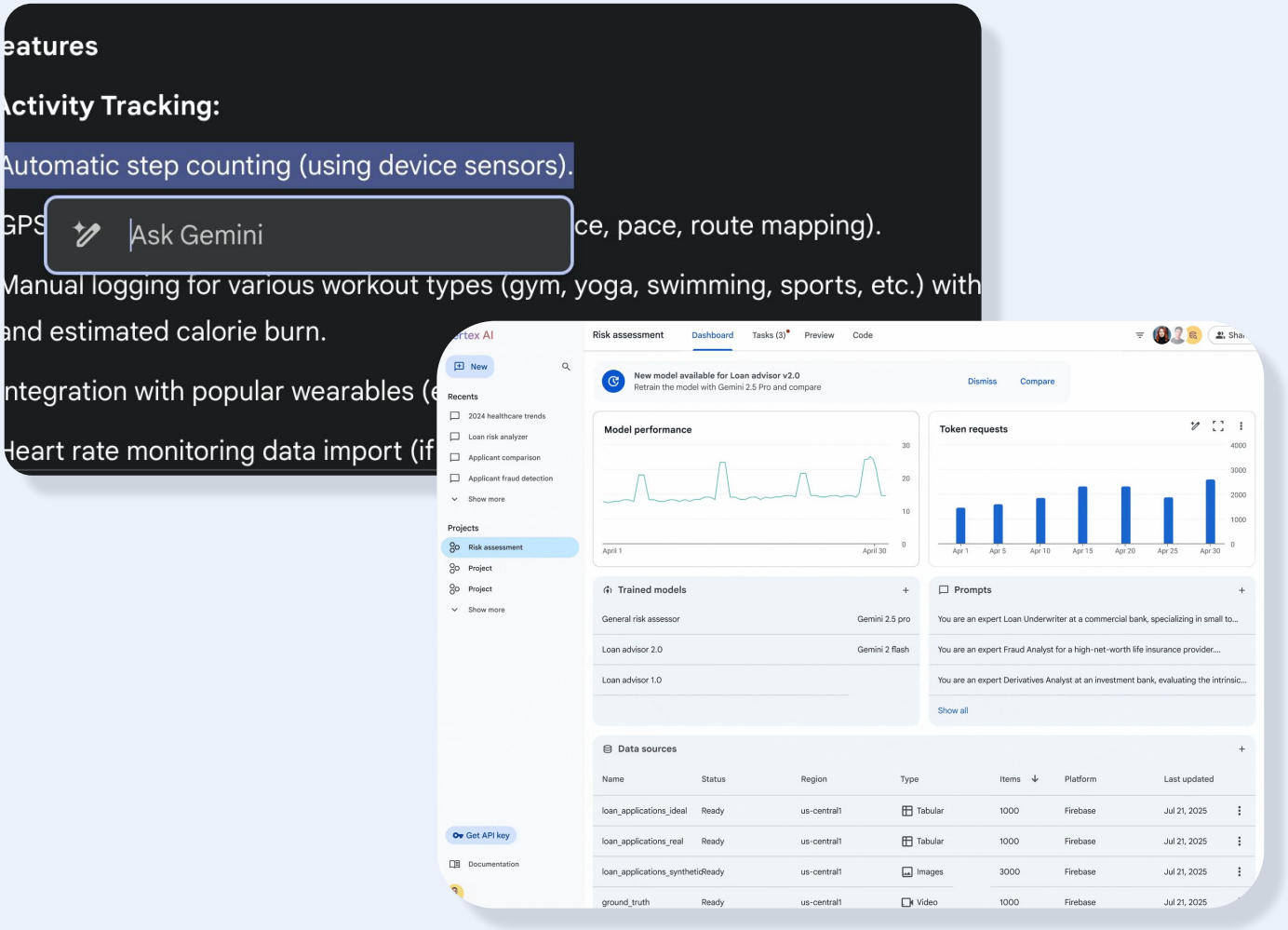
Provides direct, **local access** for power users and developers. It allows for precise control, scripting, and direct interaction with local files, serving as an efficient and unmediated interface for advanced AI tasks.



