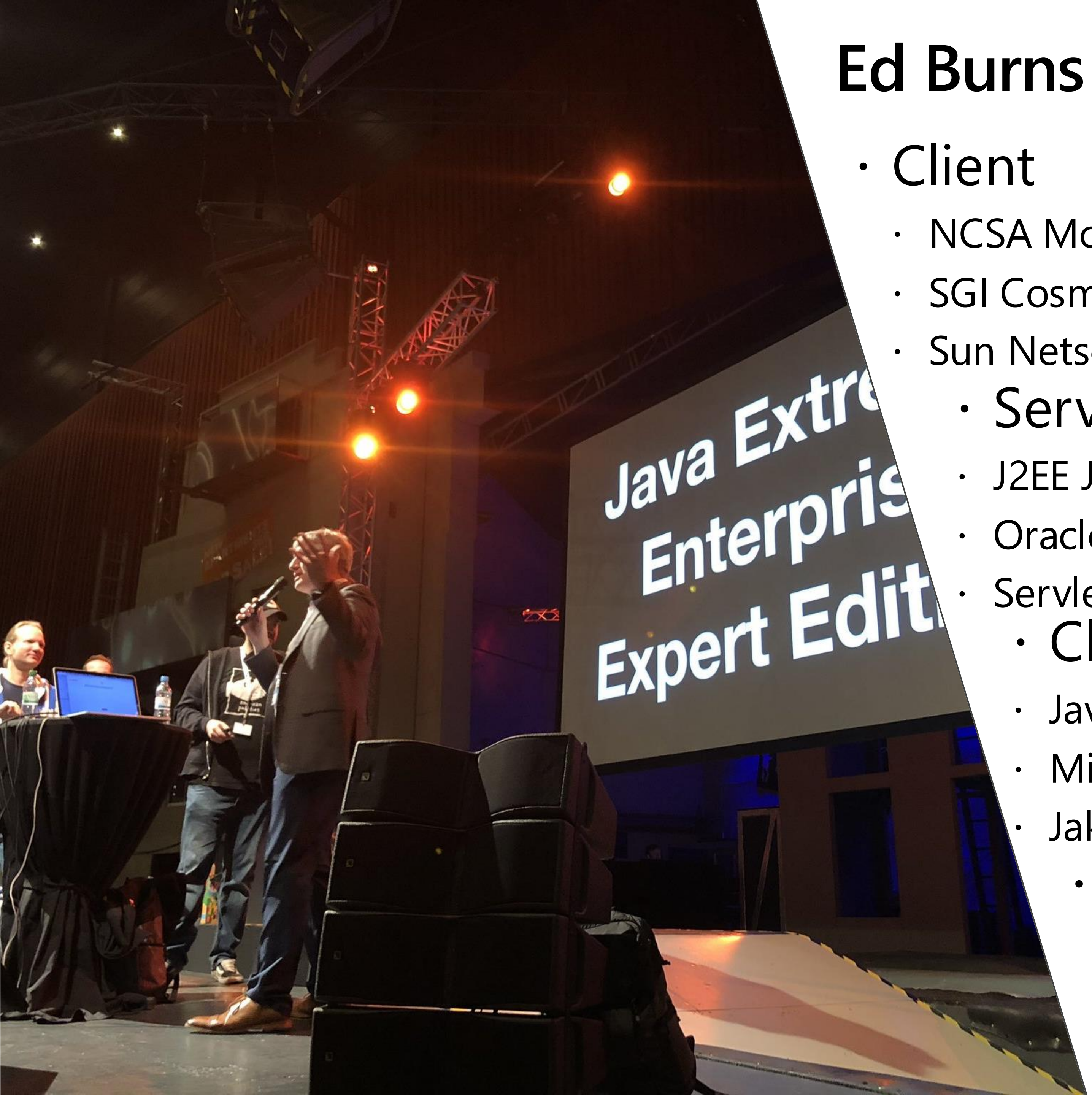
Microsoft

Java and AI with LangChain4j

Code the future with AI

Ed Burns @edburns.bsky.social
Principal Architect

https://aka.ms/edburns/jugdarmstadt2025



Ed Burns professional biography

- Client
  - NCSA Mosaic (1994)
  - SGI Cosmo Web Authoring
  - Sun Netscape 6 OJI
- Server
  - J2EE JSF (2002)
  - Oracle Java EE
  - Servlet, JSF, Bean Validation, etc.
- Cloud
  - Java Champion
  - Microsoft Azure Cloud (2019)
  - Jakarta EE 11 committer and release coordinator
- Books

