

# Ask Breezy

## What we do

We're building the first and last employee that any solo service pros will need to run their business, starting with a business phone number + AI frontend that can handle inbound/outbound calls + text messages to customers across their entire lifecycle. We have hundreds of customers ranging from verticals like home services, instructors, consultants and professional services. We believe in a world where millions of people can offer their services and grow their business without needing to grow headcount for operational work. We recently raised millions of dollars from top vcs, as well as some key openai employees (chief product officer, chief technical advisor to ceo, head of bd) and are looking to grow.

## Why "Ask Breezy"

Better search and insight => More value from conversations on breezy => customer takes more calls with Breezy

### **Ideally, we could answer questions like:**

"I talked to someone last week about fence painting. Who was it again?"

"What are the action items from my call with john last night?"

"When was the last time alice called me?"

"Who called me the most this month?"

"How many of my calls last week were from clients in Santa Clara?"

"What do you think I should do to grow my customer base?"

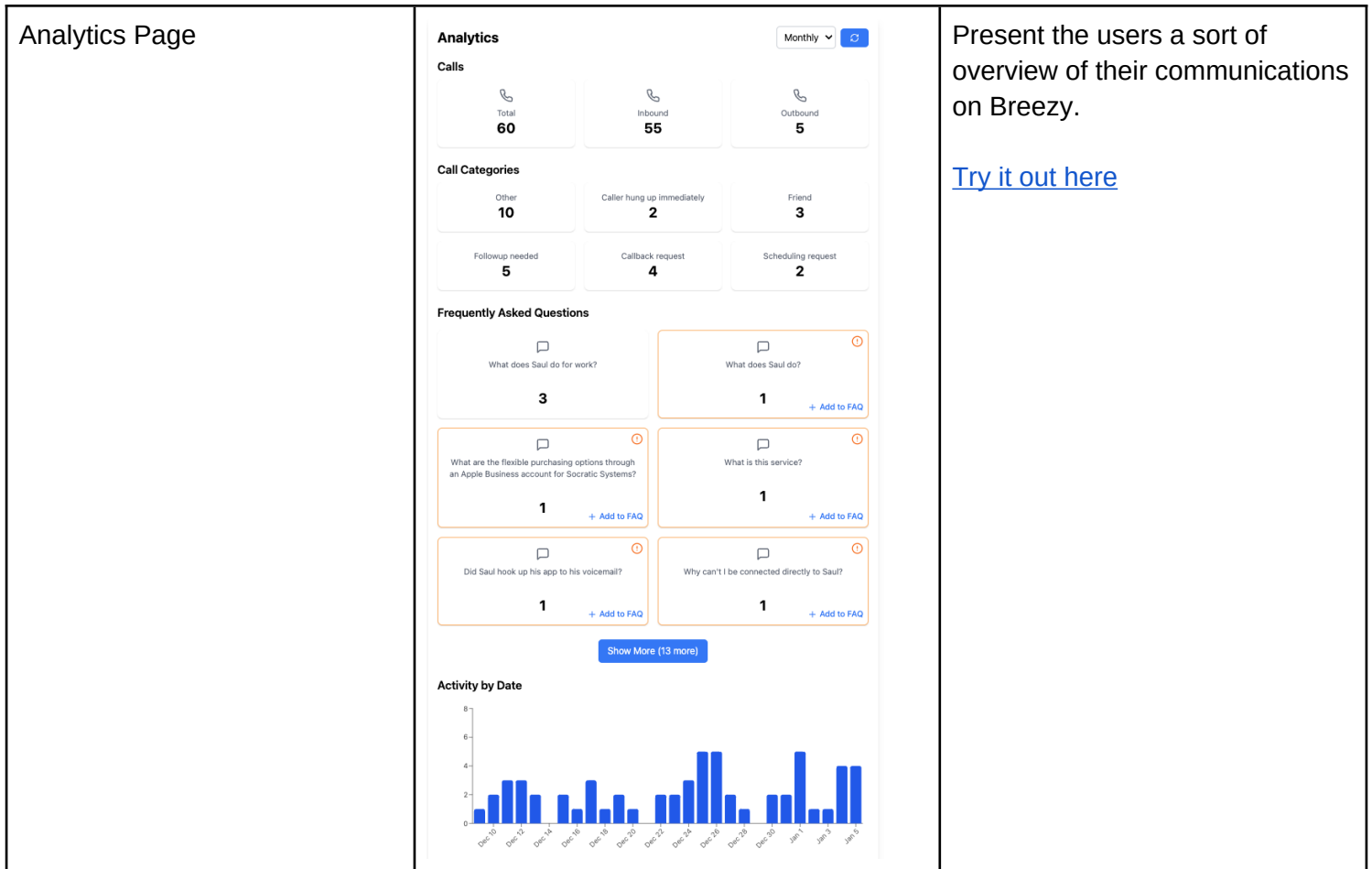
"What are the most common complaints my customers have?"