Operational

Embedded

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

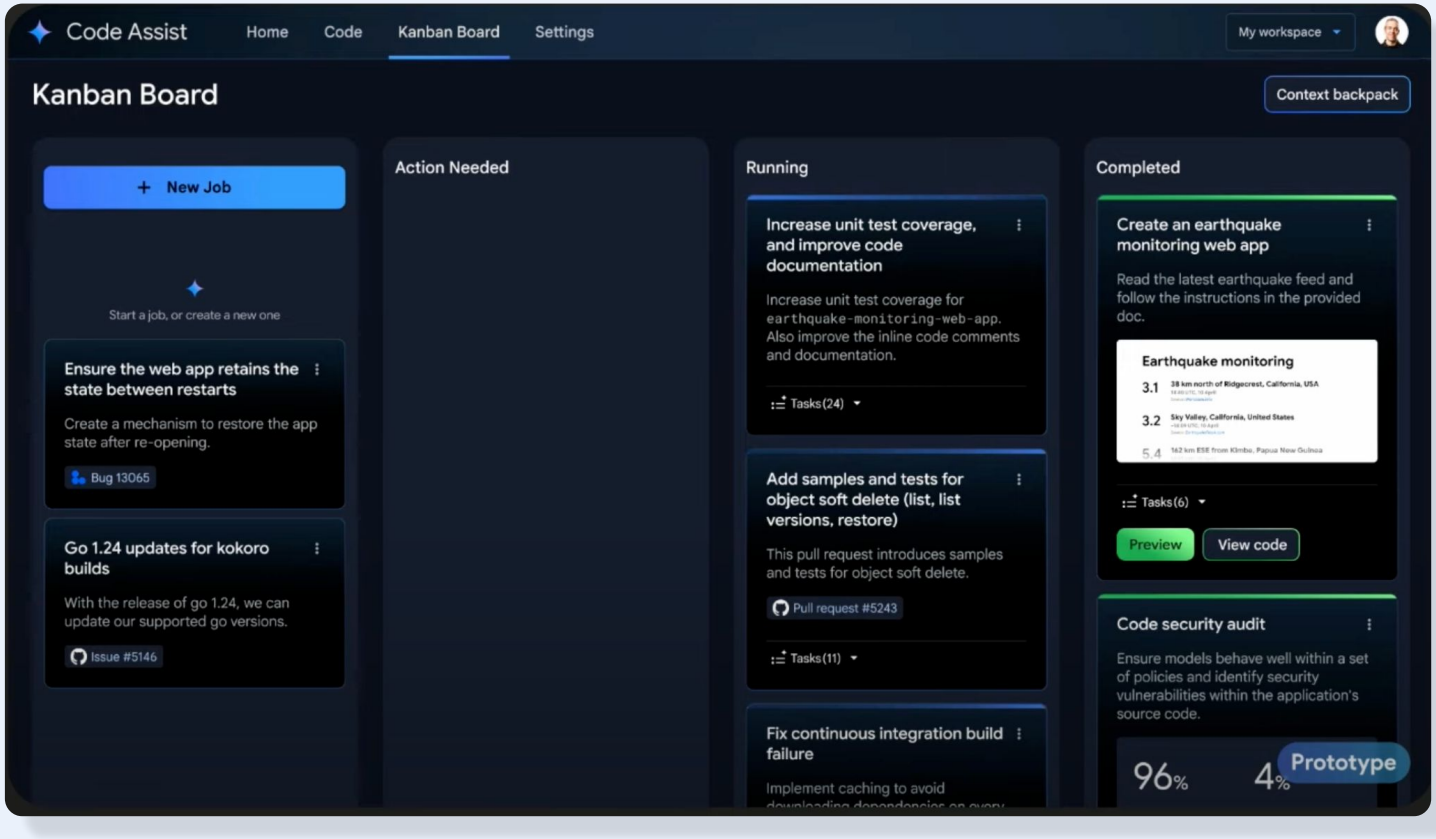


Strategic

Kanban

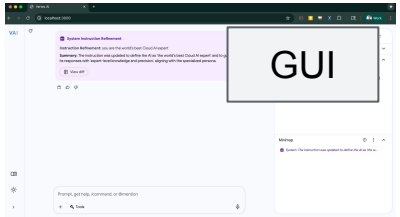
A high-level environment to orchestrate and execute multiple tasks or goals (e.g. project) augmented by agents.
Example: Gemini Code Assist w/ SDLC (Software Development Lifecycle) agents.

Google Cloud Next 25 Developer Keynote



Surface orchestration

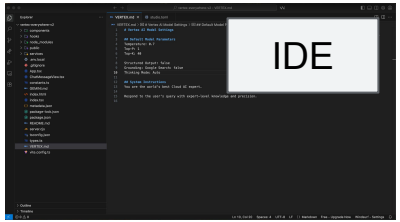
WIP



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

SPEC.md

Agent Loop
Executes commands aligned with specs



Validation Loop
review generated code and tweak if necessary

Cloud AI

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.

Details and examples go/agents-next-gen

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.

Source: Help Wanted: AI Designers

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

Prototyping

Title: body...

Title:

- Concepting
- Design
- Research
- Iteration

Realistic prototypes

Thanks to vibe coding toolkits like go/material-vibe, designers, PMs, researchers, and Eng can all mockup realistic prototypes within minutes

Title

body...body...body...body...body...body...body...body...body...body...ody...body...body...body...body...body...body...body...body...body...body...

Title

body...body...body...body...body...body...body...body...body...body...body...

Title

body...body...body...body...body...body...body...body...body...body...body...

UXR

UX Research

Studies

Gemini-First

TODO

Thank you

Appendix

Mindset Shift

Proprietary + Confidential

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.

Proprietary + Confidential

Choosing the familiar over the bold: The 10X dilemma

Mindset shifts

50% Improvement – Cater to a 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.