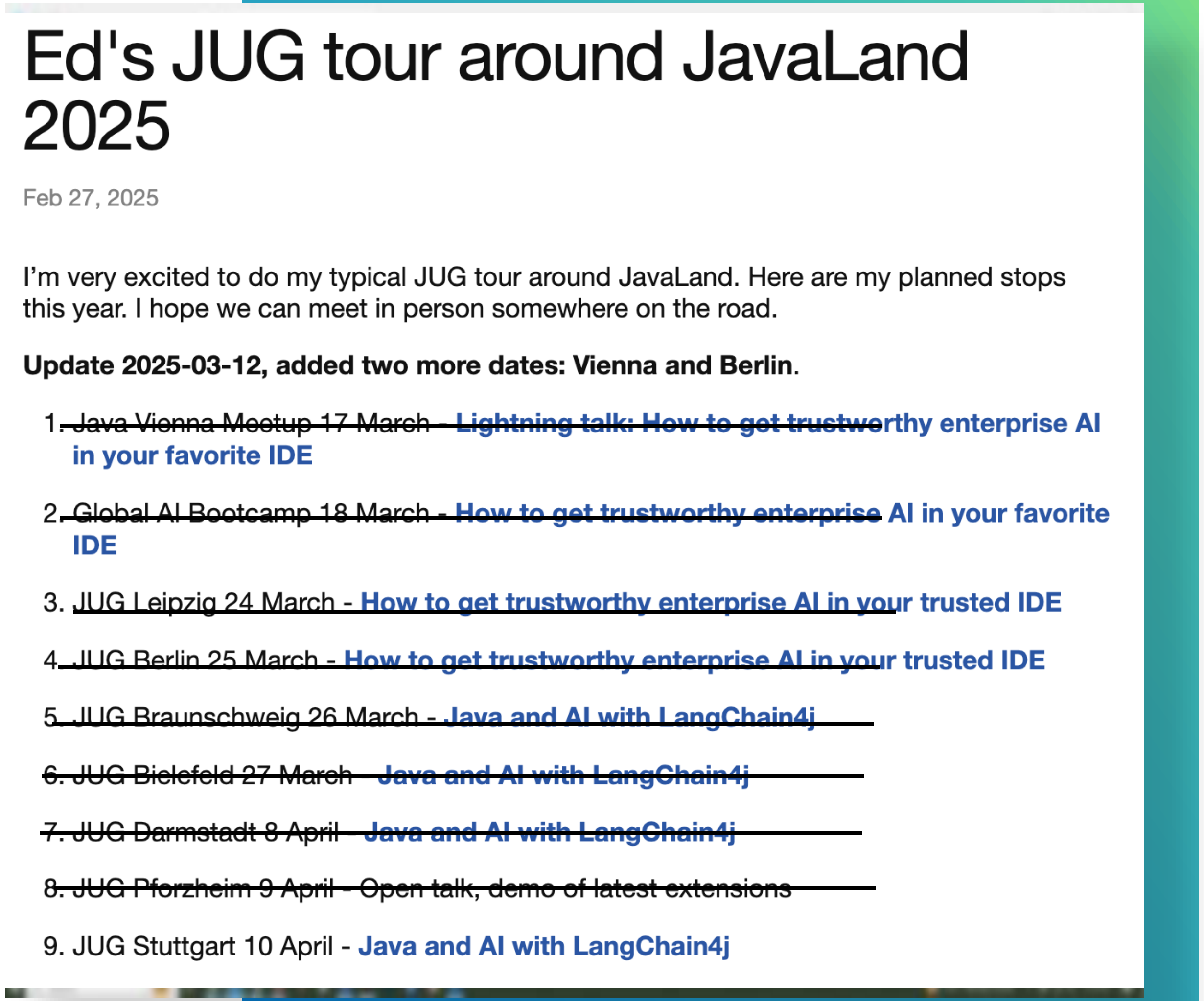
The Complete Reference
JavaServer Faces
Secrets of the Rockstar Programmers
Ed Burns
The Complete Reference
JavaServer Faces 2.0
Oracle
Hudson Continuous Integration in Practice
packt
Developer Career Masterplan
Heather VanCura | Bruno Souza

Change in content from posted agenda

- Still 100% Java and AI with LangChain4j.
- Removed content:
  - After a brief look at the landscape.
  - How you use it to perform the most popular AI usage pattern: Retrieval Augmented Generation (RAG).
- Added content:
  - langchain4j-microprofile-jakarta.
- If you want the RAG content, I recommend Easy RAG with LangChain4J and Docker by Julien Dubois.

JUG Tour Around JavaLand 2025

https://aka.ms/edburns/2025JavaLandJUGTour



Ed's JUG tour around JavaLand 2025

Feb 27, 2025

I'm very excited to do my typical JUG tour around JavaLand. Here are my planned stops this year. I hope we can meet in person somewhere on the road.

Update 2025-03-12, added two more dates: Vienna and Berlin.

