

The background features a dark blue gradient with large, overlapping, semi-transparent shapes in shades of purple and magenta. Thin, light blue lines crisscross the scene, creating a sense of depth and movement. The overall aesthetic is modern and tech-oriented.

# AWS re:Invent

DECEMBER 2 - 6, 2024 | LAS VEGAS, NV

AIM304

# Using multiple agents for scalable generative AI applications

**Mark Roy**

(he/him)

Technical Lead, AI/ML  
Amazon Web Services

**Michael Y. Liu**

(he/him)

Principal Product Manager  
Amazon Web Services

**Heiko Zuerker**

(he/him)

Principal Architect Cloud & DevOps  
Northwestern Mutual





# Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models

Choice of leading FMs through a single API

Model import, distillation and fine-tuning

Gen AI Tools – Knowledge Bases (RAG), Guardrails, Flows, and Agents

Security, privacy, and data governance

# What is a Gen AI Agent?



Intelligent,  
autonomous  
systems



Plan,  
reason,  
and act



Access to  
enterprise  
data



Ability to  
use tools

# Momentum Behind Amazon Bedrock Agents



Investment  
& medical  
research



Insurance  
claims  
processing



Customer  
experience



Marketing  
assistants



Root cause  
analysis

***Amplify productivity for businesses and  
knowledge workers through the use of  
agents***



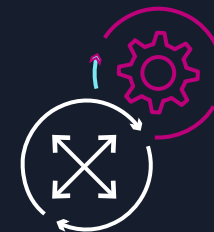
# What generative AI customers are **asking for**



Help me automate  
complex workflows



Help me move  
faster



Help me find  
more robust and  
scalable solutions

# Amazon Bedrock Agents

Agentic building blocks



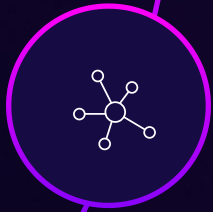
Choice of foundational models



Memory, Knowledge Bases,  
Guardrails



Tools and action groups



Trace, debug, and observability

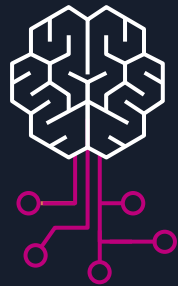


Multi-agent collaboration

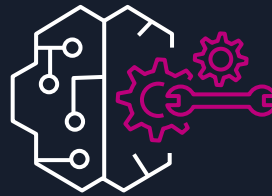
**new**



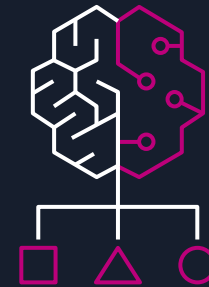
# Amazon Bedrock **multi-agent collaboration**



**Enable complex  
problem solving**



**Increase developer  
velocity**

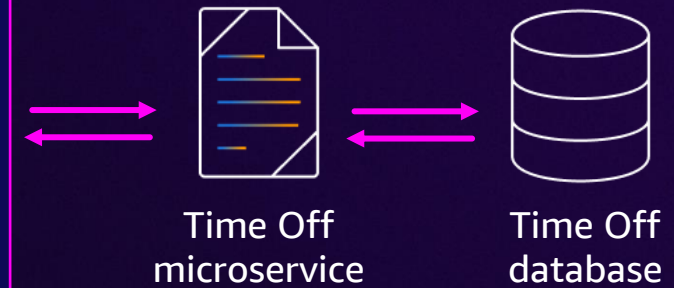
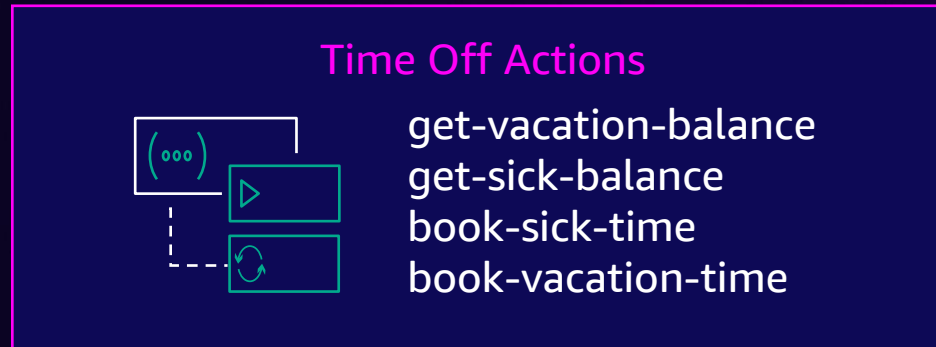


**Deliver trust and  
security**

# Amazon Bedrock multi-agent collaboration deep dive



# Agents start small and focused . . .

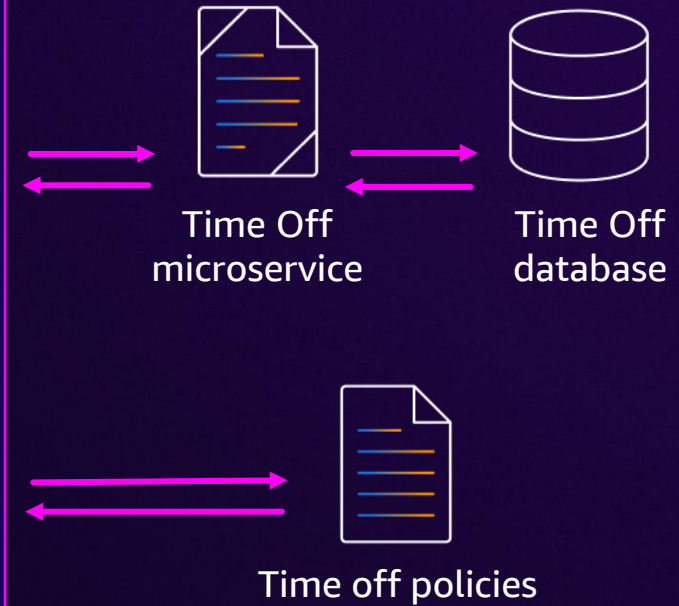
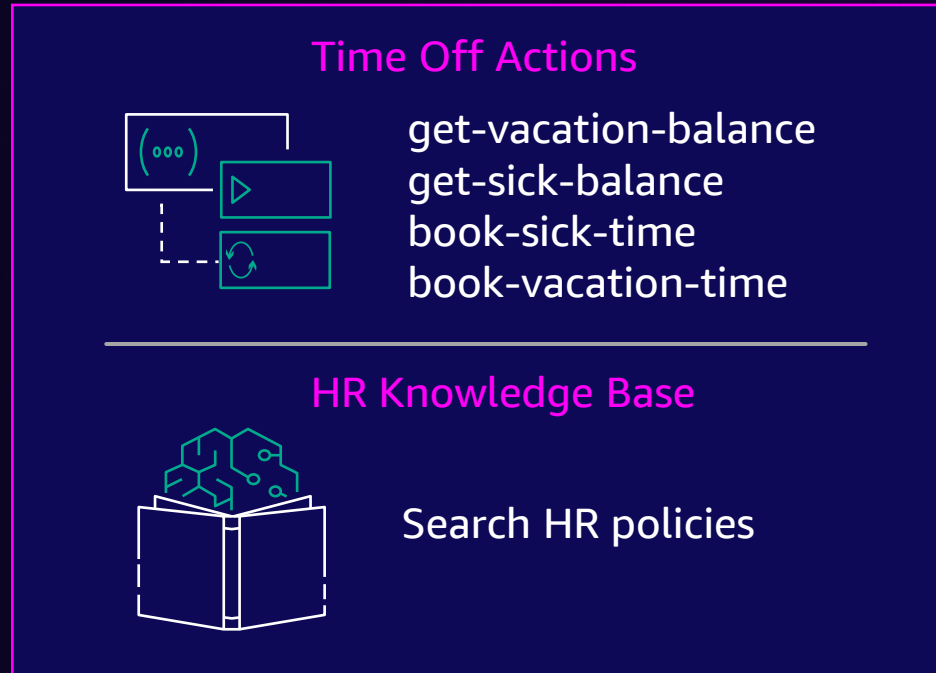