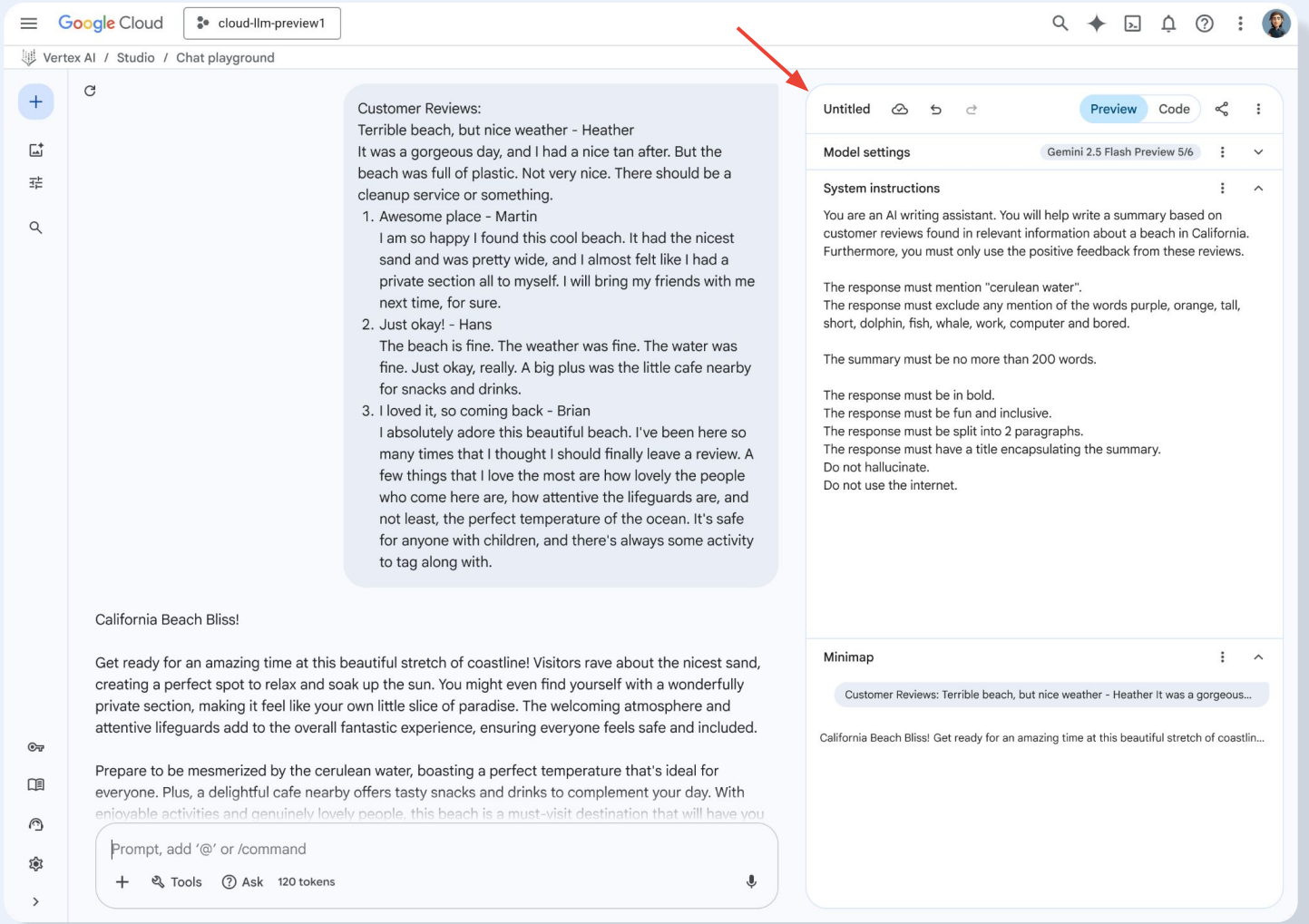
Cloud AI

UX Patterns

UX Patterns

Canvas

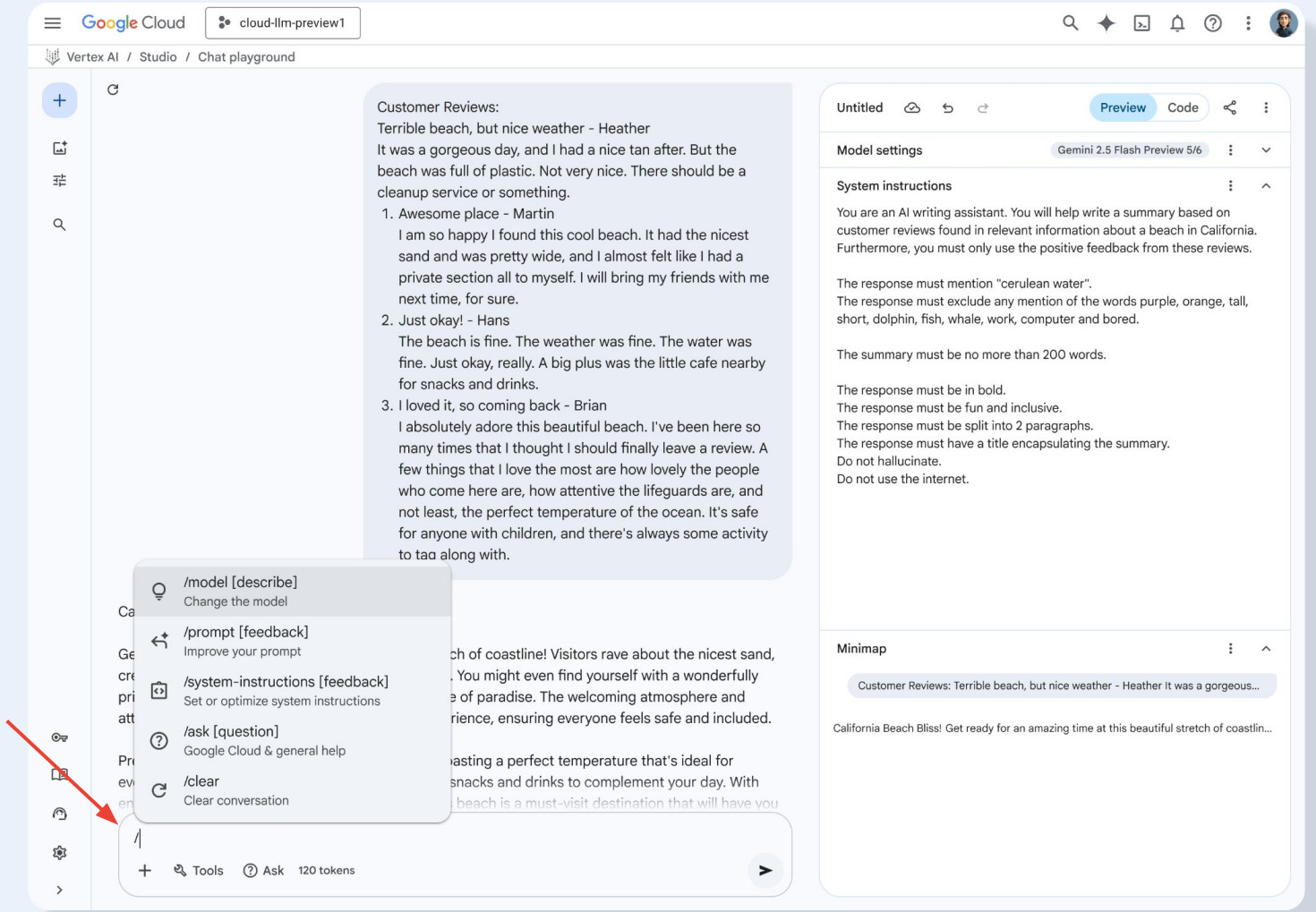
The canvas component is adaptable for anything that needs visual display or interactions (e.g. code editor, settings, form fields, app preview, doc preview, etc)