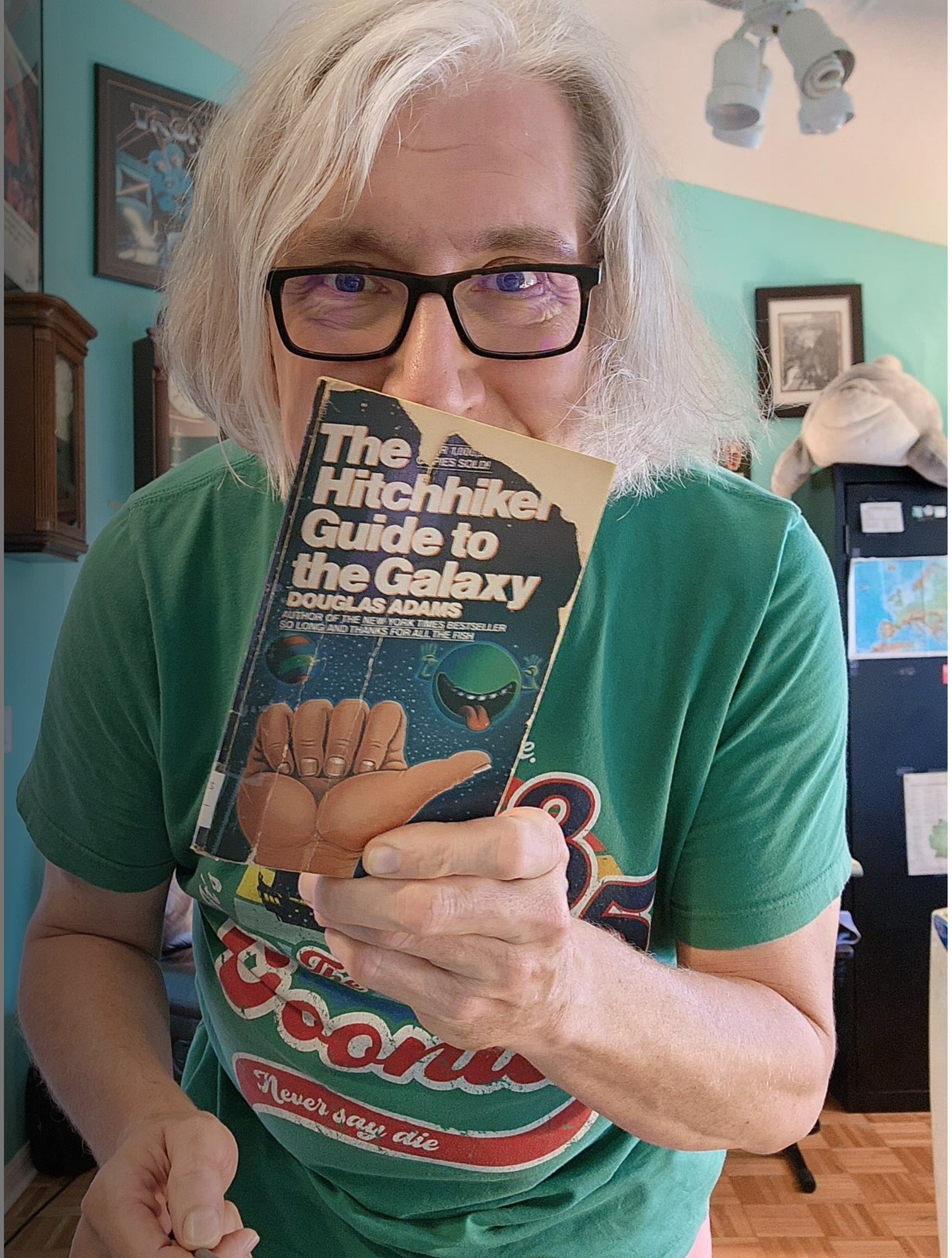
1. ~~Java Vienna Meetup 17 March - **Lightning talk: How to get trustworthy**~~ **enterprise AI in your favorite IDE**
2. ~~Global AI Bootcamp 18 March - **How to get trustworthy enterprise**~~ **AI in your favorite IDE**
3. ~~JUG Leipzig 24 March - **How to get trustworthy enterprise AI in your**~~ **trusted IDE**
4. ~~JUG Berlin 25 March - **How to get trustworthy enterprise AI in your**~~ **trusted IDE**
5. ~~JUG Braunschweig 26 March - **Java and AI with LangChain4j**~~
6. ~~JUG Bielefeld 27 March - **Java and AI with LangChain4j**~~
7. ~~JUG Darmstadt 8 April - **Java and AI with LangChain4j**~~
8. ~~JUG Pforzheim 9 April - Open talk, demo of latest extensions~~
9. JUG Stuttgart 10 April - **Java and AI with LangChain4j**

Breaking Jakarta EE news

- Jakarta EE 11 Web Profile released this past Friday.
  - https://github.com/jakartaee/specifications/pull/802
  - https://deploy-preview-802--jakartaee-specifications.netlify.app/specifications/webprofile/11/
- Many EE 12 component specifications in Plan Review stage.
- MicroProfile is discussing whether and how to host MicroProfile in Jakarta EE.
  - The need for a unified Config solution for EE 12.
  - Coincidentally timed soon after Evolving our middleware strategy blog post from Red Hat.

New to AI?

Don’t panic!