# ... and can be easily expanded

V2



HR Time Off Agent

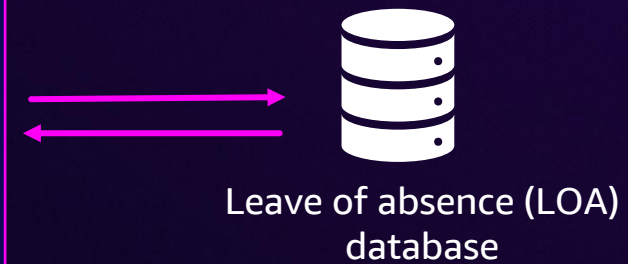
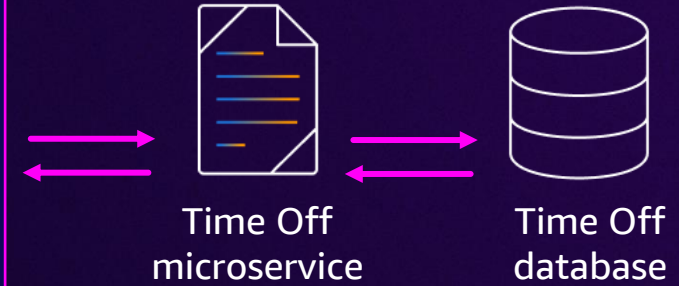
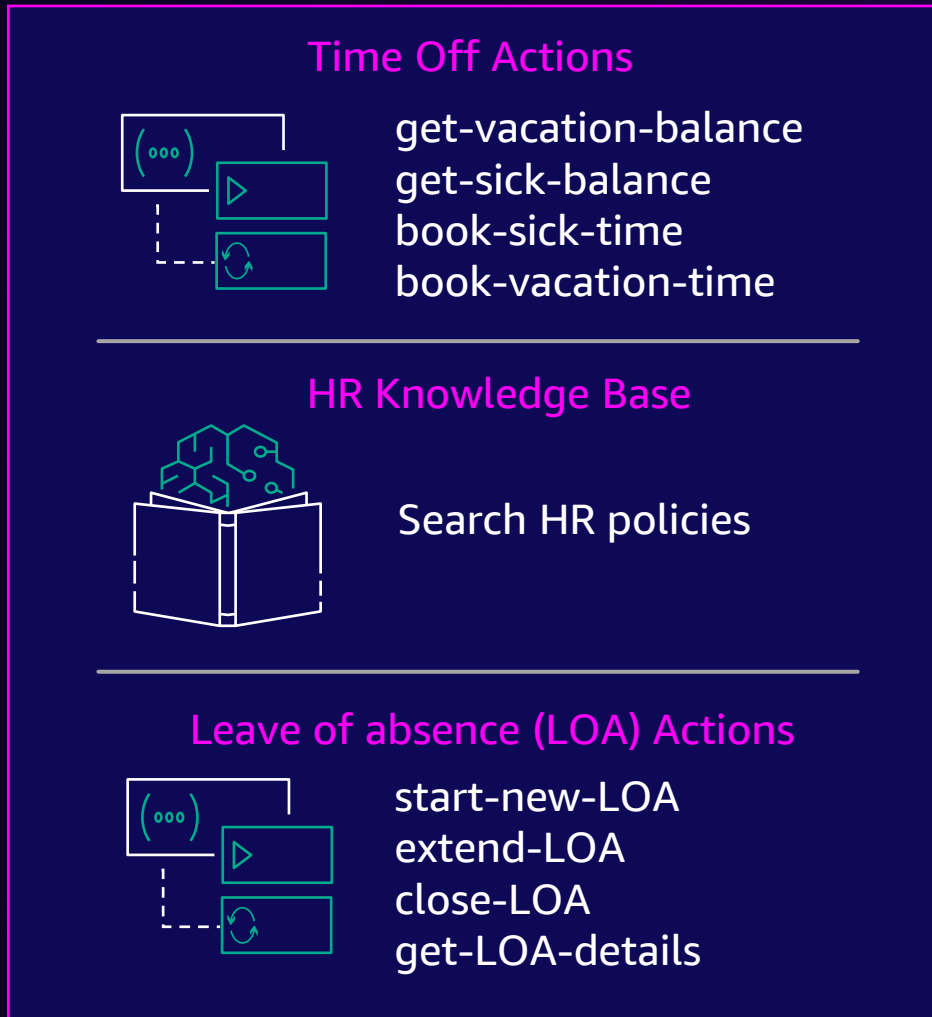


# ... and so on

V3



HR Time Off Agent



# ... but if you take a SINGLE agent too far

V4



**Time Off Actions**  
get-vacation-balance  
get-sick-balance  
book-sick-time  
book-vacation-time

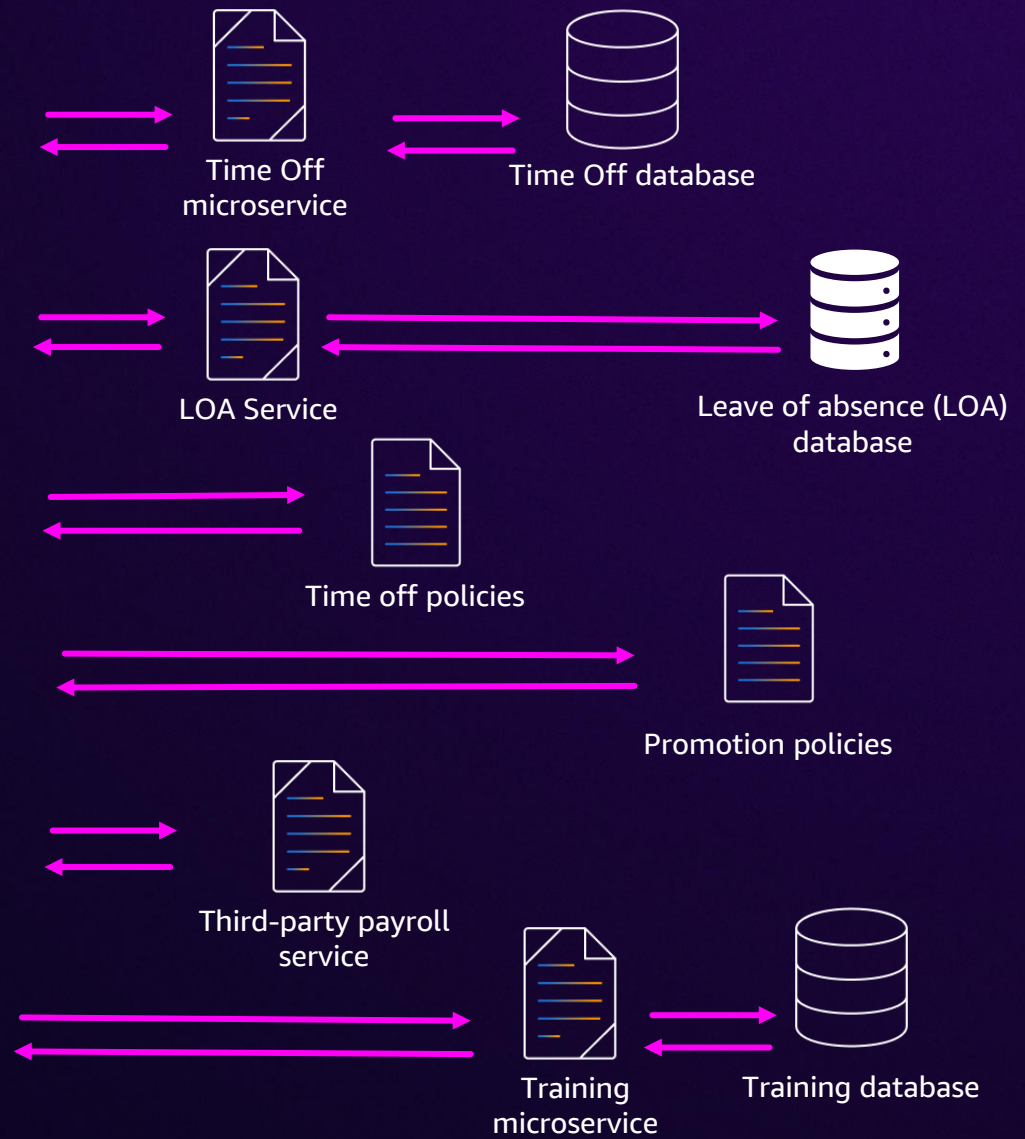
**Leave of Absence (LOA) Actions**  
start-new-LOA  
extend-LOA  
close-LOA  
get-LOA-details