## Current State of the world

You can get a test account here to try out the [product](#). Use the username and password **prod** and **test**.

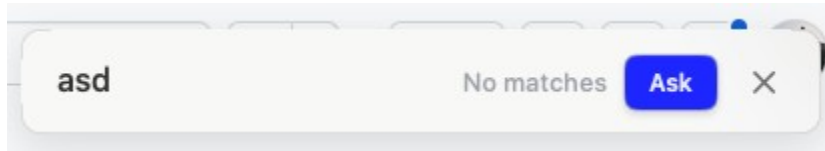
Experience	Screenshot	Description
------------	------------	-------------

<p>Simple matcher search</p>	<div data-bbox="532 210 1015 703"><p>Threads</p><div><input type="text" value="q"/></div><div><span>All</span> <span>Unread</span> <span>Hidden</span></div><div><div><div>N</div><div>Naomi</div><div>Jan 5, 10:41 AM</div><div>Identified as Ashley</div></div><div><div>R</div><div>Ross</div><div>Dec 31, 11:21 PM</div><div>Interested in technology or app integration</div></div><div><div>S</div><div>Saul local</div><div>Dec 27, 1:42 PM</div><div>Hello, I'm calling to schedule an appointment with doctor Smith. Certainly, I'm calling on behal...</div></div><div><div>J</div><div>Justin Fortino</div><div>Dec 4, 4:12 PM</div><div>Reaching out from Sigma Computing again. We're just hoping to have a discussion. Yeah, I not...</div></div><div><div>D</div><div>Diva (Beacon House)</div><div>Oct 24, 8:40 PM</div><div>I'm here with the sign. It gets all, so I can give him these signs.</div></div></div></div>	<p>This allows a user to search through their threads, but the search only works by fuzzy with name or phone number.</p> <p><a href="#">Try it out here</a></p> <p>Code:</p> <pre>breezy-fe/components/contacts_page/ContactsPage.tsx</pre>
<p>Algolia indexing</p>	<div data-bbox="573 819 1006 1764"><p>Search Customer Records</p><p>Phone calls, messages, and tasks</p><p>Search records... <input type="text"/></p><div><div><div>phone_call</div><div><div>+16504178824</div><div>9/27/2024, 6:26:52 PM</div><div>Listen to Audio</div></div><div><div>Task Details</div><div>Title: No message from caller</div><div>Description:</div><div>Urgency:</div><div><div>Conversation Transcript</div><div>Agent: hello this is saul's personal assistant what do</div></div></div></div><div><div>phone_call</div><div><div>+16504178824</div><div>9/27/2024, 6:26:08 PM</div><div>Listen to Audio</div></div><div><div>Task Details</div><div>Title: No message from caller</div><div>Description:</div><div>Urgency:</div><div><div>Conversation Transcript</div><div>Agent: hello this is saul's personal assistant what do you need</div></div></div></div><div><div>phone_call</div><div><div>+12406013642</div><div>9/27/2024, 10:25:52 AM</div><div>Listen to Audio</div></div><div><div>Task Details</div><div>Title: Request for callback from Saul</div><div>Description: The caller requested a callback but the conversation was interrupted, and they confirmed their contact number before the call dropped.</div><div>Urgency: low</div><div>Follow-ups:<ul style="list-style-type: none"><li>Pass the message to Saul for a callback</li></ul></div><div><div>Conversation Transcript</div><div>Agent: hello this is saul's personal assistant what do you need</div><div>User: hey I'm hoping you can call me back just have a hot</div><div>Agent: sure</div><div>User: I hear that</div></div></div></div></div></div>	<p>This is a more sophisticated search system that actually looks at call transcripts. It's really fast. It uses Algolia (which is basically a managed Elasticsearch service). This is in production but there is no entry point to it, so users basically never look at it. We can say it's never been released to the users.</p> <p><a href="#">Try it out here</a></p> <p>Frontend Code:</p> <pre>breezy-fe/app/service_provider/dashboard/[professional_token]/search/page.tsx</pre> <p>Backend Indexing Code:</p> <pre>socratic_systems/app/actions/search/index_phone_call_actio n.rb</pre>