The Hitchhike
Guide to the Galaxy
DOUGLAS ADAMS
AUTHOR OF THE NEW YORK TIMES BESTSELLER
SO LONG AND THANKS FOR ALL THE FISH
Never say die

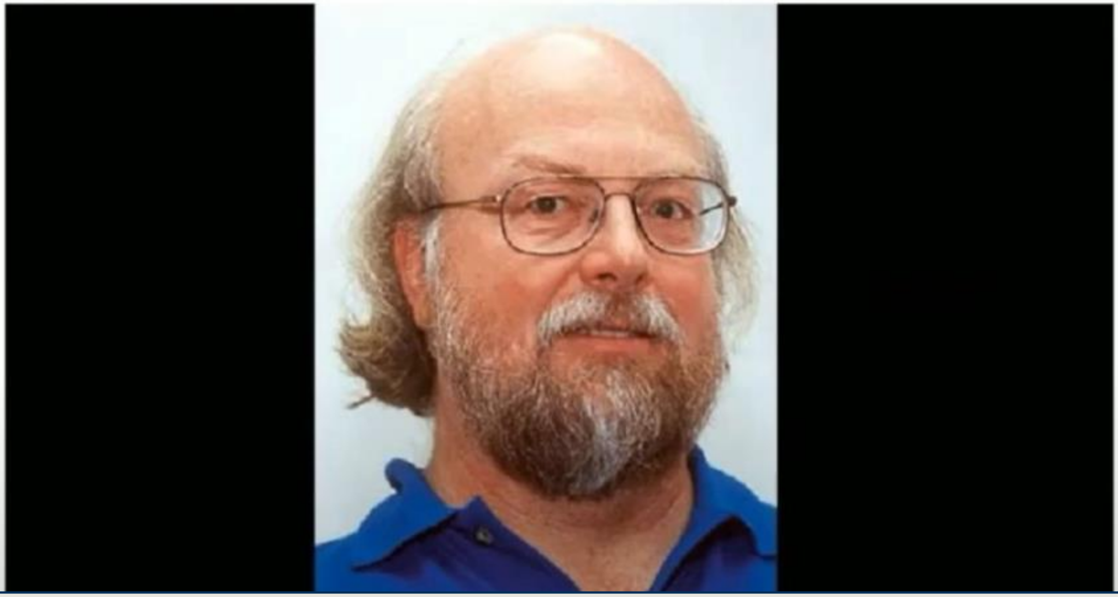
Ed’s fairly obvious observations

- Recognize the market forces that have always been in play.
- Vendors want adoption.
- They’ll make it easy to adopt.
- This is also true with AI.

...but there's more to life than making
shallow, fairly obvious observations.

```
publ   ass    ViewRoot implements Serializable { public Ajax  wRoot(PartialTraversalViewRoot viewRoot) {if (viewRoot == null)
 th     galArgumentException("PartialTraversalViewRoot "  "cannot be null!");}setUIViewRoot(viewRoot);
         onents = new ArrayList<UIComponent>();    markImmediate = new Util.TreeTraversalCallback() { public boolean
take      de(FacesContext context, UIComponent comp) throws FacesException { if (comp instanceof ActionSource) {
Act                                                 (true); modifiedComponents.add(comp); } } else if
(co   instanceof EditableValueHolder) { EditableValueHolder ev = (EditableValueHolder)comp; if (!ev.isImmediate()) {
ev.setImmediate(true); modifiedComponents.add(comp); } } return true; } };
```

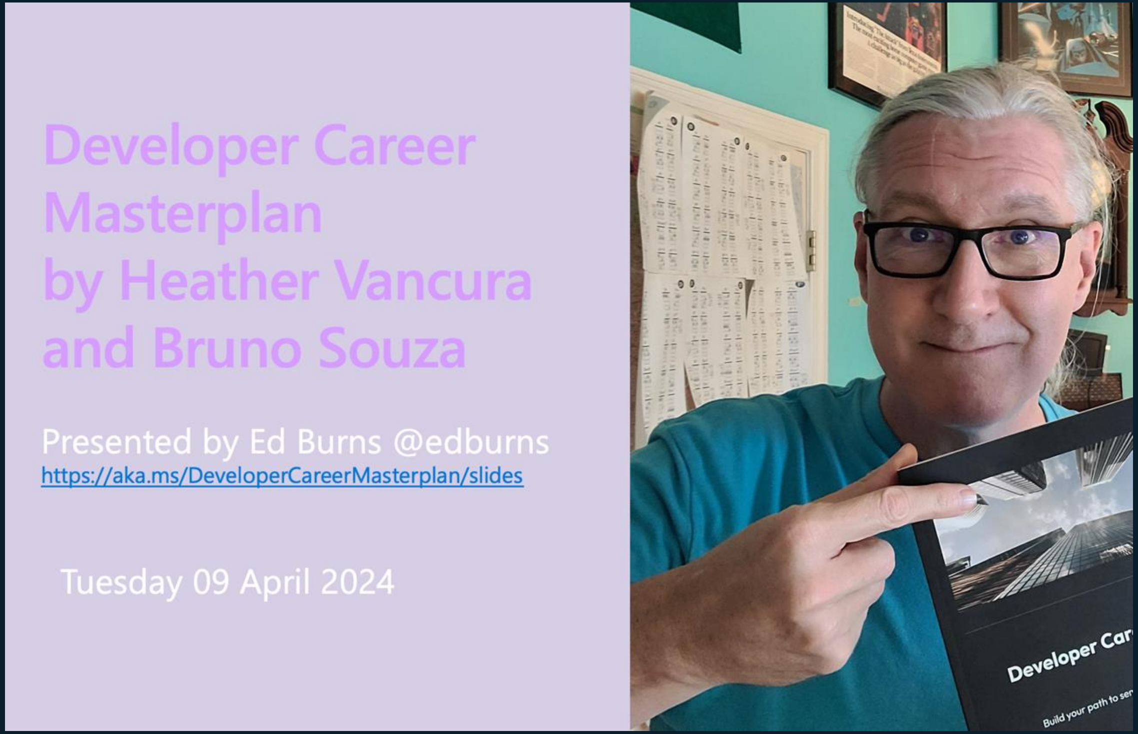
Packaging a new idea for adoption: deliver it on the back of existing ideas.