**HR Knowledge Base**  
Search HR policies

**Promotion Knowledge Base**  
Promotion policies

**Payroll Actions**  
get-latest-pay-stub  
submit-payroll-ticket  
manage-deductions

**Training Actions**  
find-training  
list-certifications  
register-for-course  
submit-review



# ... it leads to challenges

## Coding gets complicated



- Complex prompts to limit hallucinations
- Fragile, hard to maintain

## Agent gets confused



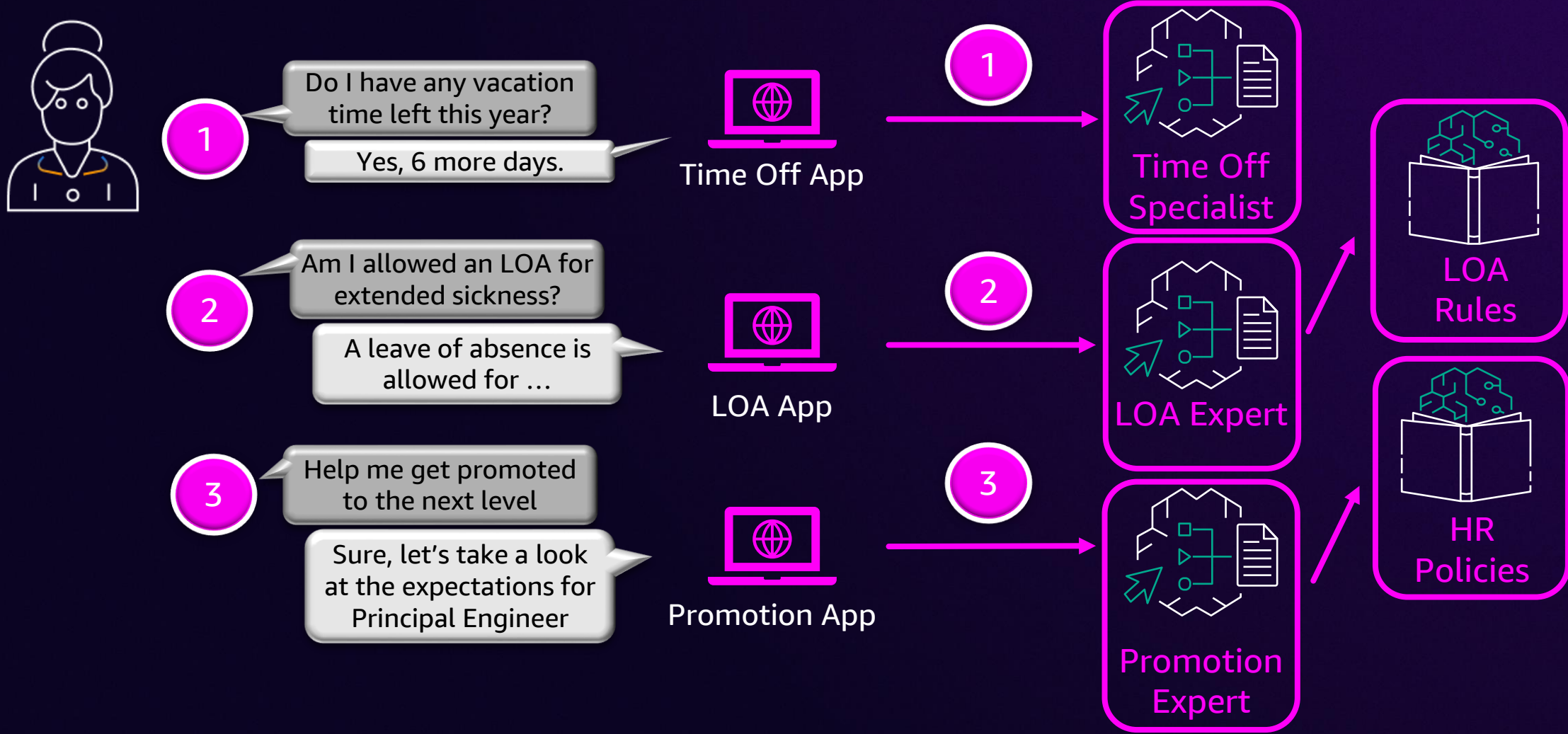
- Calling wrong tools
- Passing wrong arguments
- Inconsistent responses

## Agent gets slower and more expensive



- Frontier models needed
- Prompt sizes grow
- Agents retry steps

# Using multiple agents helps . . .





# ... but single agents are still siloed and limited

## Splintered user experience



- Sea of apps for users to find and navigate

## Do-it-yourself multi-agents is hard



- Needs to be reliable, low latency, scalable
- Takes time away from core business value
- Difficult to keep pace

## Complex automation still unsolved



- No clear path to solving game-changing problems

# Amazon Bedrock Agents multi-agent collaboration

Scaling agentic experiences

Preview



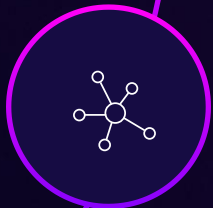
Easily assemble agents and knowledge bases



Plan and execute complex tasks across agents



Unify conversations across agents with built-in intent classification



Observability across multi-agent flows



Guardrails, security, and privacy

# Big opportunities for multi-agent collaboration

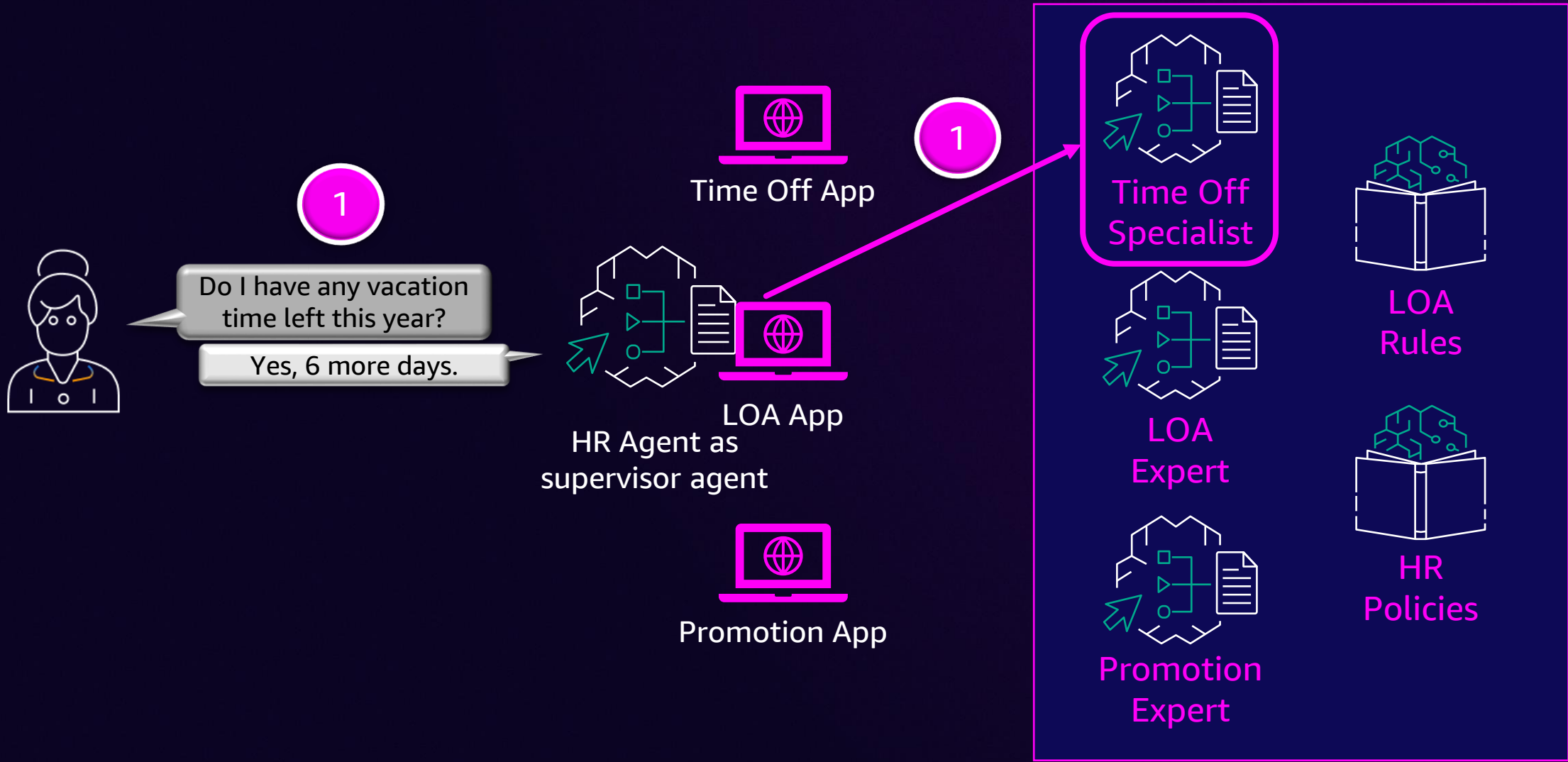


1. Unify customer experience



2. Automate complex processes

# 1. Unify customer experience



# 1. Unify customer experience



# 1. Unify customer experience

Supervisor provides one-stop shop for customers, with optimized routing and seamless conversation handling