UX Patterns

/commands

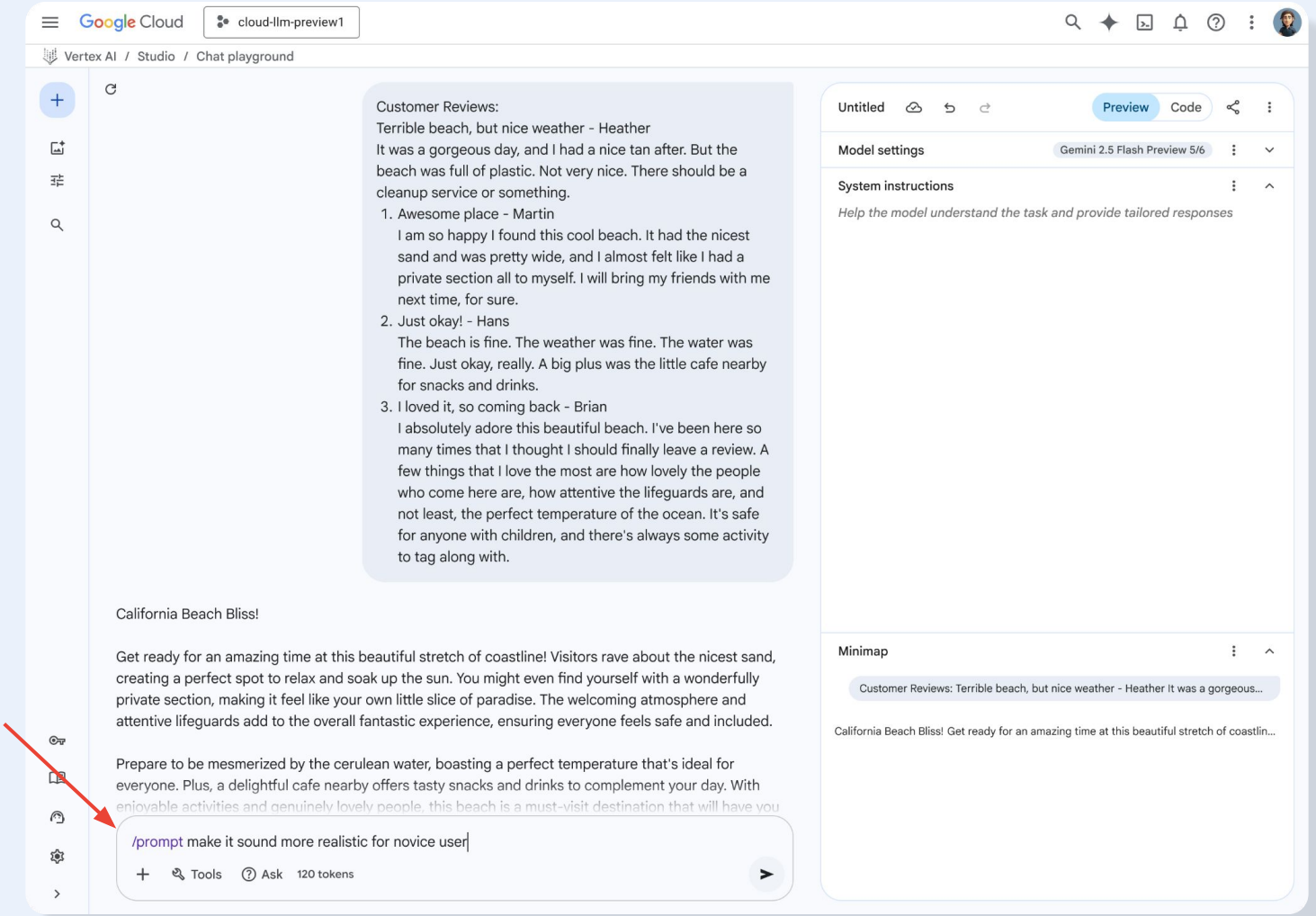
Slash commands within the chat input field allow the user to invoke advanced actions or shortcuts (e.g. change the model, refine a prompt, clear the conversation, etc)