James Gosling: I realized if I wanted bankers to use automatic garbage collection, I needed to give them curly braces rather than lots of irritating silly parentheses (LISP).

AI related career advice

Everyone I know admits: AI is moving faster than anything that has come before

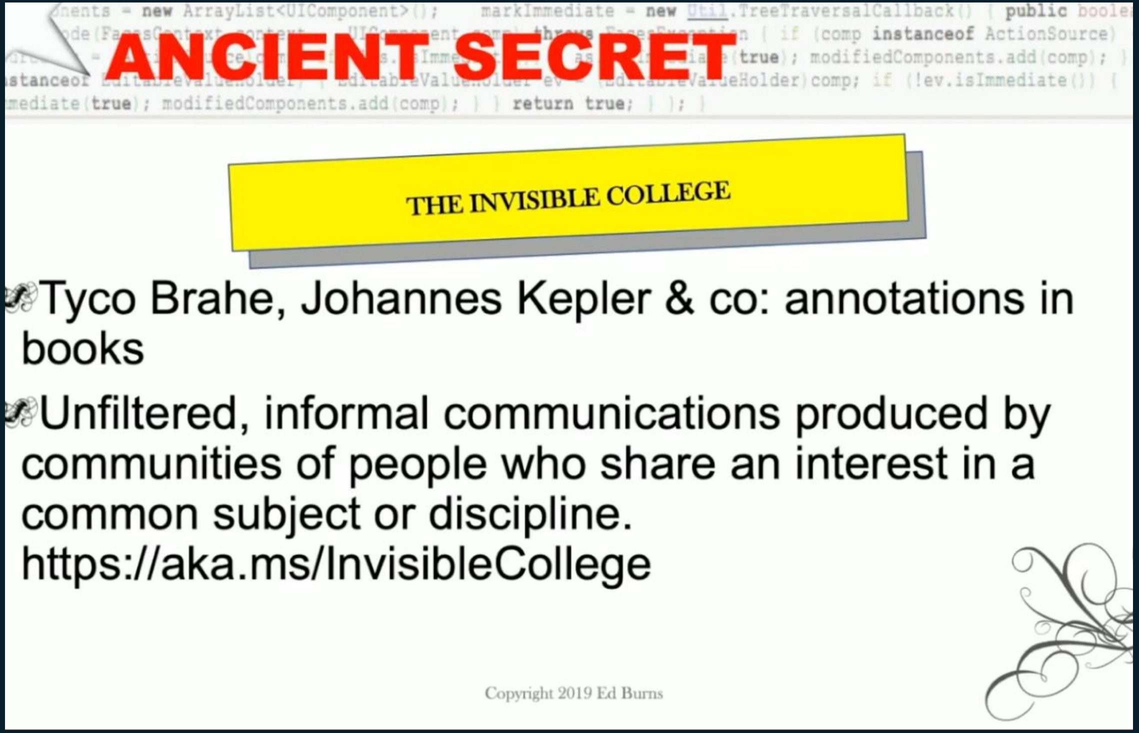


Developer Career Masterplan
by Heather Vancura
and Bruno Souza

Presented by Ed Burns @edburns
https://aka.ms/DeveloperCareerMasterplan/slides

Tuesday 09 April 2024

https://aka.ms/DeveloperCareerMasterplan/slides



ANCIENT SECRET

THE INVISIBLE COLLEGE

- Tyco Brahe, Johannes Kepler & co: annotations in books
- Unfiltered, informal communications produced by communities of people who share an interest in a common subject or discipline. https://aka.ms/InvisibleCollege

Copyright 2019 Ed Burns

https://aka.ms/RockStarProgrammerSecrets

Scott Hanselman: Let’s all just chill

From a trust perspective

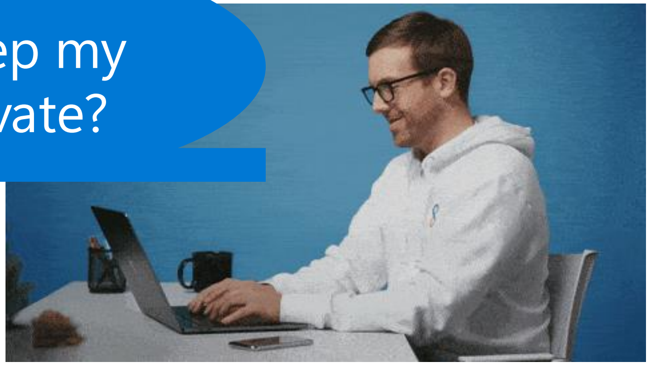
- Very important to keep the “intelligence” in AI in perspective.
- LLMs cannot answer simple questions such as “Which episode of Gilligan’s Island was about mind reading?”
  - The question that no LLM can answer and why it is important

https://www.mindprison.cc/p/the-question-that-no-llm-can-answer

How to chill? Building some trust helps.

