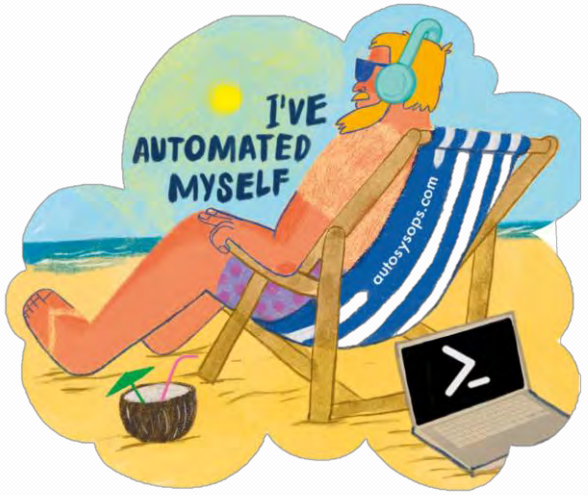
Clean and Controlled AI

with API Management

Leo Visser



Who am I

AUTO-SYS-OPS

- Consultant @ OGD
  - AI, Automation, Cloud
- DuPSUG Organizer
- Azure MVP (PowerShell)

Blog: www.autosysops.com
Socials: @AutoSysOps

Meet: nature reserves management group

Case:

Use an AI chatbot for employees to interact with different systems.

Meet: nature reserves management group

Requirements:

- Use AI to make work easier

Meet: nature reserves management group

Requirements:

- Use AI to make work easier
- Don’t overuse the AI systems

Meet: nature reserves management group

Requirements:

- Use AI to make work easier
- Don’t overuse the AI systems
- Prevent certain topics

Meet: nature reserves management group

Requirements:

- Use AI to make work easier
- Don’t overuse the AI systems
- Prevent certain topics
- Make the AI more sustainable

Show of hands

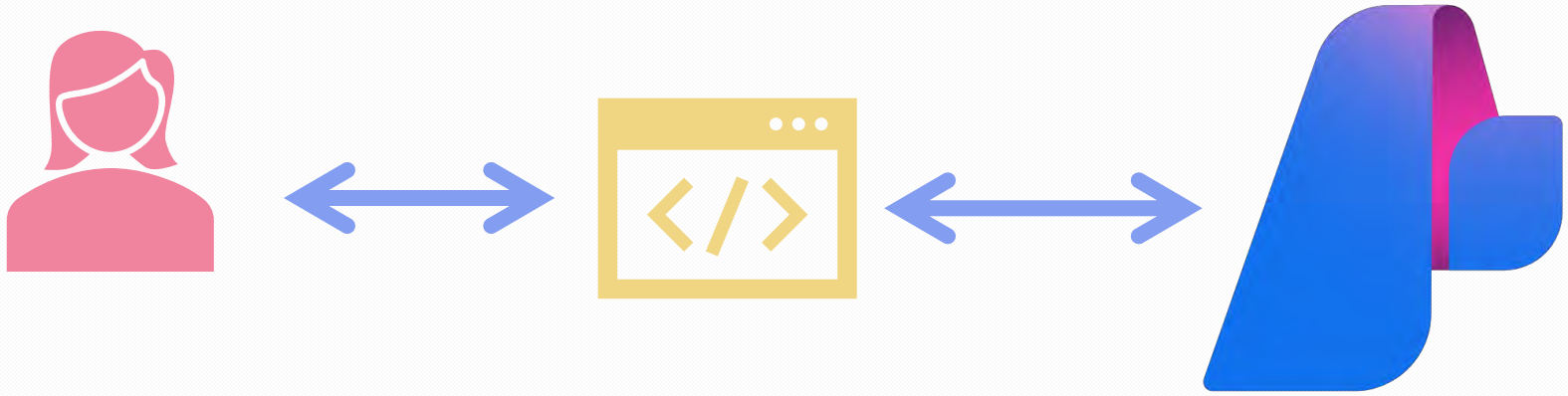
Who here has had questions about limiting the AI topics?

Show of hands

Who here has had questions about making AI more sustainable?

