



Leveling Up Agents: Copilot Studio for Enterprise Solutions

Kristen Holmes



About me



Kristen Holmes
Associate at **Morgan Stanley**
Application Infrastructure — Low Code Fleet



kristen.holmes@morganstanley.com

Agenda

Which Copilot?

Key features of Copilot Studio

Building an Agent

Topics, Agent Flows, levels of autonomy

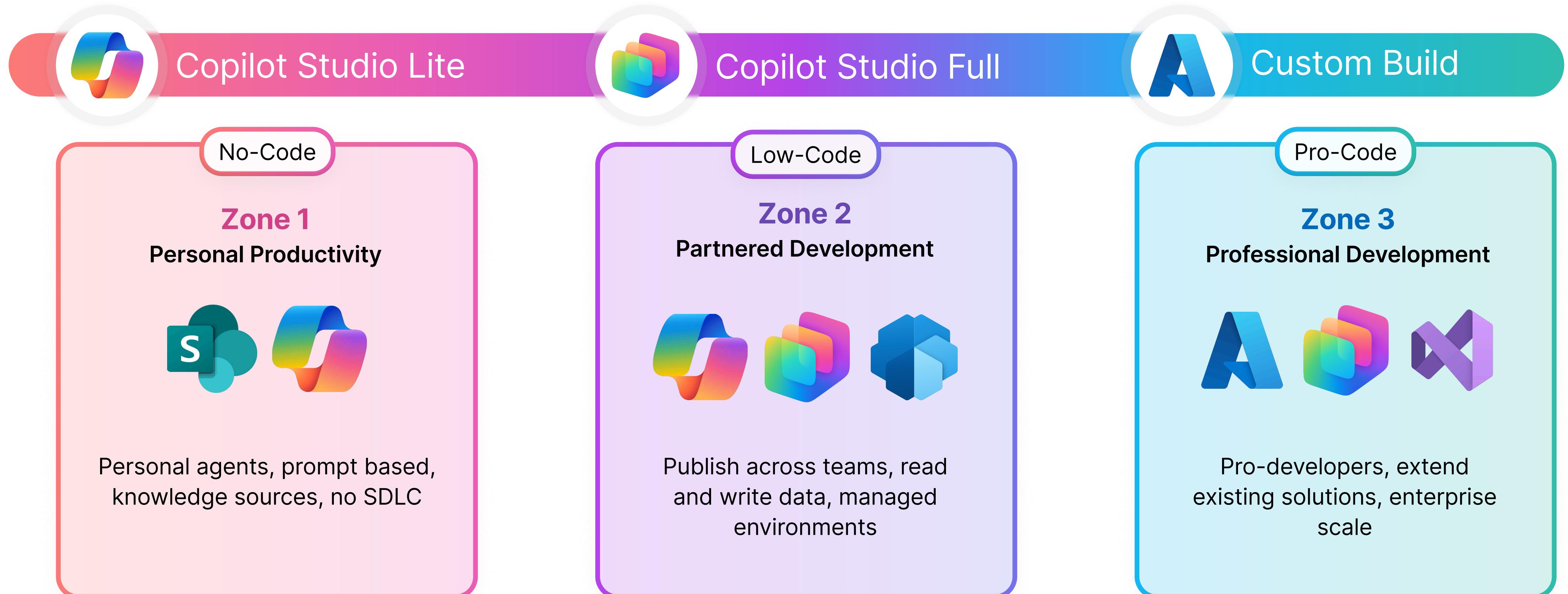
Multi-Agent

Multi-Agent Patterns, Orchestration

Evaluation

Tools for Evaluation, Agent as a Judge

Which Copilot?



