Mapi: Natural Language to API Execution

Ayotomiwa Akinyele, Daniel Bessonov, Justin Zhou, Patrick Li, Saaketh Narayan

Advisor: Prof. Jonathan M. Smith

Executive Summary

Mapi is a platform that lets people talk to any application and have it complete an action. For example, you can talk to your calendar app to schedule a coffee with an old friend without clicking through the scheduling process.

Unlike Siri or Alexa, developers do not need to update their app to support natural language input. Mapi automatically understands how to use an application, and can convert a users' natural language request into a sequence of API requests for that app.

With Mapi, businesses and consumers alike enjoy fast response times, a consistent experience across channels, and an enhanced online experience. Mapi, as part of the artificial intelligence revolution, unlocks your time to focus on the things that matter.

Value Proposition

Mapi is a technology platform that gives users the ability to generate and execute actions against application interfaces.

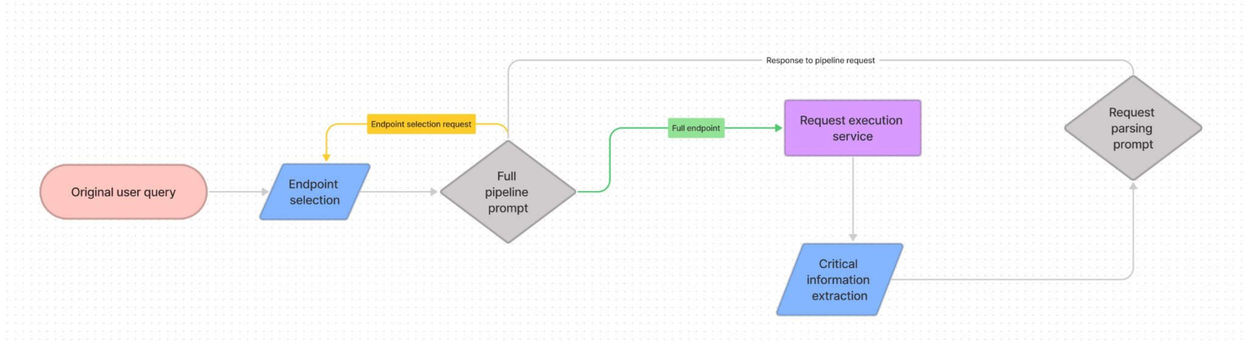
For example, let's assume Google Calendar used Mapi. A user could prompt, "I want to update my meeting with Saaketh to 3:30 PM." Mapi would read that prompt, structure the correct API request to reschedule the meeting, and execute the request against the Google Calendar API.

On a high level, interaction with software applications mostly happens in two ways. First, people interact with some visual interface over an API. Second, applications interact with other applications through standard procedures like HTTP or GRPC. There have been efforts to create voice or free-text interfaces, but those efforts are typically restricted to the most well-resourced development teams, like Amazon's Alexa or Apple's Siri.

Mapi allows any developer to add a free-text interface without a world-class NLP team. They simply need to expose their application's API and authentication credentials to Mapi, and their application is now natural-language-ready.

Engineering Innovation

Our main contribution is a novel dataflow model that iteratively handles multi-endpoint selection, request construction, response handling, and error fixing. The model diagram can be found below.


Response to pipeline request
Endpoint selection request
Original user query
Endpoint selection
Full pipeline prompt
Full endpoint
Request execution service
Request parsing prompt
Critical information extraction

While LLMs can be used for classification tasks, in order to accurately understand a user's query, break this down into a sequence of API calls, execute calls, use the information from responses correctly, and fix errors, a more sophisticated approach than simple prompting is needed. Our new pipeline is inspired by cutting-edge research into LLM tool use, and self-evaluation. A user first submits a query, which is fed into the endpoint selection model in order to select a set of candidate endpoints that may be useful for completing the query. We combine this with known information about the user and application, and pass this into a full pipeline prompt that fills out the request itself. The pipeline prompt may optionally call the endpoint selection model again, enabling multi-endpoint functionality. The full pipeline prompt can also identify information the request may still need, resulting in subsequent calls to the endpoint selection model to retrieve that information. These requests are executed in sequence, critical information is extracted, and if needed, send back to the full pipeline prompt. Errors are handled gracefully, and LLM hallucination is addressed by constraining the LLM output to deal only with the pipeline and API calls.