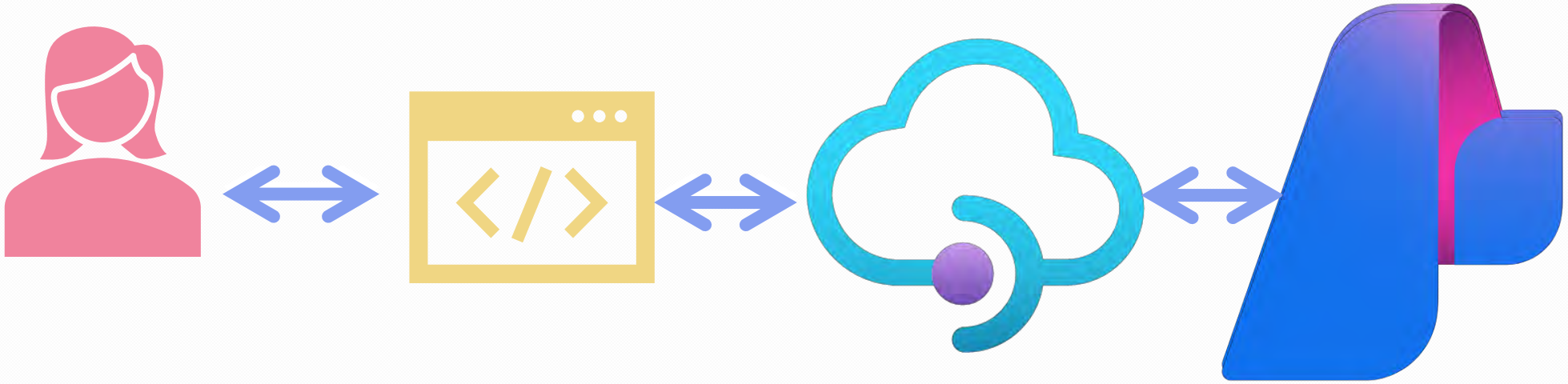
API Management + AI Endpoints

Api Management + AI Endpoints



AI Foundry allows for many models and ways to interact for an app.