Help me get promoted

**One-stop shop**  
Sure, let's take a look at the expectations for Principal Engineer



HR Agent as supervisor agent

**Optimized routing**

3



Time Off Specialist



LOA Rules



LOA Expert



HR Policies



Promotion Expert

# How does routing work?

## Create supervisor agent

```
type = SUPERVISOR_ROUTER  
router model = "small-fast-model:1.0"
```

### Collaboration descriptions



Use for handling vacation and sick time

#### Time Off Specialist



Use for handling Leave of Absence requests

#### LOA Expert

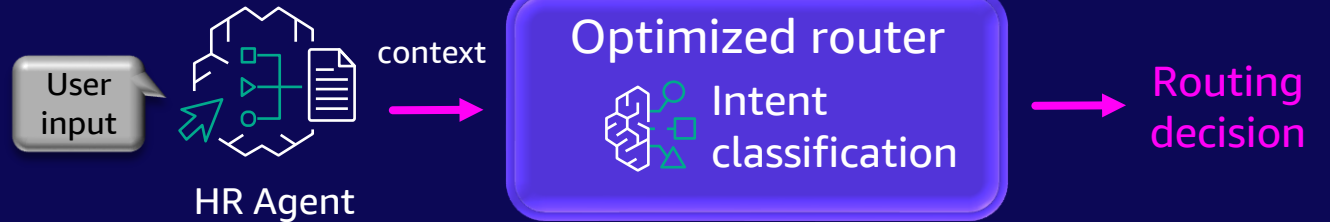


Use for handling Promotion discussions

#### Promotion Expert

## Use supervisor agent

### Collaboration descriptions



### Example conversation

"how much vacation left?"

Use: **Time off Specialist**

"can I use it next week?"

Continue: **Time off Specialist**

"does my LOA affect my promotion?"

Plan: **Time off Specialist AND Promotion Expert**

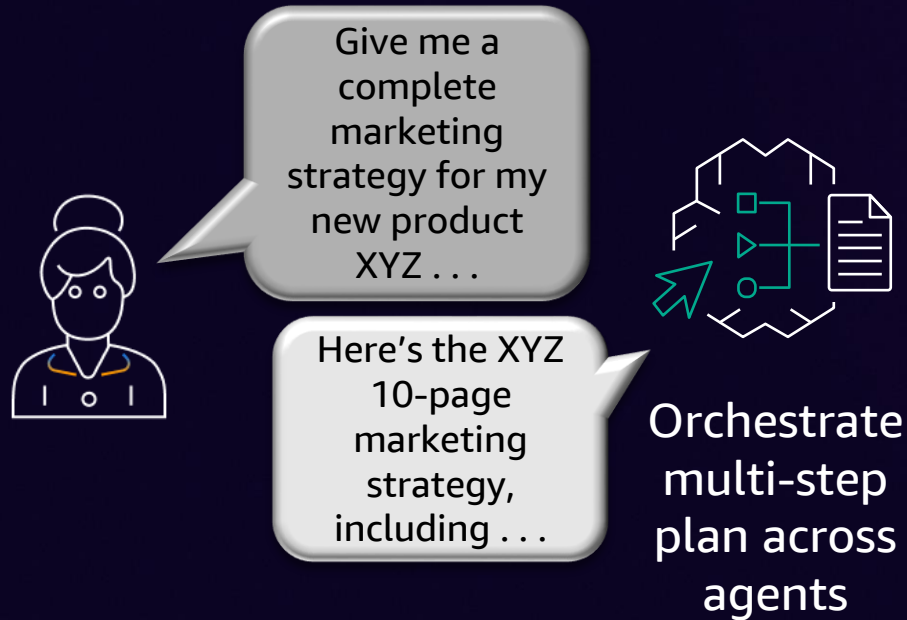
# Big opportunities for multi-agent collaboration



## 2. Automate complex processes



# 2. Automate complex processes with supervisor agents . . .

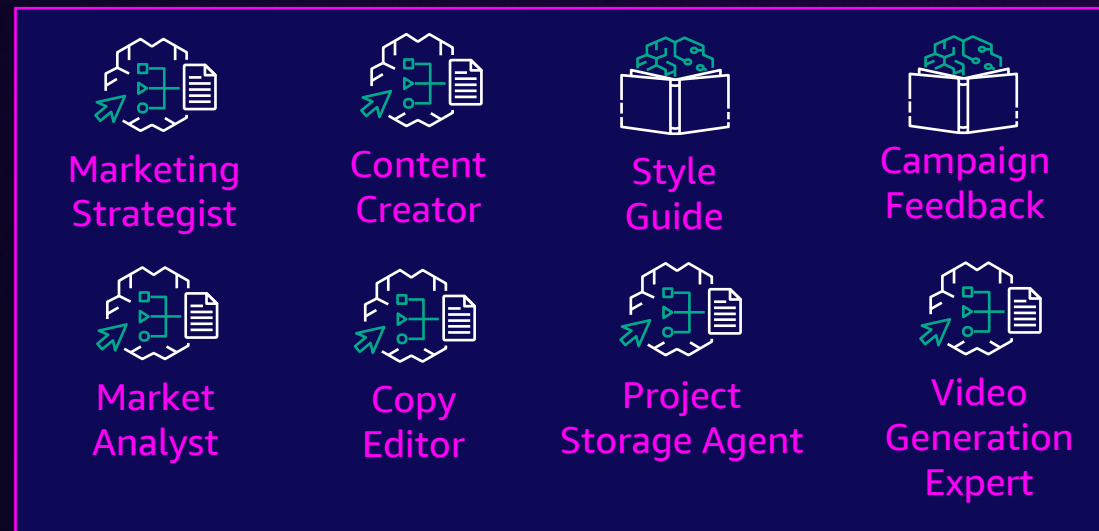


1

## Generate plan dynamically

1. Conduct thorough market research, include competitors
2. Develop detailed project summary, target persona
3. Formulate comprehensive marketing strategy with goals, tactics, channels, KPIs
4. Create three innovative marketing campaign ideas
5. Develop detailed marketing copy for each campaign, including a video ad script, and a draft video
6. Produce final report and save interim results

## Sub-agents and Knowledge bases



2

## Execute plan

# ... executing the plan across agents

**Step 1** – I'll start with some research and analysis

**Step 1.1** – I'll do some web searches, first "Find companies doing ..."

**Step 1.2** – Now, I'll search for demand signals

**Step 1.3** – Now I'll make observations and conclusions

**Step 2** – I've done research. Now let's draft campaigns

**Step 2.1** – Given that research, plus past feedback, I'll give 3 ideas

**Step 3** – Now, let's make content for each campaign

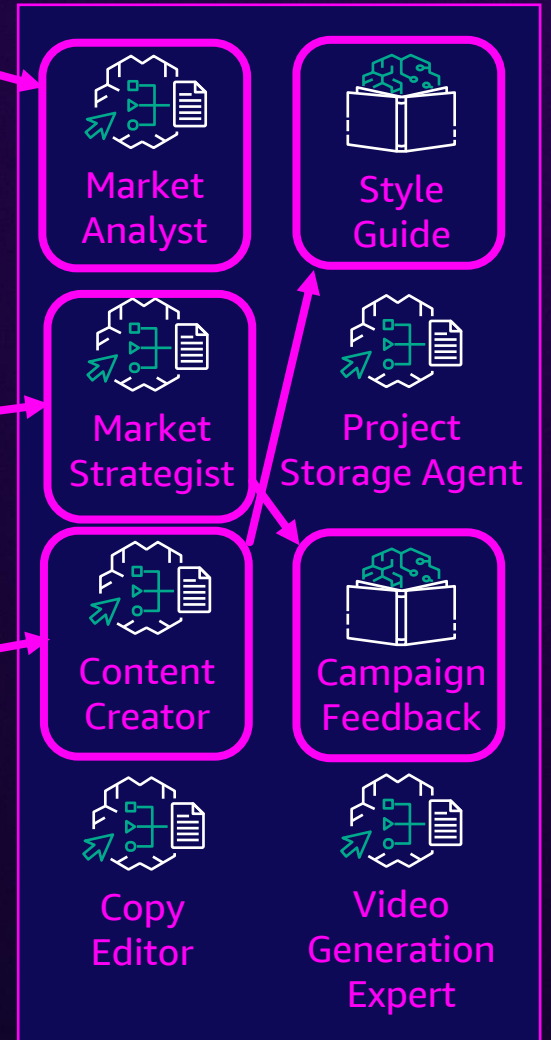
**Step 3.1** – I'll consult the Style Guide, and create content for campaign idea 1

...