This method is generalizable to any API – we have demonstrated its use on a sample Shopify storefront here.

Stakeholders

In context of Mapi's technology platform, relevant stakeholders include

- **Platform teams within companies** that will be able to interact with Mapi's mapping from natural language to APIs by connecting custom playbooks and resources.
- **Users of the platforms** will rely on Mapi service to communicate their desired action through natural language, with proper verification before irreversible actions are taken.
- **Third party services** are services or platforms that Mapi may need to interact with in the servicing of a user request to a specific platform.

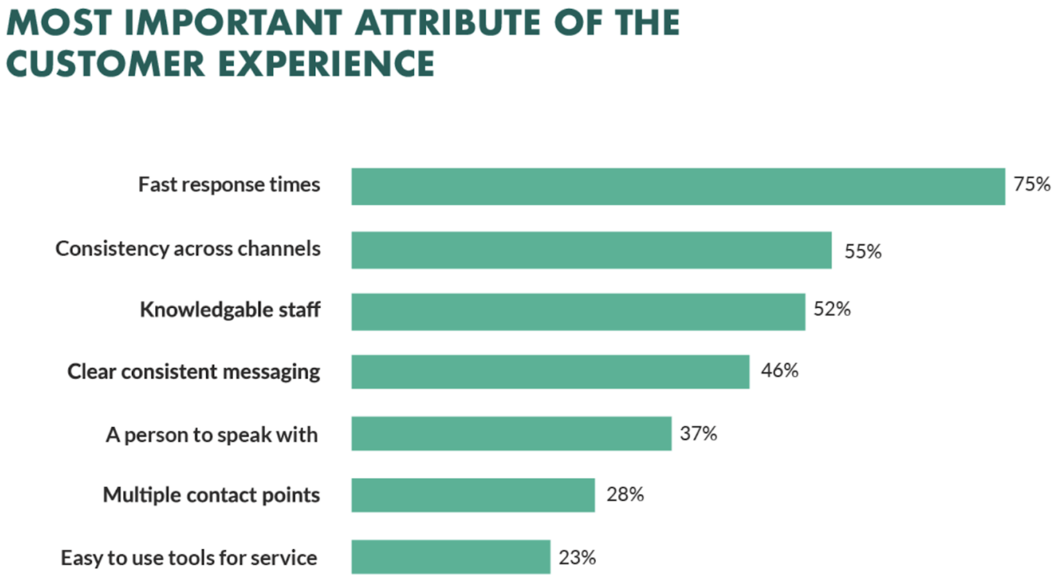
Market Research

Research advancements in Artificial Intelligence in the past few years has produced entirely new ways of interacting with the applications we're familiar with. Voice-to-text technology such as OpenAI's whisper have become highly accurate and less error prone, making them reliable for a variety of applications. Moreover, OpenAI's ChatGPT and Google's Bard have pioneered a new form of interaction with AI that resembles that of an Artificial General Intelligence(AGI) – an Oracle that can provide reasonable and relevant responses across a wide range of subjects.

However, one missing step in the journey to AGI is including the capabilities to affect the real world. Google's PaLM-SayCan is directionally relevant as it combines the new powers of Artificial Intelligence, with the real world execution of robots. You input a command like "get me coffee" and the robot proceeds to implement the user request.

Sayscan is evidence of the need for leveraging AI to actually do things (as opposed to just respond) and Mapi is solving for this problem in the software realm. As far as we can tell, the market includes use cases for traditional customer support actions, travel and hotel booking, internal tooling, and so on.

Specifically for customer support, we see that the most important factors are fast response times and consistency across channels. By being faster than human-mediated customer support, more accurate than existing automatic customer support, and interacting with an API that is common across channels, Mapi has the opportunity to enhance customer experiences across online services.