Replace the UX in the “Threads” screen

We can use the Algolia search thing for most queries, and then use RAG on more complex ones. Like how the arc browser does it (“No matches [Ask]”)

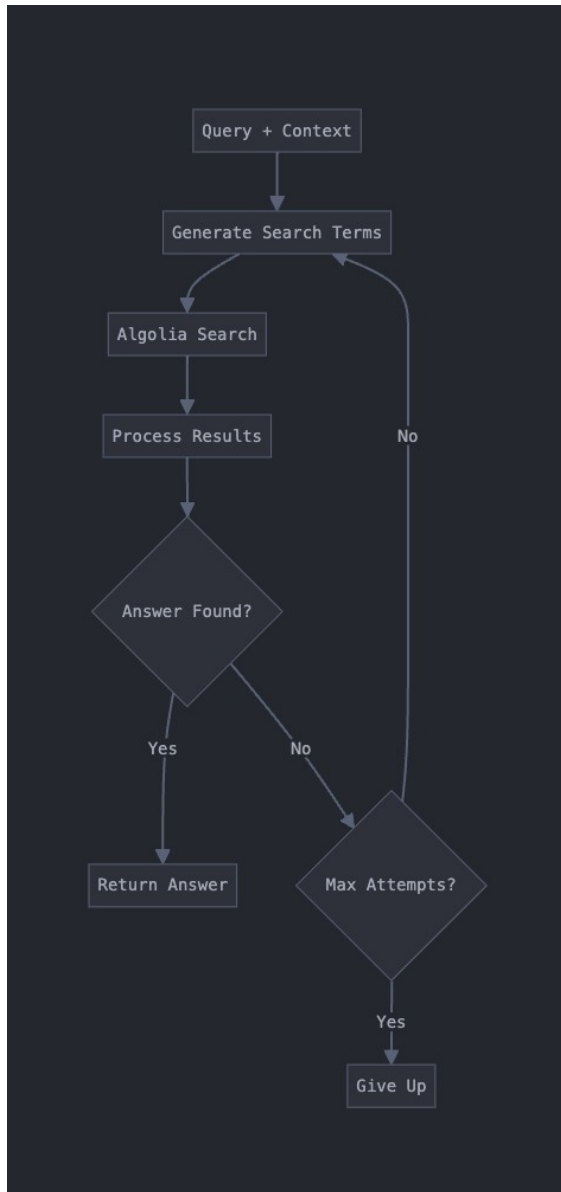


Ask should trigger a basic RAG sort of thing (LLM decides what search queries to make, makes them, synthesize the result into a final answer)