Api Management + AI Endpoints



API Management allows for extra controls in between

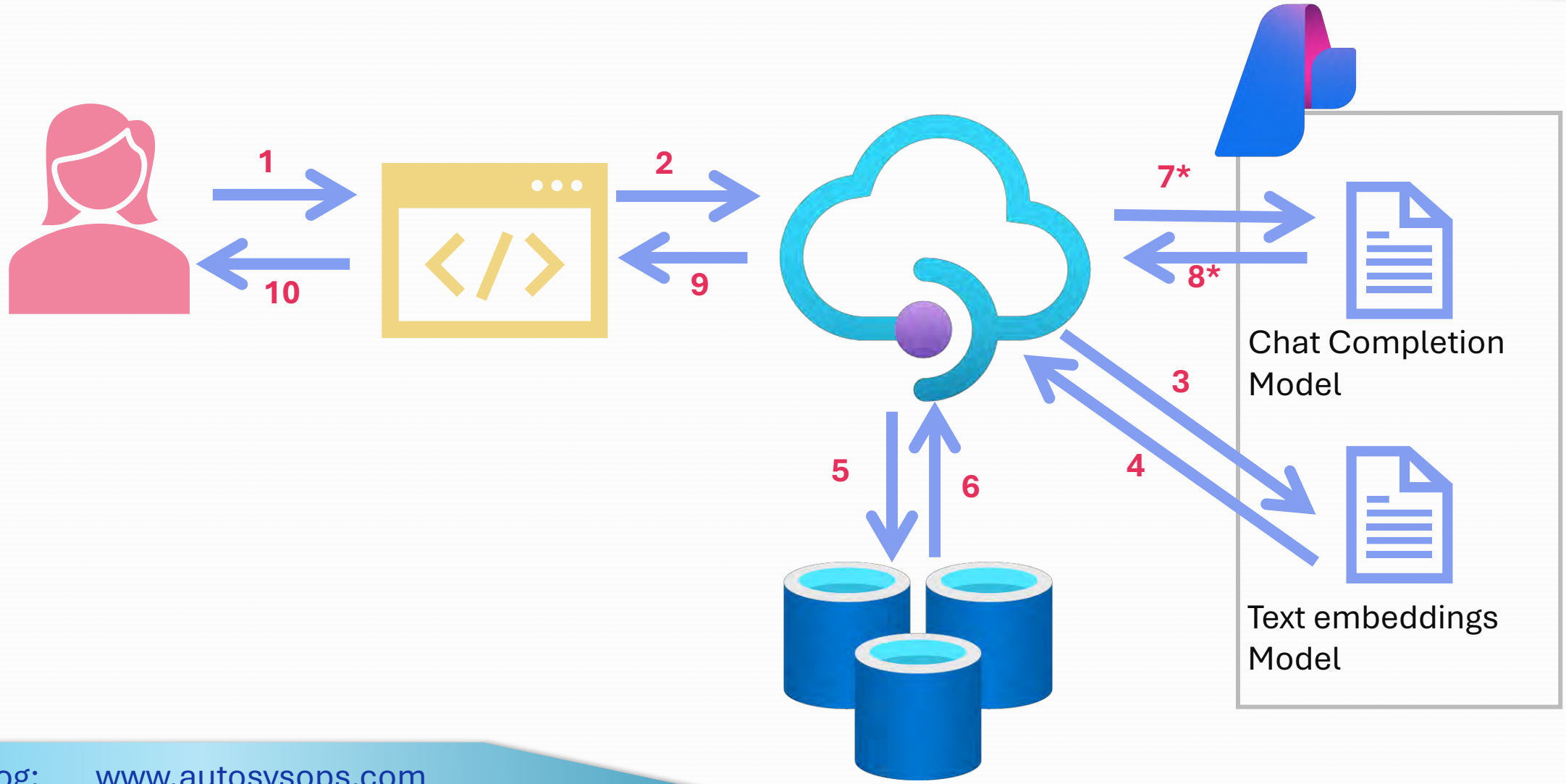
Demo connecting

Rate limiting

Content Safety

Caching

Api Management + AI Endpoints



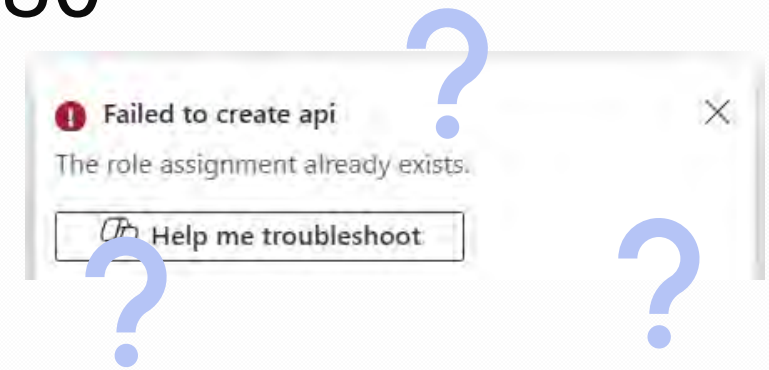
But money!

Money?

- API Management basic v2 is ~150 EUR per month

Money?

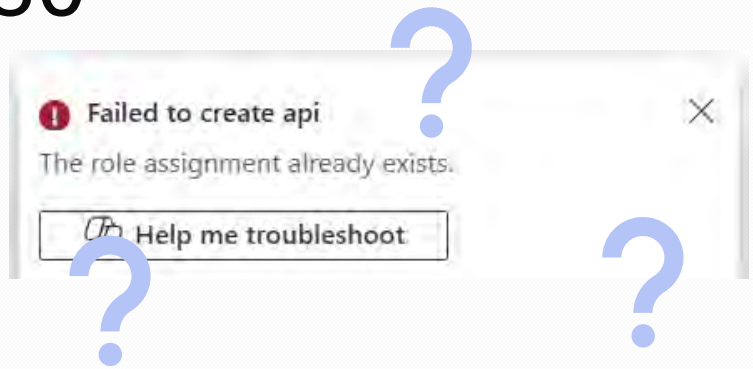
- API Management basic v2 is ~150 EUR per month



AUTO-SYS-OPS

Money?