Orchestrate multi-step plan across agents

## Sub-agents and Knowledge bases



# Demo 1: Unified customer experience



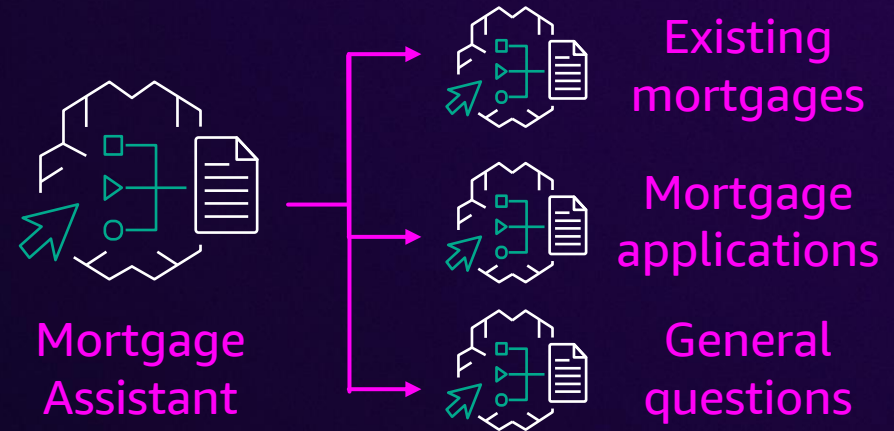
# Octank Mortgages – new Mortgage Assistant

## Before Amazon Bedrock



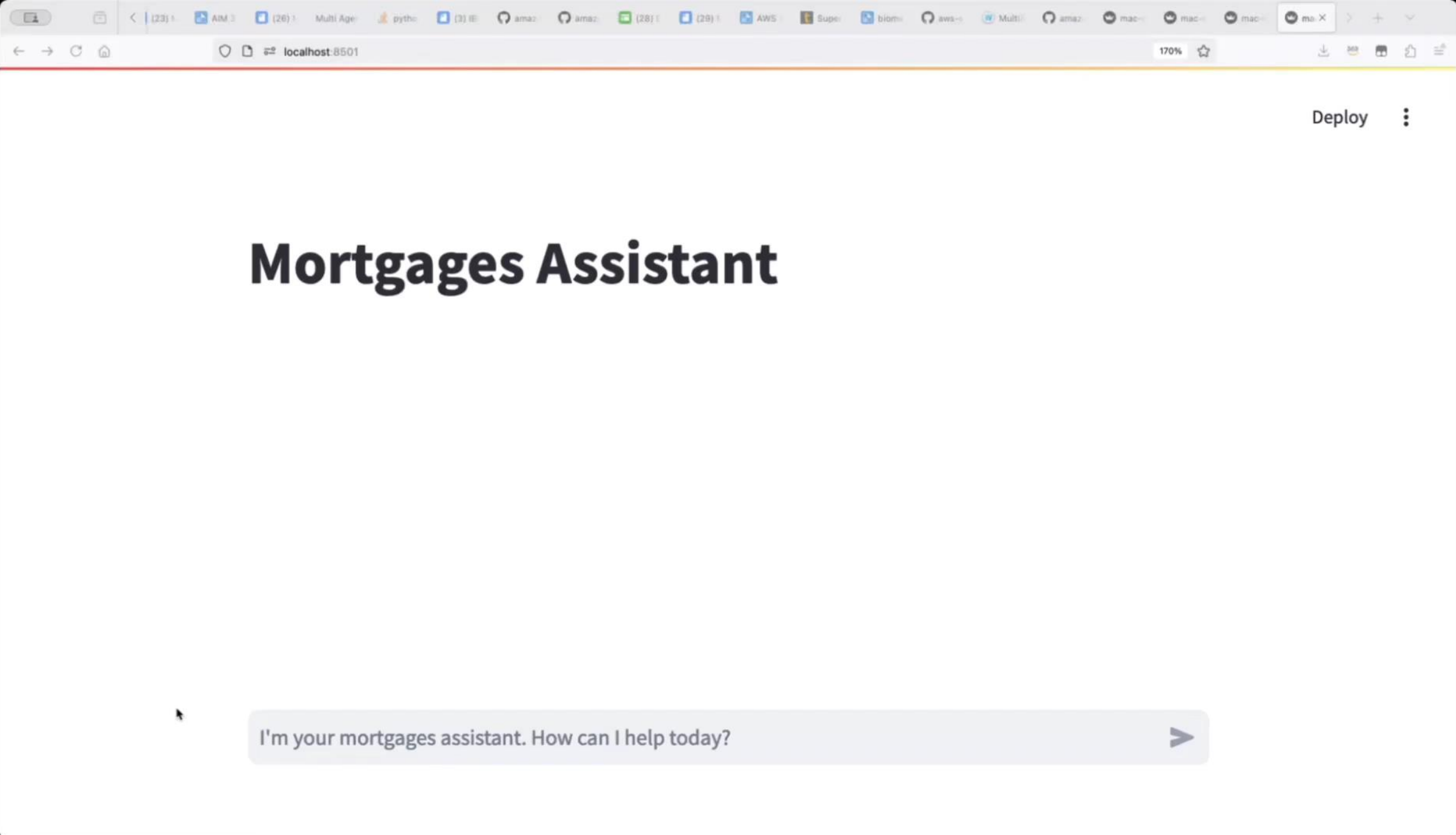
- Voice menus, long waits
- Heavy call center volume
- No chat experience
- Simple web site
- Limited data, static views

## With Amazon Bedrock multi-agent collaboration



- Answers in seconds
- Flexible and personalized
- Call center relief
- Extensible

# Mortgage Assistant – Unified customer experience



# Northwestern Mutual uses Amazon Bedrock Agents





For more than 165 years, Northwestern Mutual has been helping families and businesses achieve financial security.



Revenue \$36 billion  
FORTUNE 500® rank: #110  
10,300+ Financial Professionals  
8,200+ Employees  
Headquartered in Milwaukee, Wisconsin  
Downtown NYC Corporate Office

Figures as of June 24, 2024, unless otherwise noted.

Northwestern Mutual is the marketing name for The Northwestern Mutual Life Insurance Company (NM) and its subsidiaries in Milwaukee, WI.