GitHub Copilot Trust Center

https://copilot.github.trust.page/faq



Can I keep my code private?

I’m worried that my code is sent to Microsoft to do the suggestions. My concerns include:

- Can someone snoop on my code using a man-in-the-middle attack?
- My code is kept by Microsoft?
- My code is being used by Microsoft to train the model?

Copilot data flow

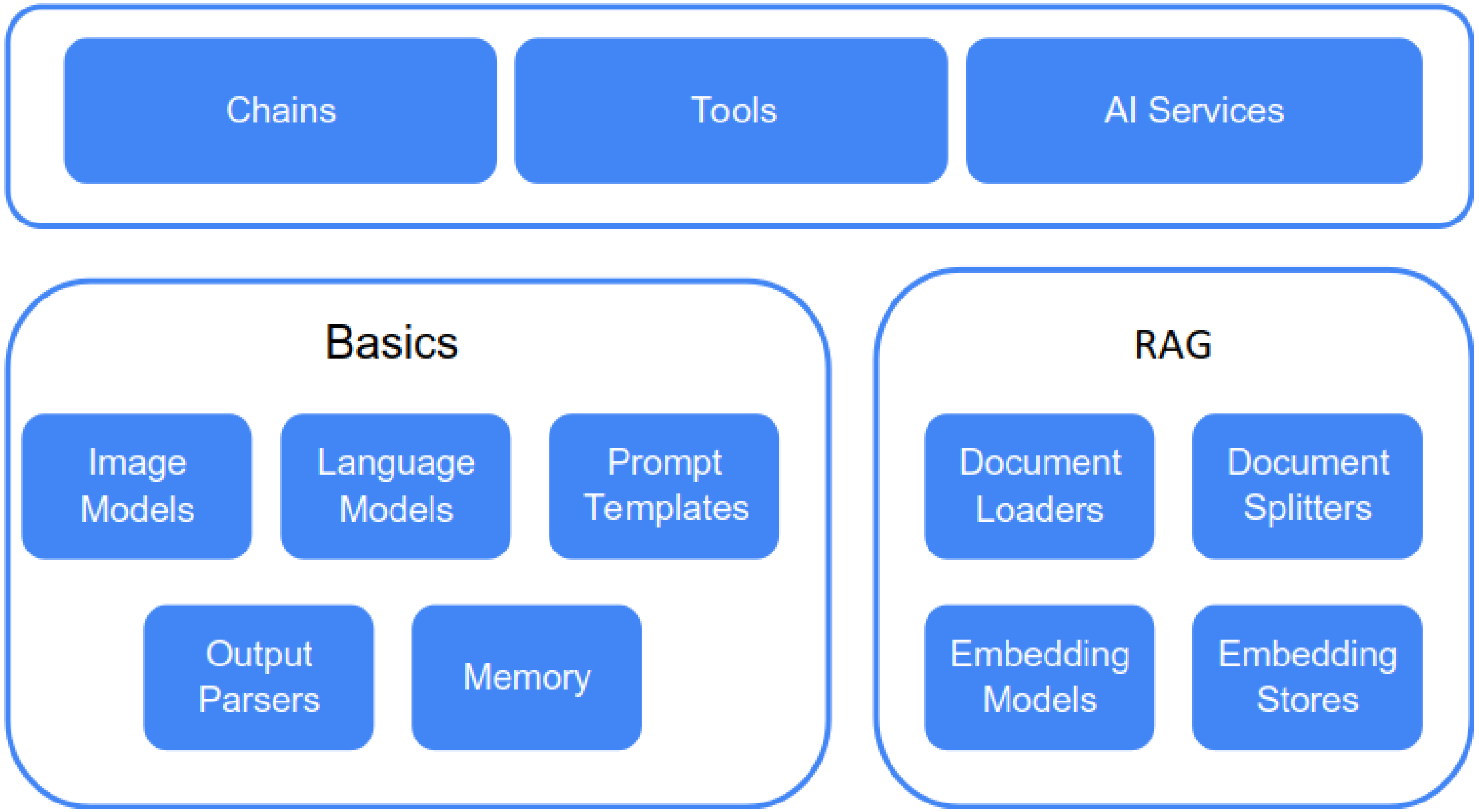
Java meets AI

Tell the LangChain4j origin story
Dmytro Liubarskyi

Java
python
LangChain

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Components of LangChain4J



Chains

Tools

AI Services

Basics

Image Models

Language Models

Prompt Templates

Output Parsers

Memory

RAG

Document Loaders

Document Splitters

Embedding Models

Embedding Stores

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Quick look at LangChain4j examples

- https://github.com/langchain4j/langchain4j-examples

Jakarta EE meets AI

smallrye-llm

langchain4j-microprofile-jakarta

CDI

JAKARTA EE

Microprofile config
Microprofile Fault Tolerance Microprofile OpenTelemetry

Set up telling the value prop. A few slides from now

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Value proposition for Langchain4j-microprofile-jakarta