UX Patterns

/commands

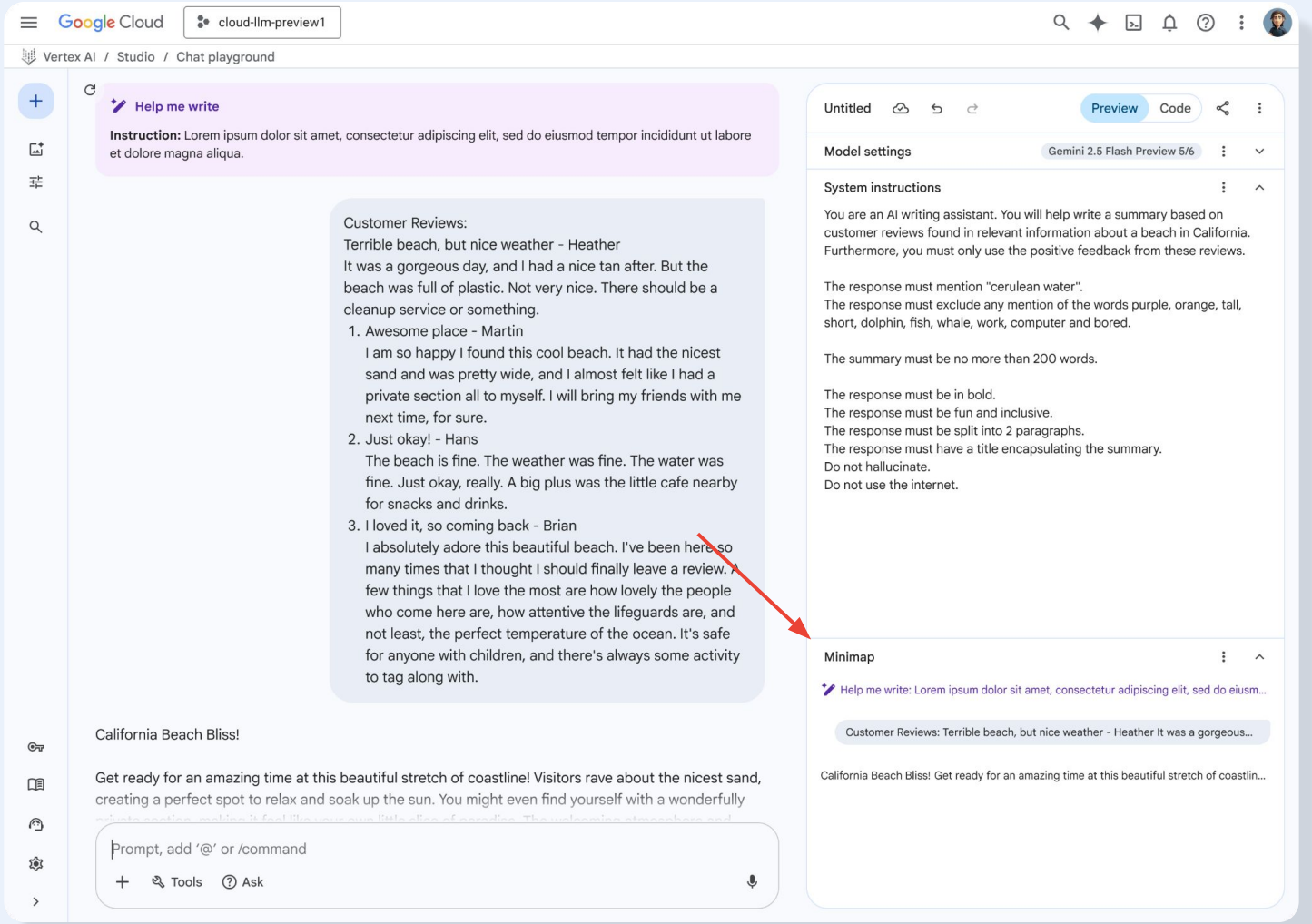
Slash commands within the chat input field allow the user to invoke advanced actions or shortcuts (e.g. change the model, refine a prompt, clear the conversation, etc)



UX Patterns

Minimap

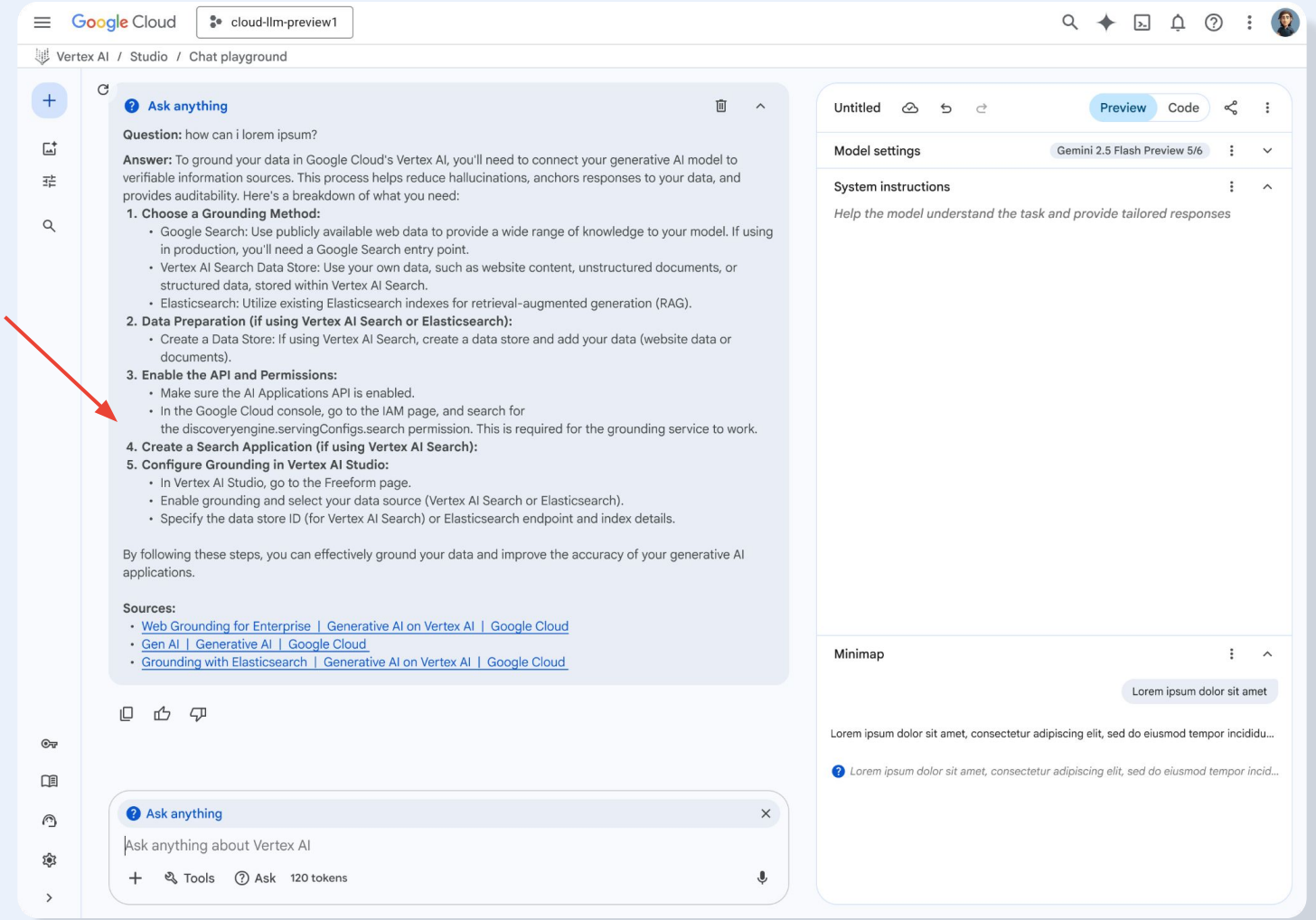
The conversation minimap makes it easy to traverse a long conversation by condensing it small into single line entries that are clickable to autoscroll to that point.