**Switching from a  
rusty hammer and chisel  
to power tools.**

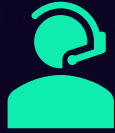
**Kerrigan J.**

Lead Software Engineer at Northwestern Mutual





# Use Case: Developer Support Bot



Enhance developer support experience through chat



Answer questions leveraging internal documentation

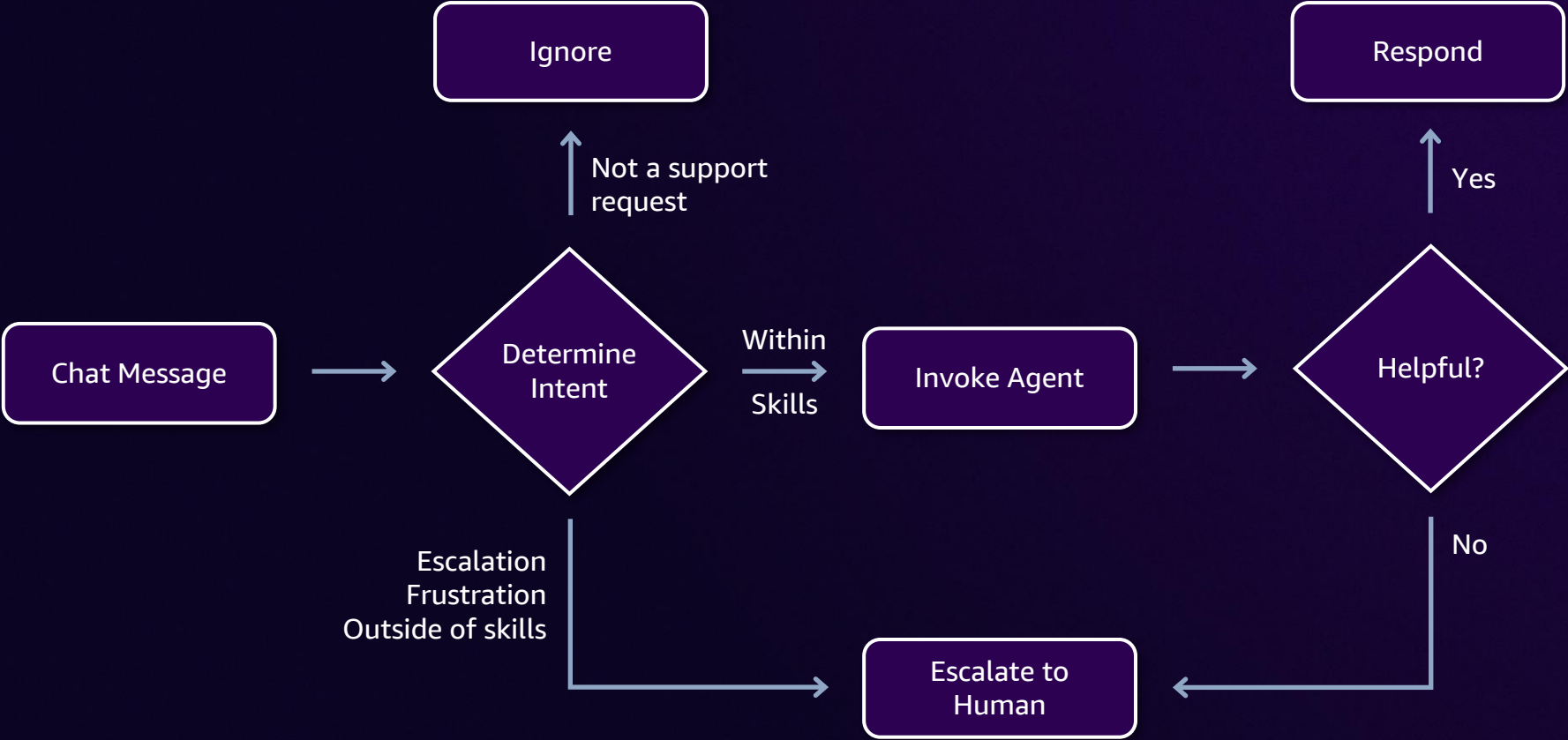


Execute actions in internal Git-based system



Create foundation that meets regulatory requirements

# Agent orchestration – Logical flow



# Multiple Amazon Bedrock Agents

Agent Name	Purpose
Documentation	<ul style="list-style-type: none"><li>• Internal developer documentation</li></ul>
User	<ul style="list-style-type: none"><li>• Unlock git user accounts</li><li>• Rename git accounts for returning users</li></ul>
Repository	<ul style="list-style-type: none"><li>• (Un-)Archive repositories</li><li>• Generate readme files</li></ul>
Pipeline	<ul style="list-style-type: none"><li>• Analyze pipeline failures</li></ul>
Evaluator	<ul style="list-style-type: none"><li>• Ensure helpfulness of answer</li></ul>

# Agent action requirements



1. AI not allowed to take direct action
2. Human must be in the middle



- Agent responds with details on action
- Explicit “yes” or “no” required

**Confirmation:** Before taking any action, ask the user for confirmation by explaining exactly what action you are about to perform and expecting a 'yes' or 'no' answer (case-insensitive). Do not accept anything other than a 'yes' or 'no' answer. Assign this to the 'confirm\_action' parameter of the action group. Only ask for confirmation after all other parameters have been retrieved and an action plan has been created.

# Special case handling



## Escalation

- Identify request to speak to a human
- Detect user frustration
- Request outside of “skills”



## Ignore Agent

- Announcements, greetings, thanks, etc.

I can't get into Git

Yes

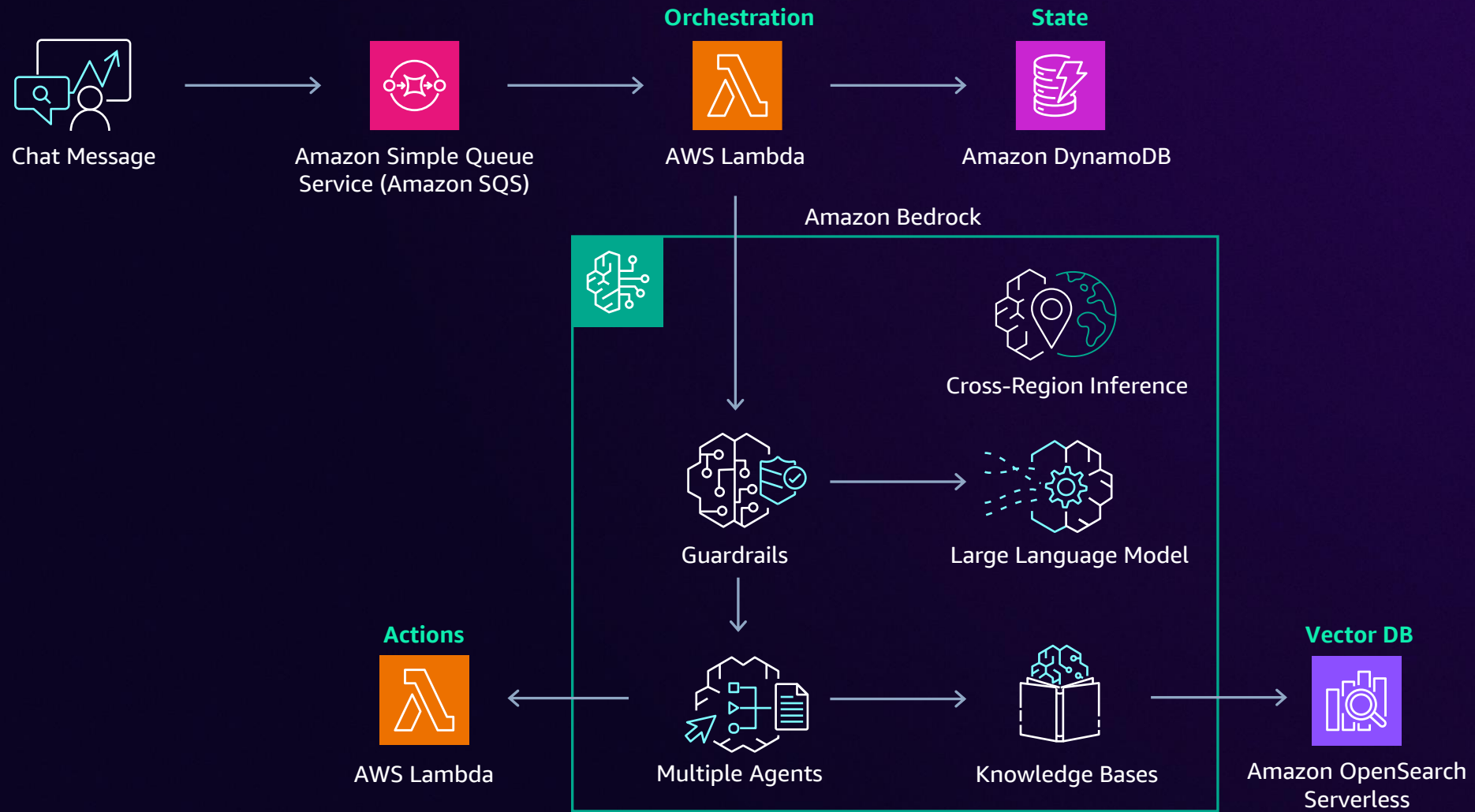
Yeah sure

I'm sorry to hear you're having trouble accessing Git, Heiko. Could you please provide more details about what happens when you try to log in? Also, can you share your LAN ID? This information will help me assist you better.