### How would the RAG system actually work?

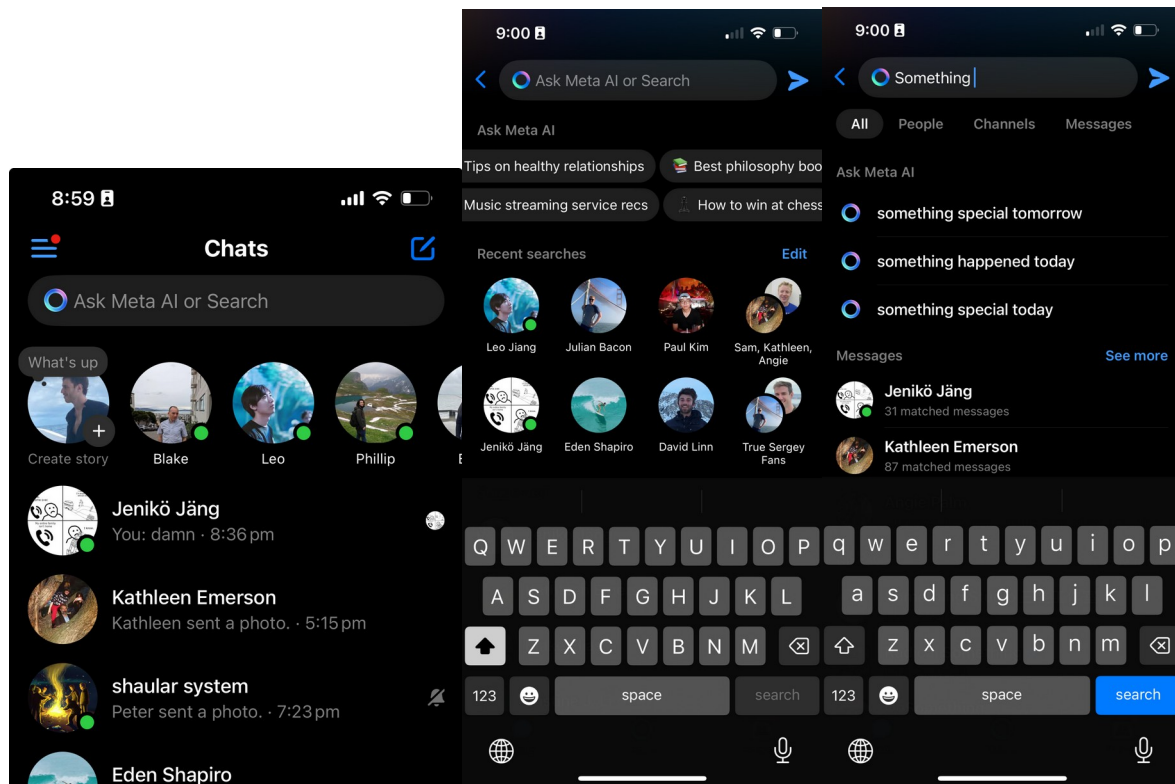
We would use [vercel ai](#) “tool use”. This will allow us to easily stream LLM output (the generation part of the RAG) in the frontend. In practice it would look something like this:

1. Take the search query and some context about the customer and generate a search term and filtering parameters.
2. Do the algolia search (using the backend api)
3. Take the algolia result + the initial query and context and run a
4. Come up with an answer or another search. If nothing works after a few searches, let the user know that you want to give up (we should use [tool use](#) for this one)



Each step of this process should be displayed to the user. I like how Cursor does this sort of stuff. [Vercel](#) has some cool examples of this sort of stuff with some code too. We might want to turn on the [algolia semantic search](#) system to improve performance.

In terms of UX, I really like what FB Messenger does. They even use similar naming to us (“Ask Meta AI” = “Ask Breezy”). Zoom in to see what it looks like:



I think the transition from the main screen to the search screen is also a really neat touch.

## Phase 2: Conversational Search

Very similar to the above. Basically providing an LLM with tools to query conversations and creating an interface for our customers to interact with it.

At that point, it might make sense to use embedding in addition to algolia to improve quality.

Define and track search quality metrics.

You can text ask breezy

## Phase 3: More modalities

In-thread / notifications:

- Ability to respond to summary notifications / thread “give me a summary”

Ability to get Breezy to take actions from notification: text XYZ back

Support aggregate questions (“how many of my customers are x”)

Send searcher to the specific point in a conversation when something was said

Support “advice” questions (e.g, what should i do to get x)

You can call ask breezy

#### **Phase 4: Support “Write” Actions**

Add some writing tools here. At this point, we are offering assistance.

Ability to trigger SMS and text calls from “Ask Breezy”

Note taking assistant via owner-mode (calling or texts)



**kevin** 10:08 AM

we had a customer who works on wrapping ve  
be able to have “owner-mode” so if he texts or  
wanted to review measurements for the helico