- API Management basic v2 is ~150 EUR per month
- Also supports MCP server (only tools at the moment)
- Allows for things like authentication checks too
- Can be managed centrally



AUTO-SYS-OPS

What is next?

What is next?

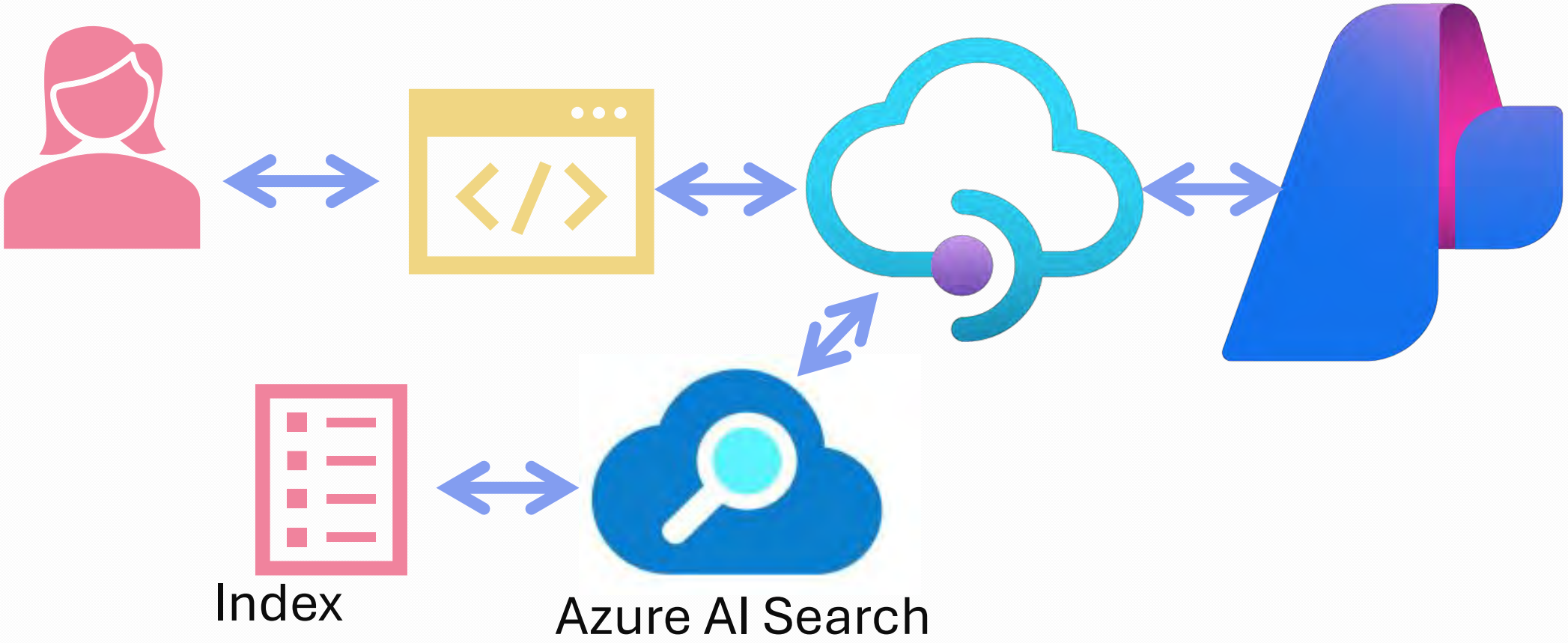
Use AI Search + APIM to:

- Detect which finetuned model needs to be selected
- Detect which model is best for the type of query to limit CPU usage (could also consider using model router in foundry)

Api Management + AI Endpoints

AUTO-SYS-OPS

Use AI Search + APIM to:



AUTO-SYS-OPS

Questions?

Blog: www.autosysops.com
Socials: @AutoSysOps