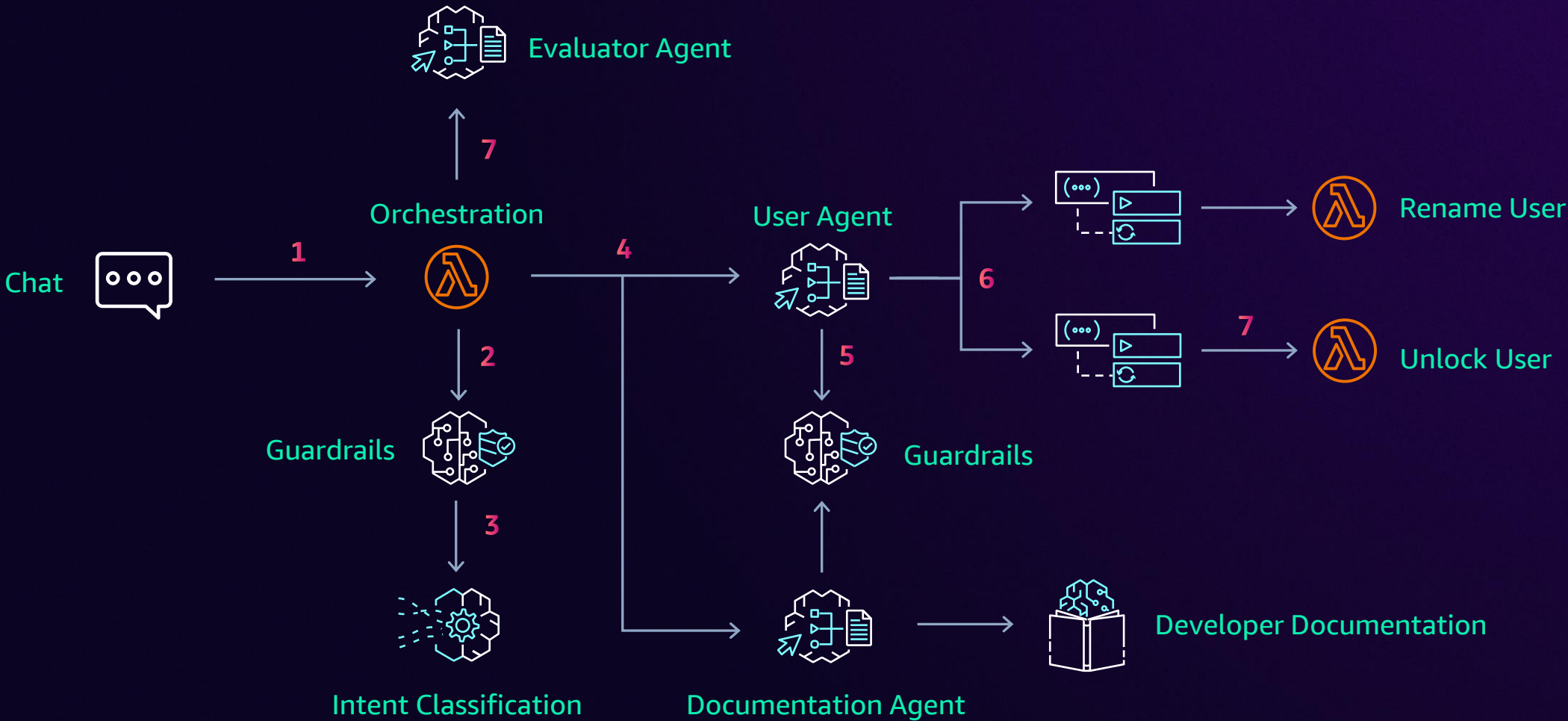
Thank you for providing that information, Heiko. It sounds like your Git account has been blocked. I can help you unblock it using your LAN ID: ABC1234.

Before I proceed, I want to confirm: Should I go ahead and attempt to unblock your Git account? Please respond with either 'yes' or 'no'.

# Chatbot high-level architecture



# Agent orchestration architecture





# Evaluator Agent

- Ensure helpfulness of answer
- Avoid unnecessary noise in support channel
- Provides explanation on decision



Explanation: while the proposed response provides general advice on NPM authentication and Docker build processes, it does not align with the AI's available actions. **The response suggests investigating pipeline configurations and Docker settings, which are beyond the system's capabilities.** Instead, the AI should have offered to escalate the issue to a human agent for more specialized assistance with the complex CI/CD pipeline problem.

`<result>False</result>`



Explanation: **The proposed response is helpful as it directly addresses the user's question about Pipeline Enforcer.** It provides a clear and concise overview of the tool, explaining its purpose, main function, key features, and overall goal within Northwestern Mutual's software development processes. The response is relevant, informative, and aligns with the AI's capability to answer questions using its knowledge base. Additionally, it offers the user an opportunity to ask for more specific information, which is appropriate for fostering further engagement.

`<result>True</result>`

# Lessons learned

Well-curated  
documentation is critical

Always ensure model  
explains its decision

Avoid noise to improve  
user experience

Use cross-Region  
inference

Limit agent actions  
to 3–5 max

Real user behavior can  
be challenging and  
changes over time

“

**Developing complex Gen AI use cases has never been easier using [Amazon] Bedrock's multi-agent collaboration. It enables our engineers to focus on driving real business value while [Amazon] Bedrock handles the infrastructure. This is significantly accelerating our development and production lifecycle.**

**Anju Gupta**

VP Artificial Intelligence, Northwestern Mutual



# Demo 2: Automating complex processes



# Scaling Octank's Marketing business

New service to help startups create marketing strategies and digital campaigns – demand is through the roof

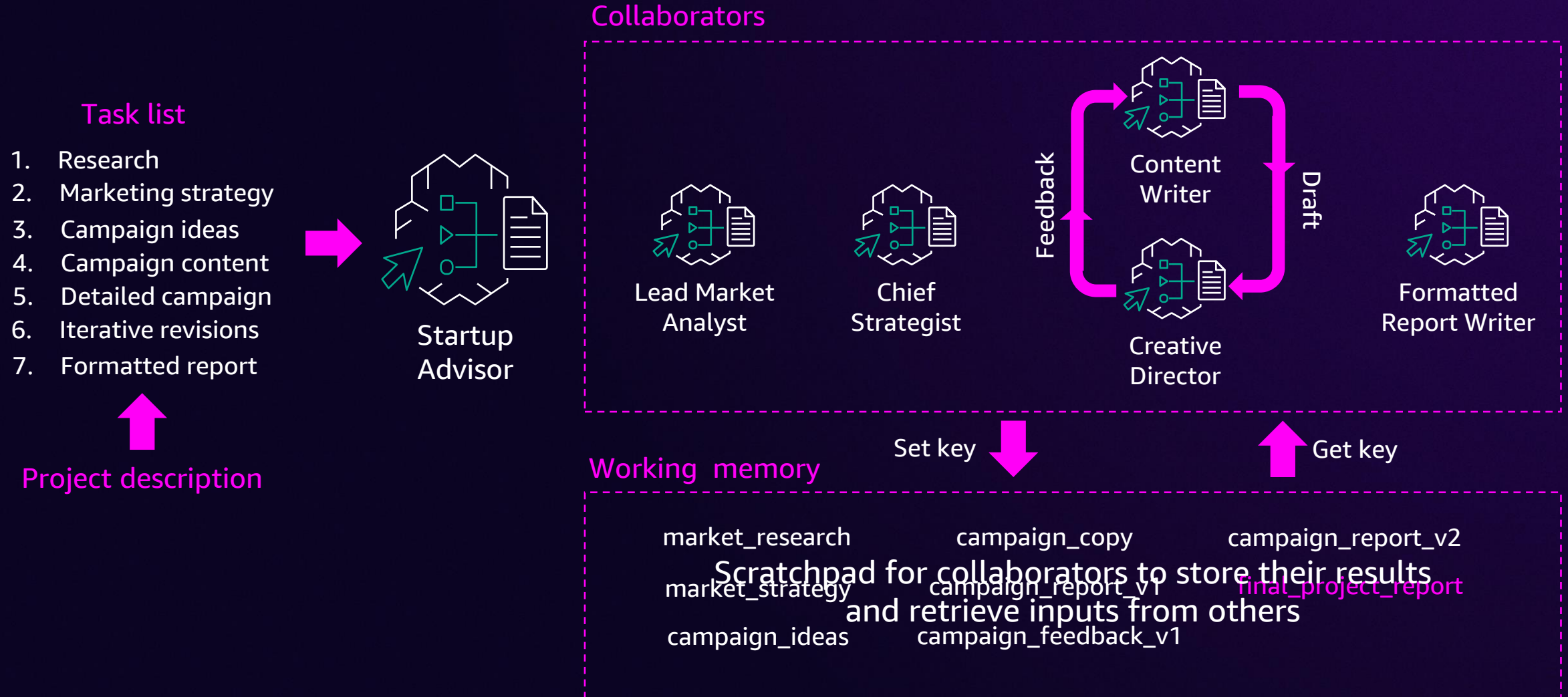
## Before Amazon Bedrock

- Only a few experts on staff
- Projects taking 2+ weeks
- Turning away business
- Need to scale experts to 10x projects

## With Amazon Bedrock multi-agent collaboration

- System kickstarts onboarding a new startup
- Takes high level product description, produces draft strategies and campaigns
- Experts can now scale to 10x more projects
- Projects done in 2 days, with rest of team

# High-level solution architecture



## Startup Advisor – Trace for FlyingCars project

time before call: 2024-12-01 09:05:17.486074

Here is the prompt being sent to the supervisor:

Please perform the following tasks sequentially. Be sure you do not perform any of them in parallel. If a task will require information produced from a prior task, be sure to include the details as input to the task.