- Bring the value of open standards to the rapidly evolving world of AI.
- AI-enable everything in Jakarta EE with all the power of LangChain4j.
- **Without introducing any new AI standards.**
- Builds on
  - Jakarta Contexts and Dependency Injection
  - MicroProfile Config
  - MicroProfile Fault Tolerance
  - MicroProfile Telemetry
- Delivered as Build Compatible CDI extension

Components of SmallRye LLM/LangChain4J Jakarta

1. **AI Service Interface Processing**:
   - Detects interfaces annotated with @RegisterAIService
   - Generates proxy implementations using LangChain4J
   - Makes them available as CDI beans
2. **Plugin Bean Creation**:
   - Reads configuration from properties with smallrye.llm.plugin.* prefix
   - Creates beans for LangChain4J components (models, embedding stores)
   - Makes these components available for injection
3. **MicroProfile Config Integration**:
   - Translates properties to LangChain4J configuration
   - Handles environment variable substitution (e.g., ${azure.openai.api.key})
4. **Fault Tolerance**:
   - Provides integration with MicroProfile Fault Tolerance
   - Adds resilience to AI service calls
5. **Telemetry**:
   - Provides observability for AI service calls
   - Integrates with OpenTelemetry

Some Code

CDI Injection

Injection.java

```java
@Inject
@Named(value = "ollama")
ChatLanguageModel chatModel;

@Inject
@Named(value = "embedding-store-retriever")
ContentRetriever retriever;
```

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

AI Service

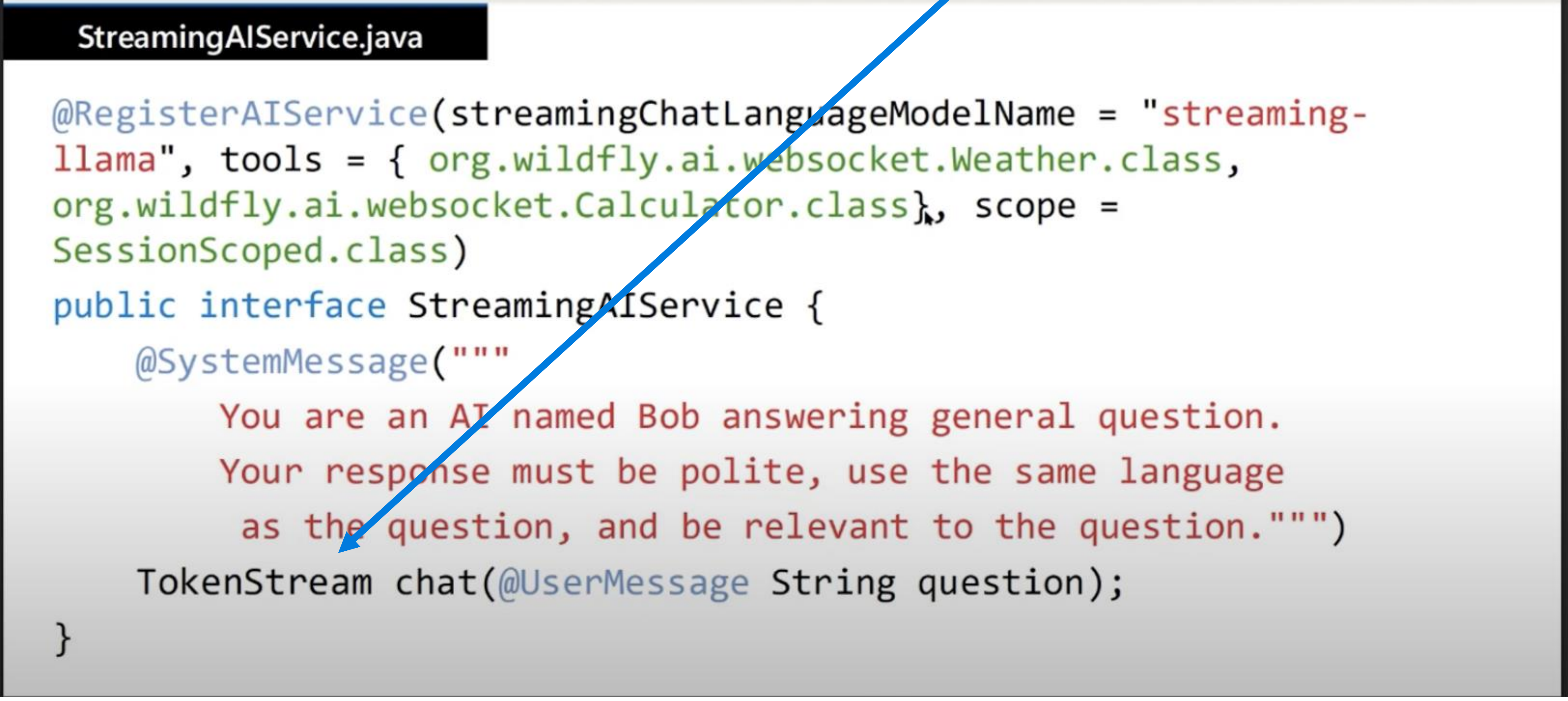
SimpleAIService.java

```java
@RegisterAIService(chatLanguageModelName = "ollama", tools =
{org.wildfly.ai.websocket.Weather.class,org.wildfly.ai.websocket.Cal
culator.class}, scope = SessionScoped.class)
public interface SimpleAIService {
    @SystemMessage("""
        You are an AI named Bob answering general question.
        Your response must be polite, use the same language
         as the question, and be relevant to the question.""")
    String chat(@UserMessage String question);
}
```