There are specific customer segments that are more attractive for Mapi – it will appeal to Platform teams that fall into one of the following cases:

- Customer Support
  - Typical use cases include returning an order from a site like Amazon, checking the delivery date for a shipped item, and so on.
- Travel and Hotel Booking
  - Typical use cases include retrieving flights for a certain condition, and so on.
- Internal Tooling & HR
  - Typical use cases include submitting routine work orders or reimbursement forms, and so on
- Call Center Integrations
  - Typical use cases include translating a customer inquiry to text, then passing it to Mapi to see if there are any actions that could spawn from that call.
- And other platform teams that rely on human hours to map prompts to apis. This usually involves some kind of playbook that dictates the rules around these mappings.

The global customer experience management market was valued at $9.5 billion in 2021 and is estimated to reach $16.9 billion by 2026, Here. The call center market is $30.3 billion in 2023. The online travel booking market is $519 billion. It's difficult to estimate what share of these industries are capturable, but we can assume that any role involving manual mapping of customer input and action is within our domain; given that a majority of jobs in these industries is consumer facing, we can assume that Mapi can address a significant portion of these global markets. As Mapi is at the intersection of these three large markets, we anticipate that the total addressable market for Mapi lies in the billions of dollars.

Competition

- Apache NLPCraft: Apache NLPCraft is an open source library for adding natural language interface to modern applications. It enables people to interact with products using voice or text and uses JVM-based languages like Java, Scala, Groovy, or Kotlin to interpret user input. It is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by the Apache Incubator and licensed under the Apache License, Version 2.0..
- Microsoft research: APIs are playing an increasingly important role in both the virtual and physical world, being used by web services and IoT devices. To make the consumer experience even easier, natural language interfaces (NLIs) are being developed to provide a unified and intelligent gateway. Microsoft

researchers are specifically studying NL2APIs, which are NLIs for individual APIs, to help users communicate with software systems, reduce integration costs, and facilitate web service discovery. While Microsoft has begun research into this area, it has not taken the research to a production level, nor has it made it available to the public as a product.

- Adept AI: They are a general intelligence research lab whose goal is to build a virtual assistant to interact with any suite of tools. They interface with any digital product, from search engines to desktop applications, and can chain together multiple actions at the request of a query. While Adept is building for end consumers, we are interfacing directly with platform teams and businesses.

IP

Our IP consists of:

1. The dataflow model itself, described earlier in section “Engineering Innovation”.
2. The implementation of the dataflow model, including ingesting user input, executing and handling requests and responses, various LLM prompts, the endpoint selection model, etc.
3. Specific integrations we have built for Chrome and Shopify.

Cost

Core assumptions: 100 users per platform per month, 50 endpoints per platform

Cost Type	Monthly Cost per Platform($)
AI Model costs: Embeddings, GPT-3 fine-tuning, GPT-3 api calls	$0, $1 per 200 words, $1 per 20k words = 0 + 50 + unknown (depends on optimality of orchestration service)
Server hosting	$50
Developer fees	$114,000 yearly salary

Calculating Average Cost per Mapi Request

# of embeddings = 3

Cost per large embedding = 0.0008192

# of chat requests = 7

Cost per chat completion = 0.024576

Avg. Cost per Mapi Req. = 7 * 0.024576 + 3 * 0.0008192 = $0.1744896

Revenue Model

Assumptions:

1. 2-5% freemium conversion rate for B2B saas
2. The average cost to outsource customer service is $2,600 – $3,400 per agent per month.
3. A customer service agent takes a month to onboard, and can make 60 mappings an hour. 180 hours of work a month. We're going to assume the cost of an agent is $3,000 per month. $3000 / (180*60) = $0.277
4. 28 cents per mapping

We present a free and premium option:

- Free model:
  - Free for first 100 user requests
- Premium model:
  - We charge 20 cents per request after the first 100 requests, that is $20 per 100 requests.

We chose the freemium model because we want as little friction as possible when connecting with customer API endpoints. By making the integration free, there is little reason for the business user to run through our free trial. Once we've demonstrated that our offering works to improve their bottom line, we can begin charging by number of requests. We chose to bill based on the number of requests because it is very easy to demonstrate how much the business is saving per request. We've estimated that each manual human mapping costs around 28 cents, so we're $20 per 100 requests (20 cents per request), cutting our clients costs by about 30%.