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.

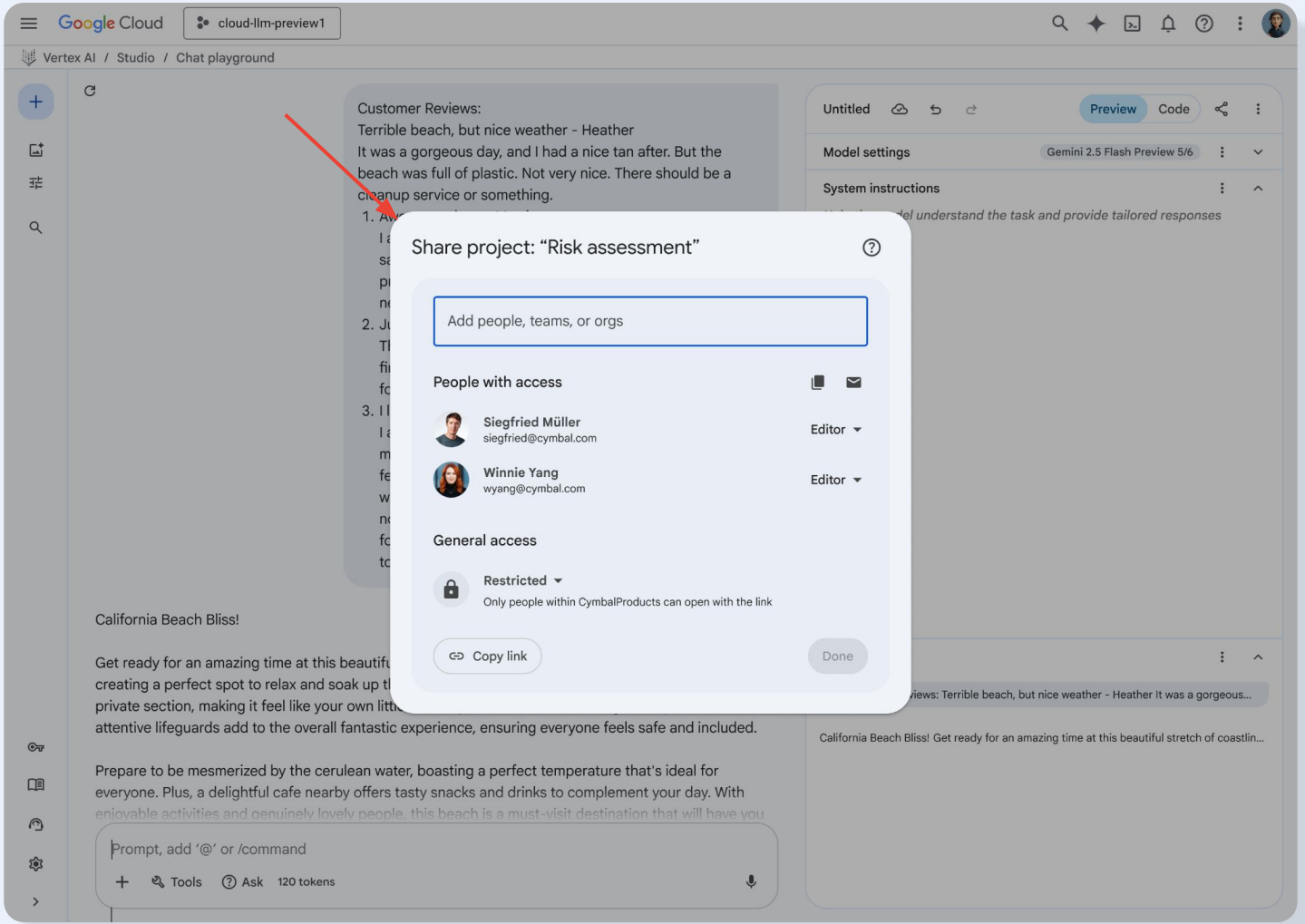


Cloud AI

UX Patterns

Collaboration

Sharing with different access levels.