2 files ▼

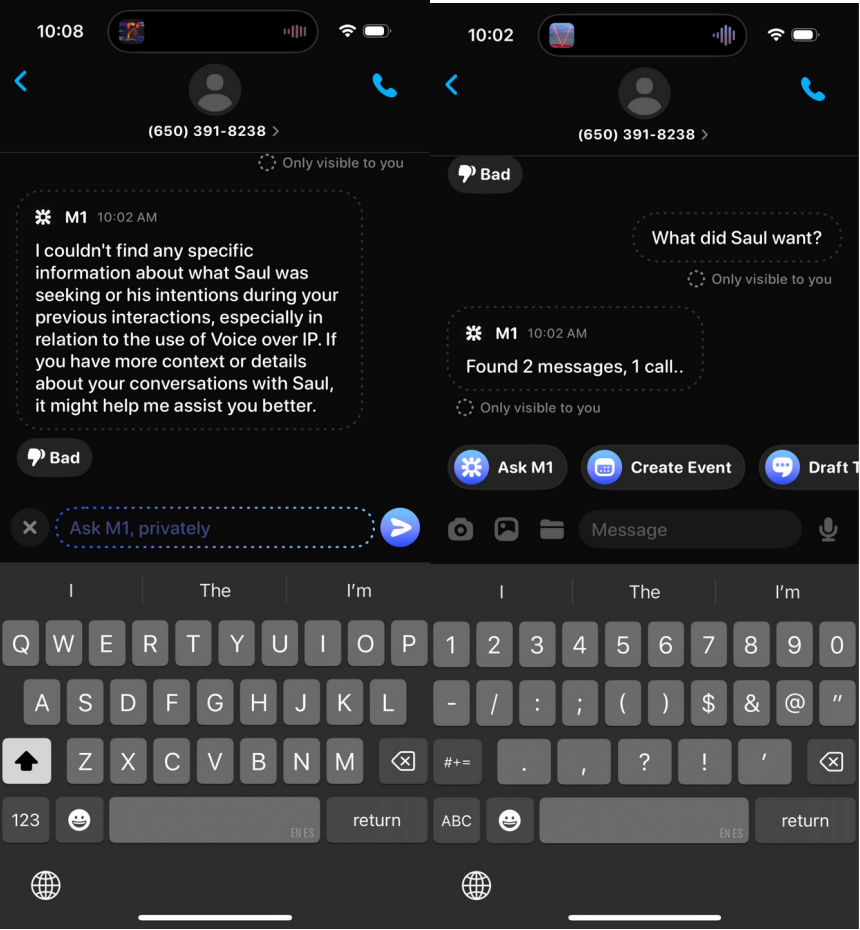


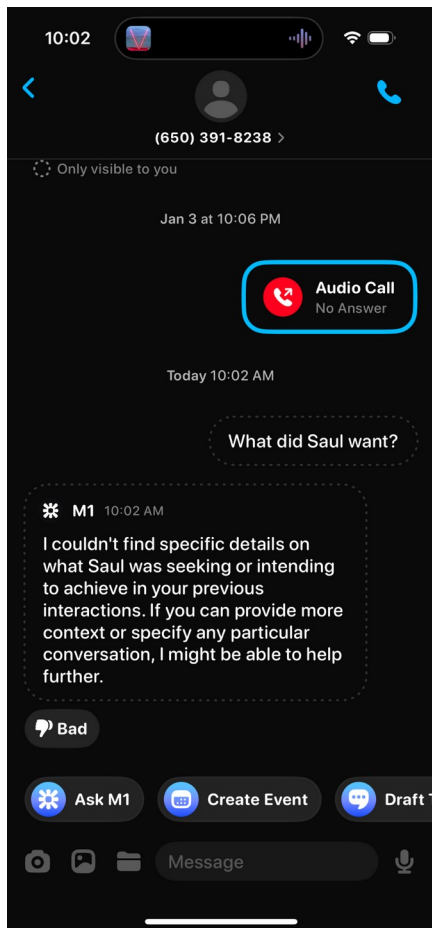
## UX Comparables

Put M1 and Google Gemini, Glean here.