Balancing Determinism & Dynamism

Dynamic

- Adaptive reasoning
- Context-aware behavior
- Ambiguity and open-ended requests
- Flexibility

Context > Certainty

Agent Flows

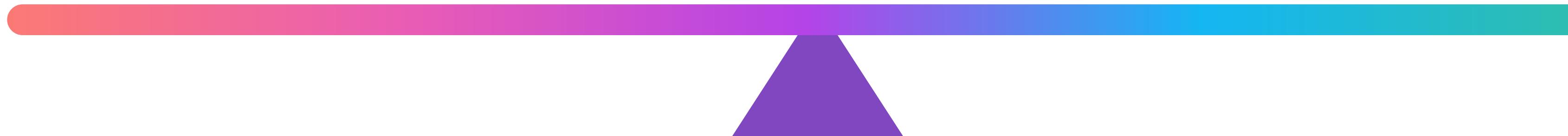
Topics

Orchestration

Deterministic

- Rules based
- Predictable outcomes
- Repeatable
- Exact data
- High impact actions

Accuracy > Adaptability



Copilot as an Orchestrator

Routing

Route between agents, tools, and topics dynamically

Tools

Instruct agents on tool suitability and guidelines

Intent

Utilize context from users to better identify requests

Context

Instructions to provide context on tasks, requirements, and capabilities



Prompt Engineering

- Default way to create agents
- More ambiguous
- Increased variability
- **Tightly coupled to a model**
- Model drift

Details Edit

 Name: Travel Agent 109/1024

Description: AI agent assisting employees with flight/hotel requests and expense submissions via conversational interface.

Orchestration: Use generative AI to determine how best to respond to users and events. [Learn more](#) Enabled

Agent's model: GPT-5 Auto (preview) ...

Instructions Edit

- Assist employees in booking flights and hotels for business travel through a conversational interface.
- Guide users through submitting travel expenses for reimbursement.
- Reference and interact with tables: Flight Booking, Hotel Booking, Expense Submission, Available Flights, Available Hotels, Employee.
- Respond to user stories such as requesting flight/hotel bookings and submitting travel expenses.
- Provide clear, step-by-step guidance for each travel-related request.
- Ensure responses are accurate, timely, and relevant to employee travel needs.
- Maintain a professional and helpful tone in all interactions.
- Do not process requests unrelated to travel booking or expense submission.
- Avoid sharing personal or sensitive information unless required for booking or expense purposes.
- Follow company travel policies and guidelines when assisting users.
- Do not answer questions outside the scope of travel booking and expense submission.

Context Engineering

Instructions

 Edit

IF USER HAS BOOKED TRAVEL:

- Extract Travel information  Extract Booking Information

IF USER HAS **NOT** BOOKED TRAVEL:

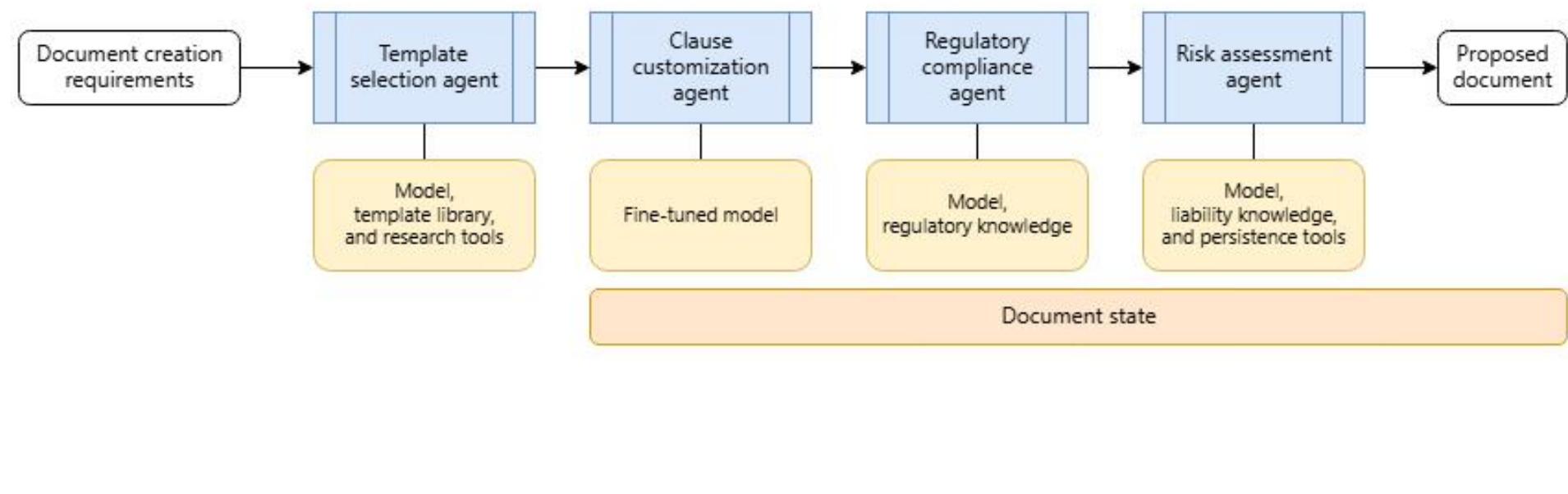
- Returns available flights  Get Available Flights
- Book flights using  Book a Flight
- Book hotels using  Book a Hotel