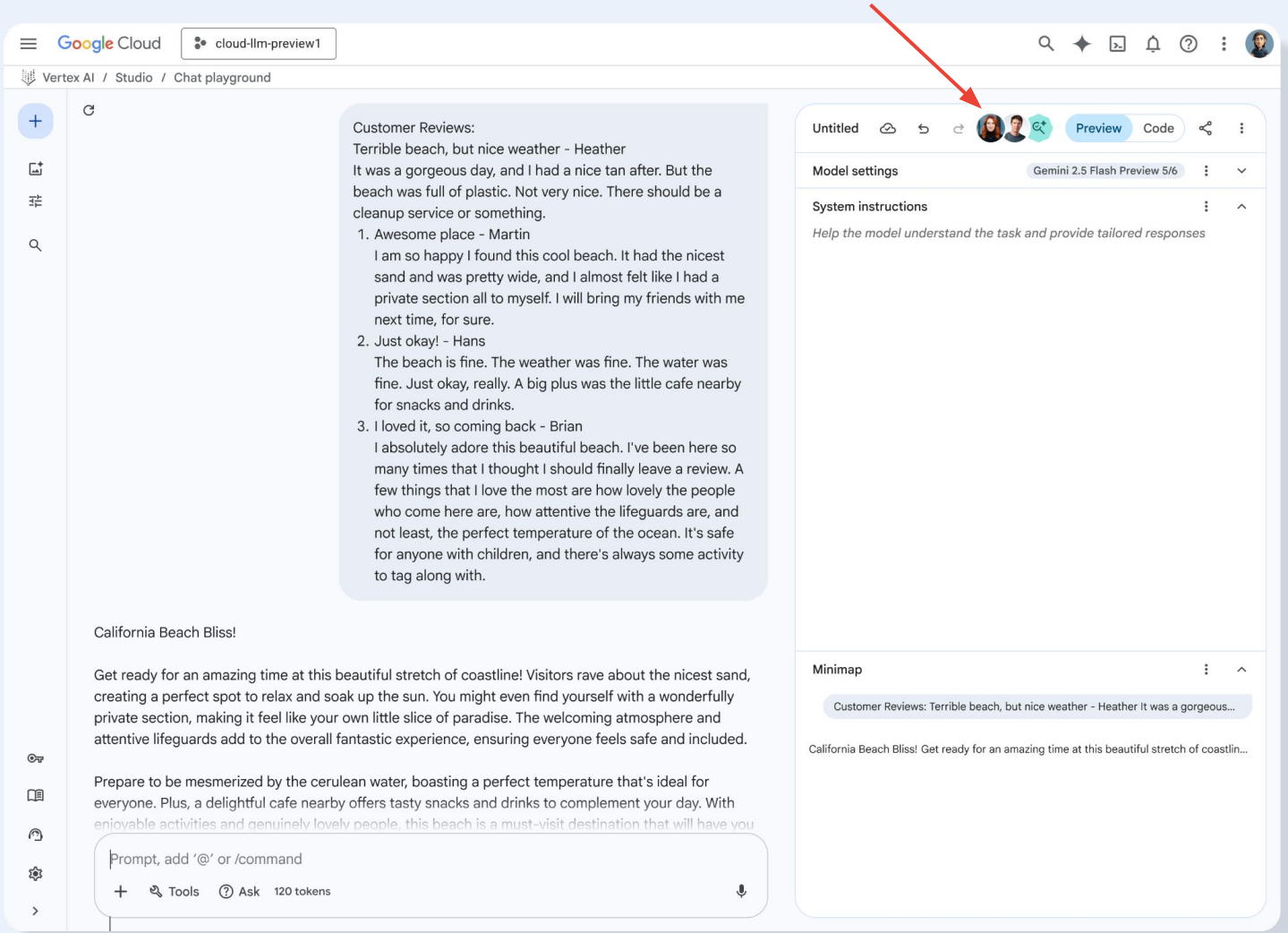


Cloud AI

UX Patterns

Collaboration

Presence indicators

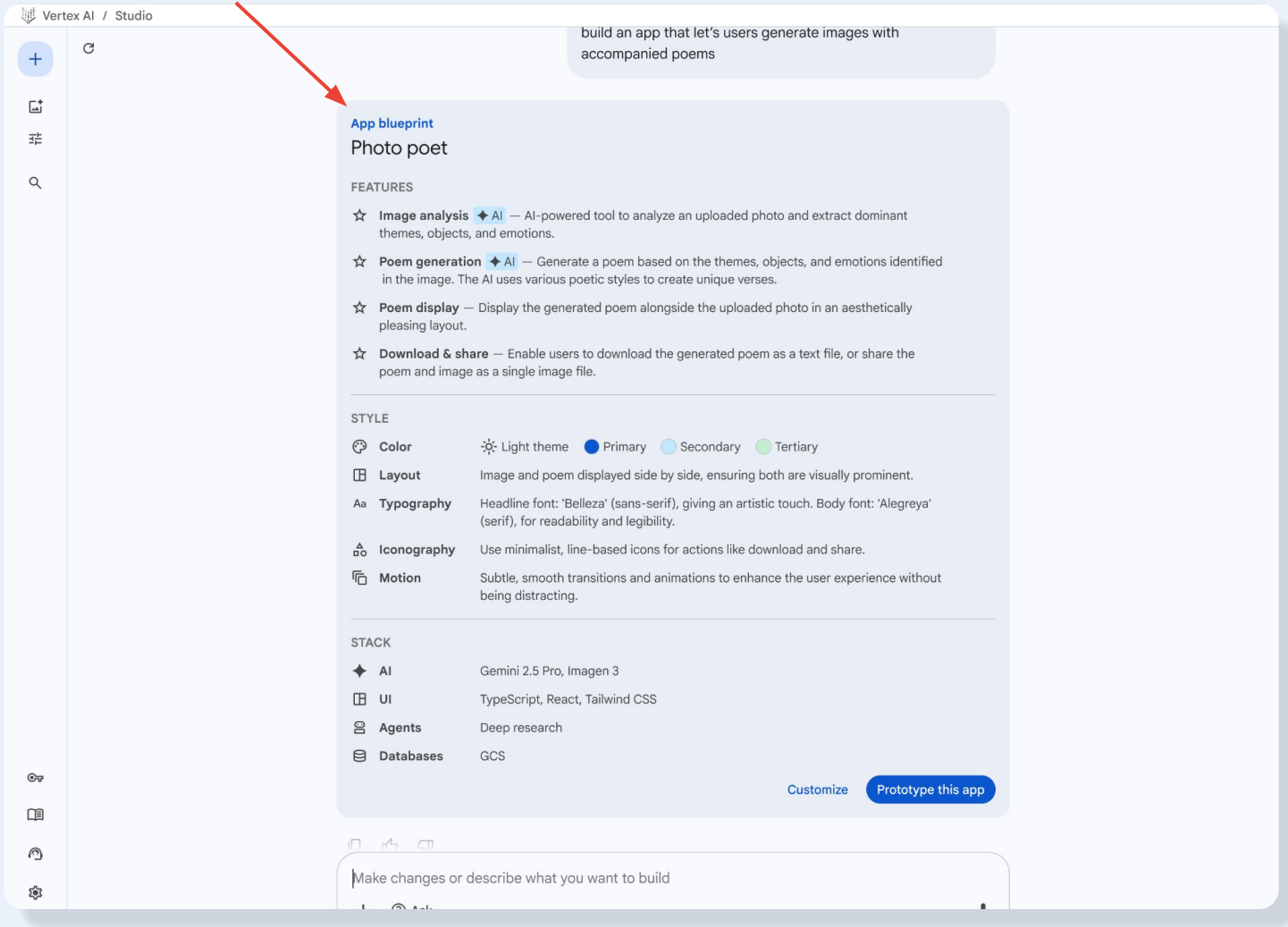


Cloud AI

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.

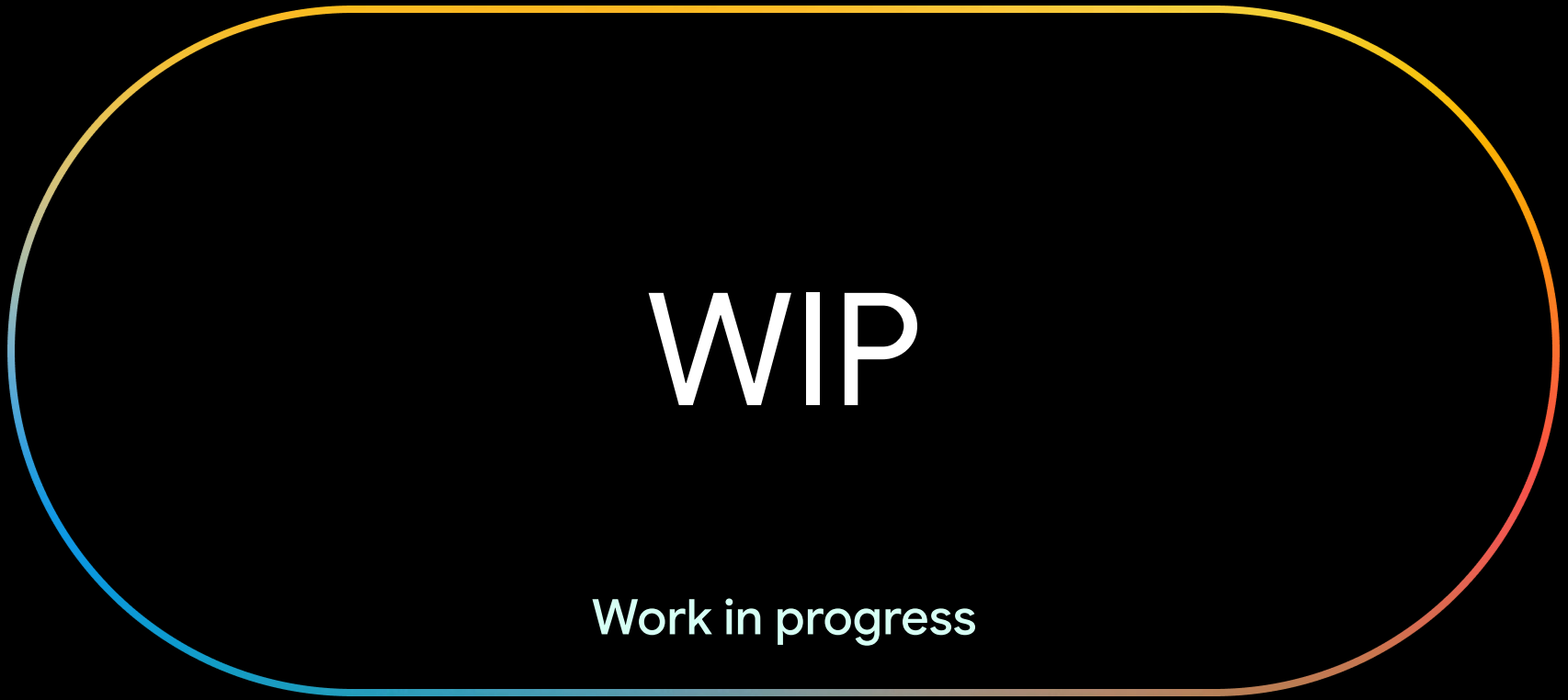
go/aistudio

GDM Labs

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

go/gdm-labs

Cloud AI


WIP
Work in progress

Make it easy to do the right thing.

Make it hard to do the wrong thing.

Cloud AI

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.

Cloud AI

Northstar pillars

01

Goal-oriented

Contextualized intent

Proactive & personalized

Product-led experience

02

Collaborative

Realźtime & asynchronous

Seamlessly integrated

AI co-creator & collaborator

Foundation

AI-Native

Agentic intelligence

Multimodal capabilities

Bespoke adaptability

Cloud AI