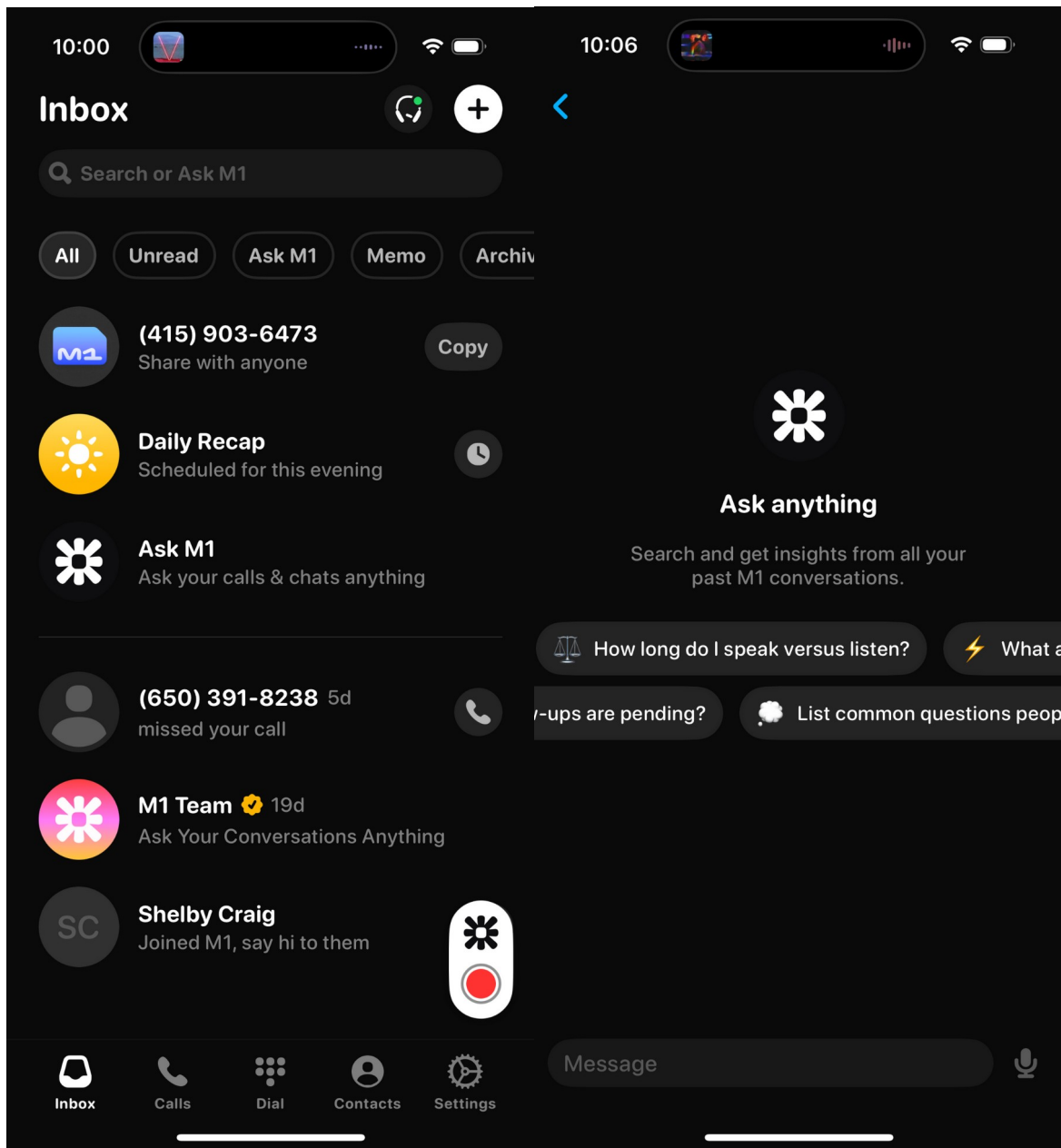


M1 in a single thread => it will review over thread content

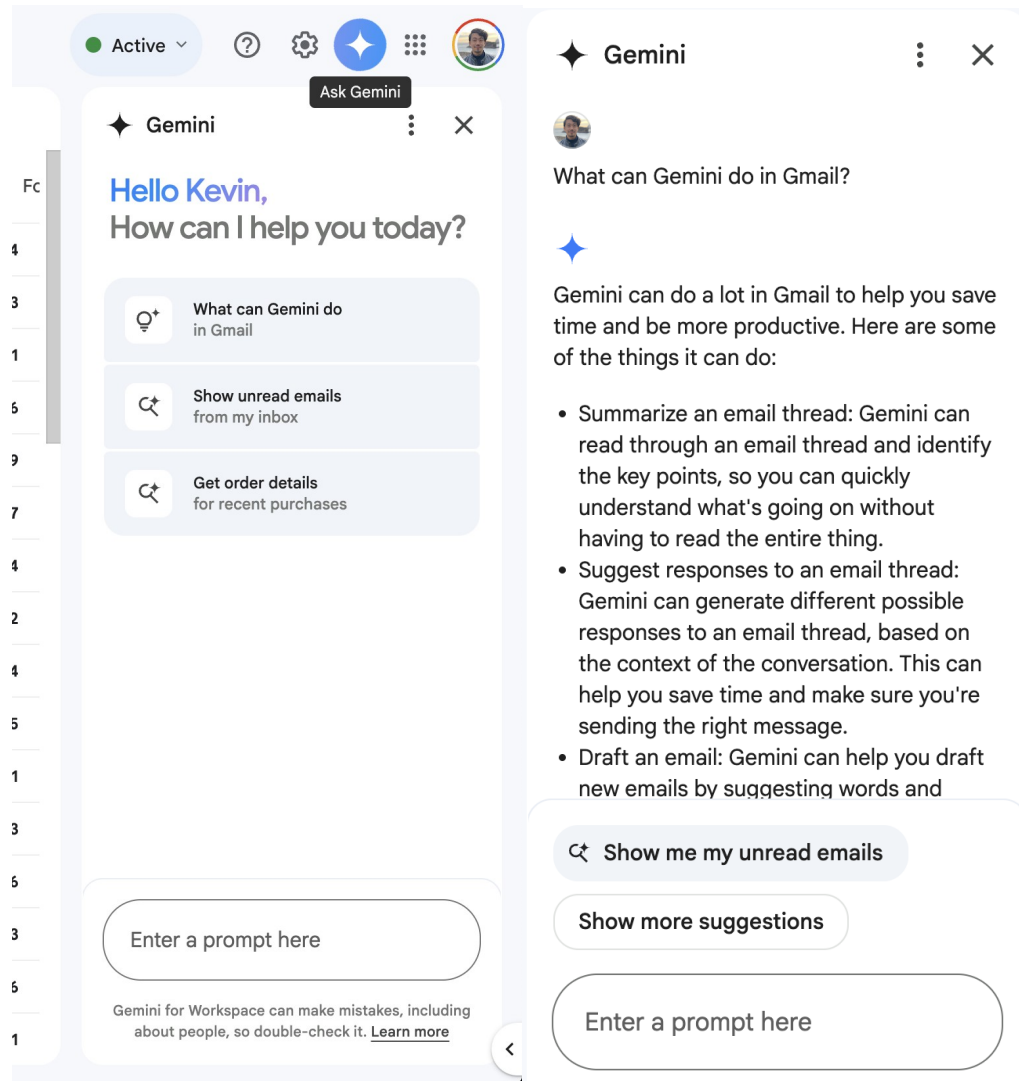




Example of higher-level / general use case (advice, what's going on, etc)



Gemini – general example



Gemini – email thread specific example

1 of 76,125

Get a head start on a healthy year

Inbox

Summarise this email

One Medical

<admin@access.onemedical.com>

Unsubscribe

10:04 (11 minutes ago)

to me

one medical

2025 looks good on you

Staying healthy in 2025 doesn't have to mean a total life overhaul. Small tweaks and habits can make a big difference and set you up for success this year. We're here to help.

Book an appointment with your provider to talk about where you're at physically and mentally — and make a plan to achieve the right health goals for you.

Book appointment

Gemini

Here are some resources to learn more about how to use Gemini in Gmail:

- Collaborate with Gemini in Gmail (Workspace Labs) - Android (Google Help):  
<https://support.google.com/mail/answer/14199860?hl=en&co=GENIE.Platform%3DAndroid>
- Collaborate with Gemini in Gmail - Android (Google Help):  
<https://support.google.com/mail/answer/14355636?hl=en&co=GENIE.Platform%3DAndroid>
- Gemini for Google Workspace | Gen AI Tools for Business:  
<https://workspace.google.com/solutions/ai/>

What's this email about?

What are the action items?

Suggest a reply

Show fewer suggestions

Enter a prompt here

Gemini for Workspace can make mistakes, including about people, so double-check it. [Learn more](#)