- Suitability of tools
- Increased consistency
- Looser coupling from model
- Readability

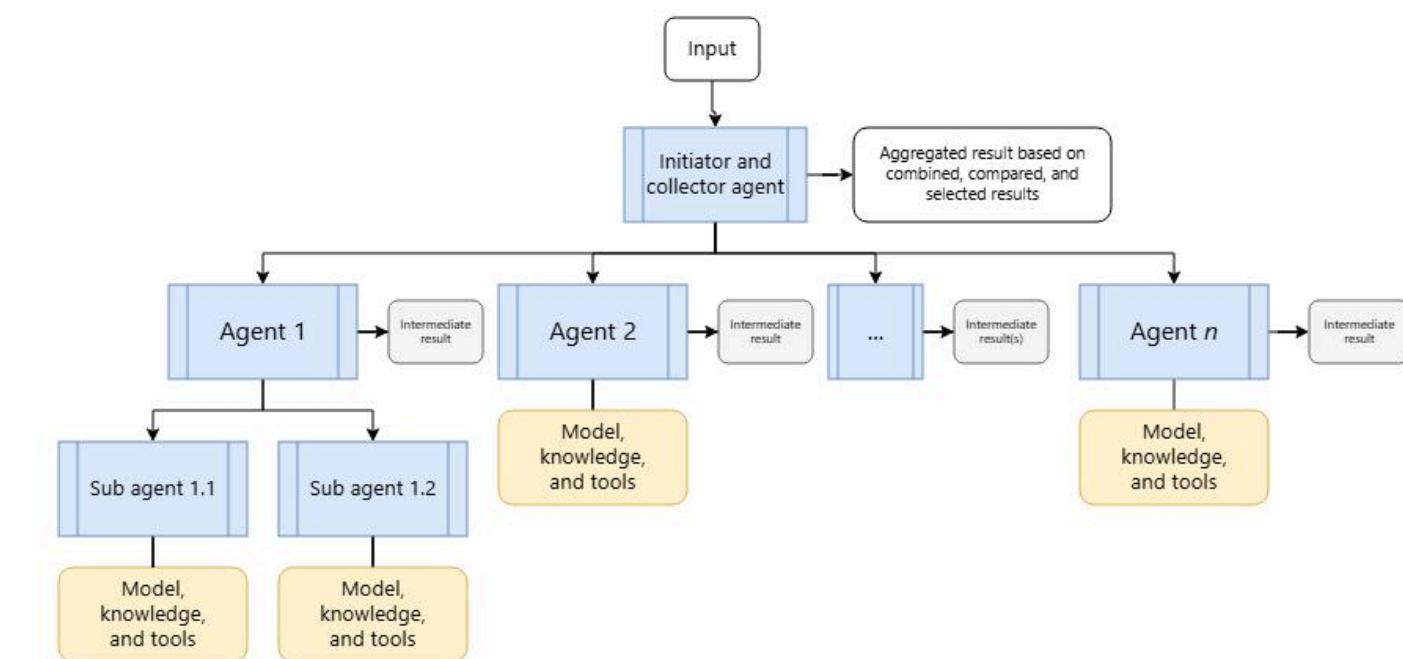


Multi-Agent Orchestration

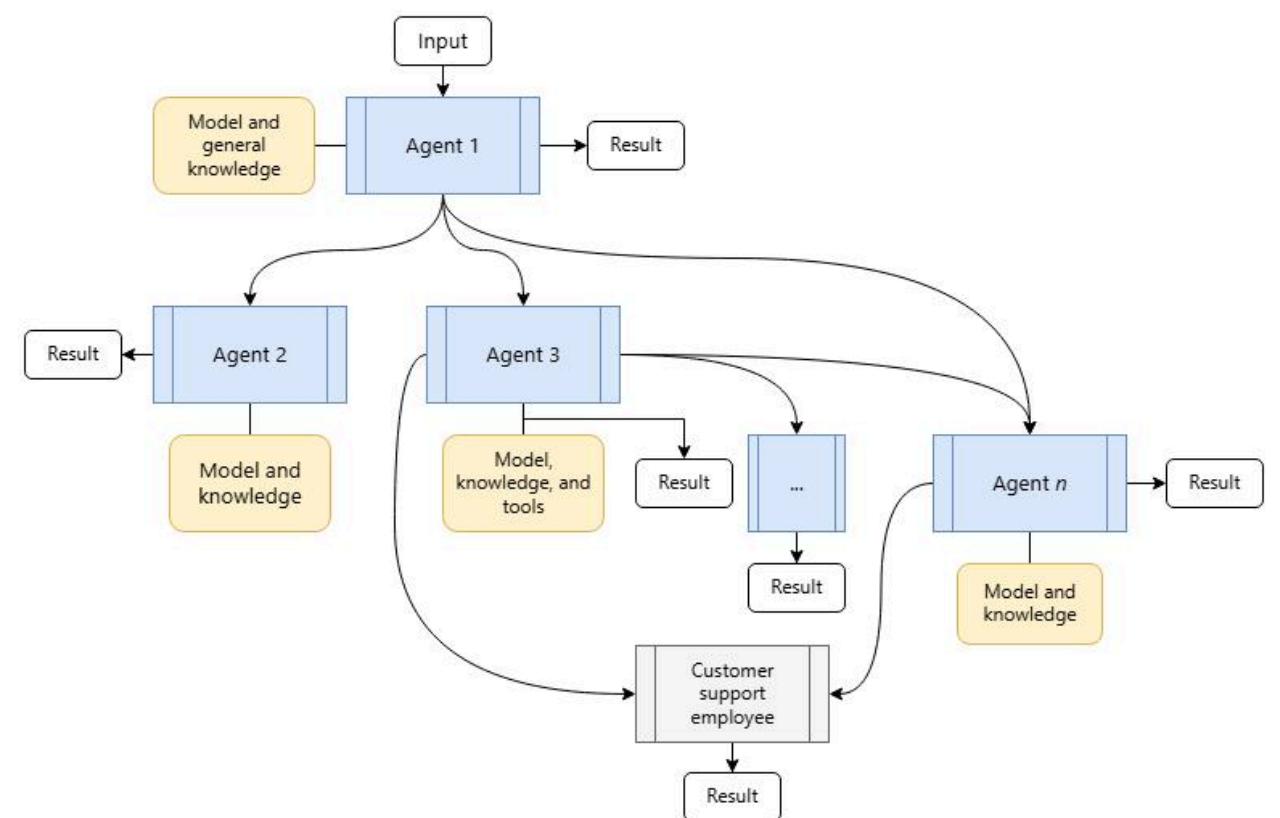
Sequential



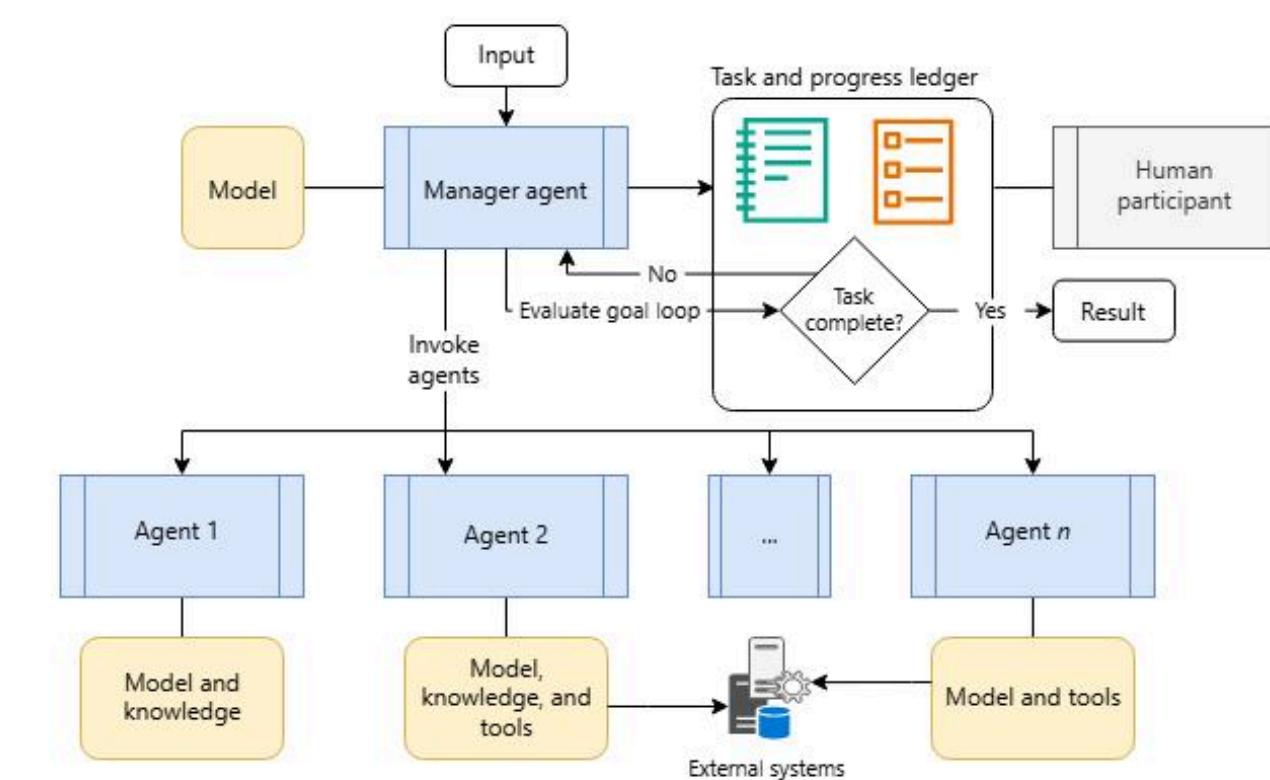
Concurrent



Handoff



Managerial



Levels of Autonomy

L1

User as an
Operator

User directs and
makes decisions,
agent acts.

L2

User as an
Collaborator

User and agent
collaboratively
plan, delegate,
and execute.

L3

User as an
Consultant

Agent takes lead
but consults user
for expertise/
preferences.

L4

User as an
Approver

Agent engages
user only in risky
or prespecified
scenarios.

L5

User as an
Observer

Agent operates
with full
autonomy under
user monitoring.

Evaluation

The screenshot shows the Copilot Studio interface for evaluation. On the left, there's a sidebar with icons for Home, Create, Agents, Flows, Tools, and The main navigation bar includes Helpdesk agent (green checkmark), Overview, Knowledge, Tools, Agents, Topics, Activity, Evaluation (selected), Analytics, and Channels. The Evaluation tab shows a sub-menu with Test results, Weekly full agent eval - 11062025.3, and a link to Evaluations (preview).