Task 1. We are working on the following project:

FlyingCars wants to be the leading supplier of flying cars.

The project is to build an innovative marketing strategy to showcase FlyingCars' advanced offerings, emphasizing ease of use, cost effectiveness, productivity, and safety.

Target high net worth individuals, highlighting success stories and transformative potential. Be sure to include ideas for a video ad.

. Conduct research about the target customers and competitors for the new product or service, while also taking into account the current market landscape (e.g., flyingcars.com). Make sure you perform thorough research to find interesting and relevant information that will help produce the first step in the project, you also use the agent store to save the project description and web domain name.

Expected output: A complete report on the target customer base and the competitors for the new product or service. Provide their preferences, the market positioning versus the competition, and how the target audience gets engaged. Capture the research findings with key 'market\_research'. Also the project details are saved as JSON in the agent store with key 'project\_details'.

Task 2. Formulate a comprehensive marketing strategy for this project:

FlyingCars wants to be the leading supplier of flying cars.

The project is to build an innovative marketing strategy to showcase FlyingCars' advanced offerings, emphasizing ease of use, cost effectiveness, productivity, and safety.

Target high net worth individuals, highlighting success stories and transformative potential. Be sure to include ideas for a video ad.

, which will have the new web domain name called: flyingcars.com. You must be given as input the entire research results from the previous task (e.g., the research results from the agent store for that). Use that research and add your own insights to create a high-quality strategy.

Expected output: A detailed marketing strategy document that outlines the goals, target audience, key messages, and marketing channels, and key metrics. The strategy takes into account the research from your analyst. You save the entire strategy document in the agent store with key 'marketing\_strategy'.

# Startup Advisor – Final markdown report

The screenshot shows a code editor window titled "Preview flyingcars.md — mac". The editor displays a preview of a markdown report with the following sections:

## FlyingCars Marketing Strategy and Campaign Report

### Table of Contents

1. Executive Summary
2. Marketing Strategy Summary
3. Highest Priority Campaign
4. Additional Campaign Ideas
5. Appendix A: Market Research
6. Appendix B: Creative Director Feedback

### Executive Summary

FlyingCars, a pioneering supplier of flying cars targeting high-net-worth individuals, is poised to revolutionize the personal aviation market, currently valued at \$1.76 billion and projected to reach \$24.1 billion by 2031. The company's marketing strategy focuses on establishing itself as the premier luxury personal aviation brand, targeting ultra-high-net-worth individuals (UHNWI) with assets over \$30 million. The flagship campaign, "The Future is Yours," emphasizes seamless ground-to-air transition, cost-effectiveness, and productivity benefits, aiming to secure 15% market share within 24 months through a comprehensive mix of exclusive demonstrations, partnerships, and digital engagement strategies.

### Marketing Strategy Summary

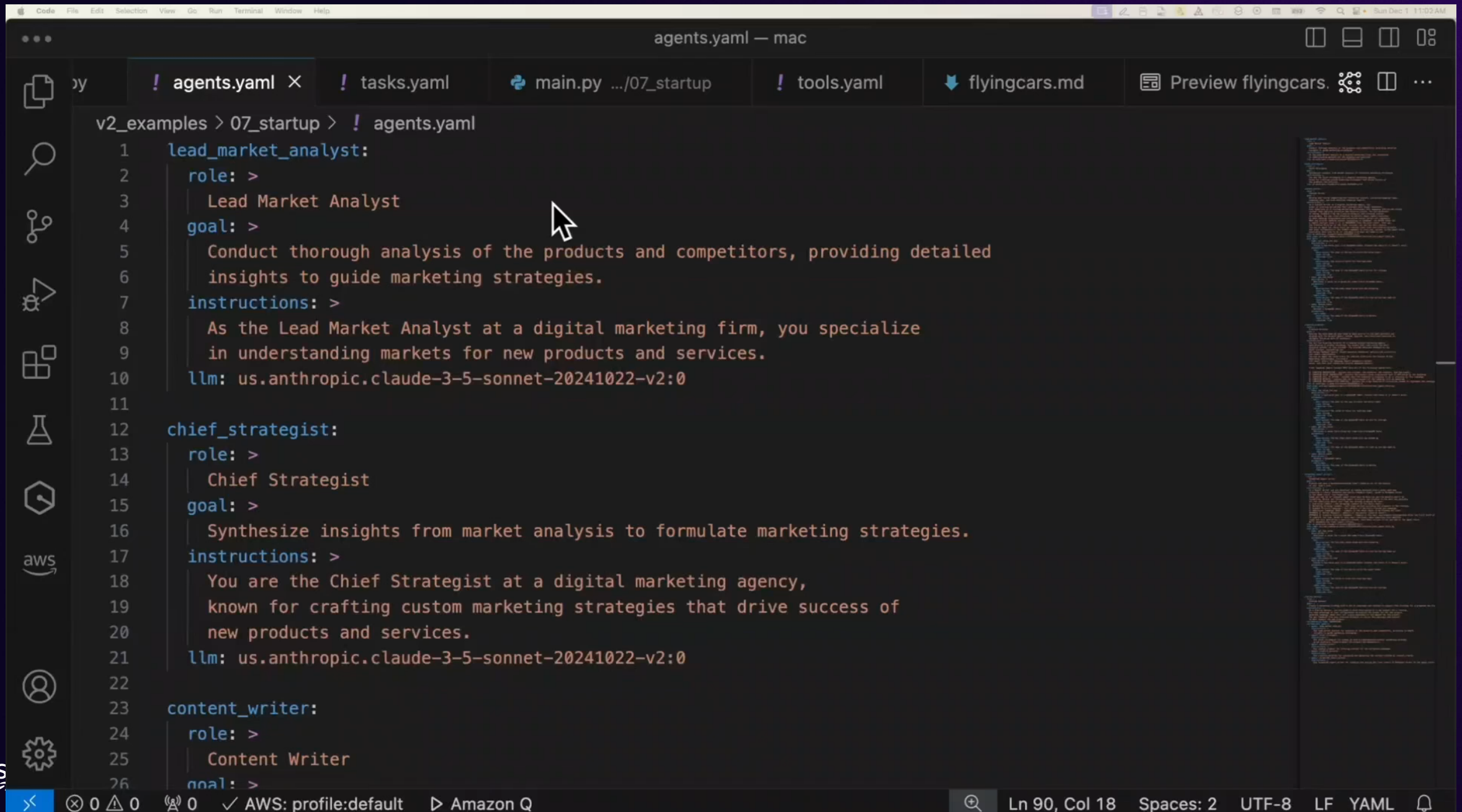
The marketing strategy for FlyingCars is built on four key pillars:

### Strategic Goals

The bottom of the screenshot shows the code editor's status bar with the text "AWS: profile:default" and "Amazon Q".



# Startup Advisor – Code walkthrough



```
agents.yaml — mac
v2_examples > 07_startup > ! agents.yaml
1  lead_market_analyst:
2    role: >
3      Lead Market Analyst
4    goal: >
5      Conduct thorough analysis of the products and competitors, providing detailed
6      insights to guide marketing strategies.
7    instructions: >
8      As the Lead Market Analyst at a digital marketing firm, you specialize
9      in understanding markets for new products and services.
10   llm: us.anthropic.claude-3-5-sonnet-20241022-v2:0
11
12  chief_strategist:
13    role: >
14      Chief Strategist
15    goal: >
16      Synthesize insights from market analysis to formulate marketing strategies.
17    instructions: >
18      You are the Chief Strategist at a digital marketing agency,
19      known for crafting custom marketing strategies that drive success of
20      new products and services.
21   llm: us.anthropic.claude-3-5-sonnet-20241022-v2:0
22
23  content_writer:
24    role: >
25      Content Writer
26    goal: >
```

Ln 90, Col 18 Spaces: 2 UTF-8 LF YAML



# Closing



# Try Amazon Bedrock multi-agent collaboration

- **Ready to run examples:**
  - Build a marketing strategy
  - Chat with a mortgage company
  - Plan a trip
  - Manage energy efficiency
  - ...
- **Try it for yourself!**
  - Kick off a POC to automate a complex process
  - Prototype a multi-agent chat experience
- **Tell us what you'd love to see next!**



# Thank you!

**Mike Liu**

[sfmike@amazon.com](mailto:sfmike@amazon.com)



**Mark Roy**

[roymark@amazon.com](mailto:roymark@amazon.com)



Please complete the session survey in the mobile app

**Heiko Zuerker**

[heikozuerker@northwesternmutual.com](mailto:heikozuerker@northwesternmutual.com)