Evaluations (preview)
Track conversation outcomes to see how well your agent is helping your users. [See preview terms](#)

Test sets

- Evaluate Website Q&A (10 test cases • last modified by Steph Kent, 16 days ago) [Active](#)
- Evaluate Website Q&A (10 test cases • last modified by Steph Kent, 17 days ago) [Active](#)

Evaluation summary

Score	94%
Pass	94%
Fail	6%

Duration: 00:15:32 | Test cases: 32 | Test set: Weekly full agent eval | User profile: daisy.phillips@contoso.com | Tested by: Mona Kane 2 days ago

New test set

Choose the Right Method

- **Automated Testing**, efficient, scalability, baseline
- **Human-in-the-loop**, subjective, high-impact tasks
- **LLM-as-a-judge**, reasoning chains, tool use

Define Metrics

- **Deterministic**, with clear, expected outcomes
- **Comprehensive Data**, representative, balanced
- **Open-ended prompts**, evaluate generative capabilities
- **Load testing**, evaluate performance



“The most effective AI implementations strategically position themselves on this spectrum based on the requirements of their individual use case rather than following industry hype”

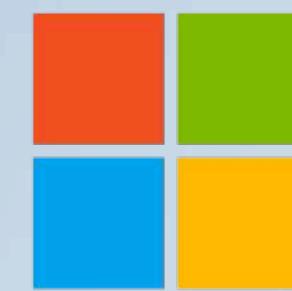
— Deepset, *AI Agents and Deterministic Workflows: A Spectrum, Not a Binary Choice*



References

- Anthropic. (n.d.). Building effective AI agents. Building Effective AI Agents Anthropic. <https://www.anthropic.com/engineering/building-effective-agents>
- Deepset. (n.d.). Ai agents and deterministic workflows: A spectrum, not a binary choice. deepset Blog. <https://www.deepset.ai/blog/ai-agents-and-deterministic-workflows-a-spectrum>
- Garypretty. (n.d.). Orchestrate agent behavior with Generative AI - Microsoft Copilot studio. Microsoft Copilot Studio | Microsoft Learn. <https://learn.microsoft.com/en-us/microsoft-copilot-studio/advanced-generative-actions>
- Gilboa, E. (2025, November 8). Build smarter, test smarter: Agent evaluation in Microsoft copilot studio. Microsoft Copilot Blog. <https://www.microsoft.com/en-us/microsoft-copilot/blog/copilot-studio/build-smarter-test-smarter-agent-evaluation-in-microsoft-copilot-studio/>
- GU, J. (n.d.). A survey on LLM-as-a-judge. <https://arxiv.org/pdf/2411.15594>
- McDonald , D., Feng, K., & Zhang, A. (2025, July 28). Levels of autonomy for AI agents. Knight First Amendment Institute. <https://knightcolumbia.org/content/levels-of-autonomy-for-ai-agents-1>
- Microsoft . (2025b, May 19). Introducing agent flows: Transforming automation with AI-first workflows. Microsoft Copilot Blog. <https://www.microsoft.com/en-us/microsoft-copilot/blog/copilot-studio/introducing-agent-flows-transforming-automation-with-ai-first-workflows/>
- Microsoft . (n.d.). AI agent Orchestration Patterns - Azure Architecture Center. AI Agent Orchestration Patterns - Azure Architecture Center | Microsoft Learn. <https://learn.microsoft.com/en-us/azure/architecture/ai-ml/guide/ai-agent-design-patterns>
- Microsoft Build smarter, test smarter: Agent evaluation in Microsoft copilot studio. Microsoft Copilot Blog. (2025a, November 8). <https://www.microsoft.com/en-us/microsoft-copilot/blog/copilot-studio/build-smarter-test-smarter-agent-evaluation-in-microsoft-copilot-studio/>

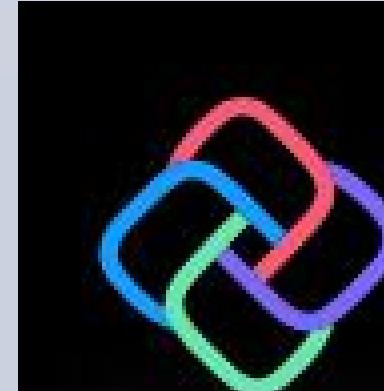
Thank You to our sponsors



Microsoft



INFRAGISTICS®



uno®
PLATFORM

Morgan Stanley

apidays

And next time -
Your Company!

Please share your event and session feedback!



<http://aka.ms/MSIgniteNYCSurvey>



<http://aka.ms/MSIgniteNYCSessionsSurvey>