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

AI Service

TokenStream instead of String



StreamingAIService.java

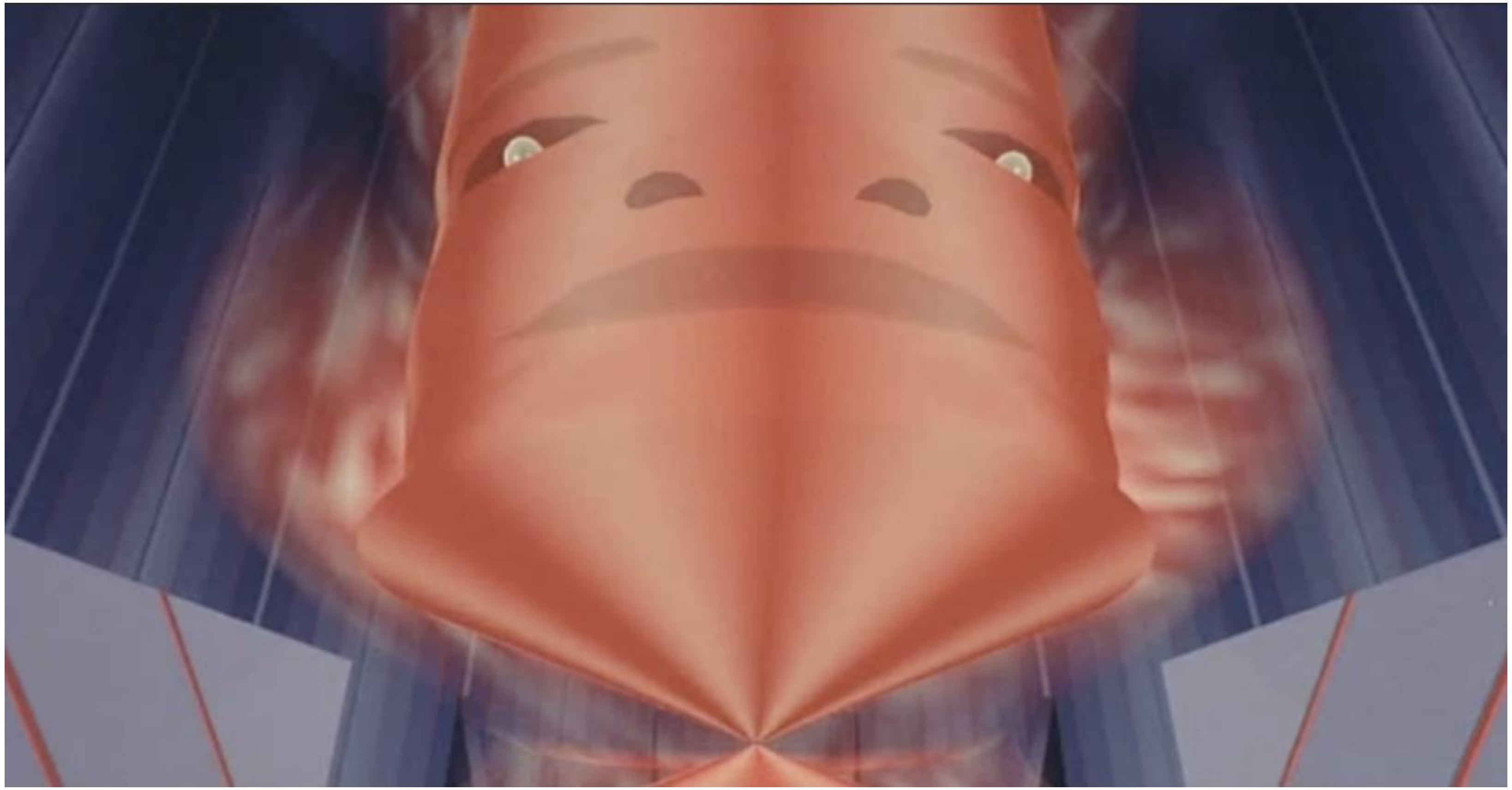
```
@RegisterAIService(streamingChatLanguageModelName = "streaming-
llama", tools = { org.wildfly.ai.websocket.Weather.class,
org.wildfly.ai.websocket.Calculator.class}, scope =
SessionScoped.class)
public interface StreamingAIService {
    @SystemMessage("""
        You are an AI named Bob answering general question.
        Your response must be polite, use the same language
         as the question, and be relevant to the question.""")
    TokenStream chat(@UserMessage String question);
}
```

Demo

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Another approach, with MCP

But what is MCP?



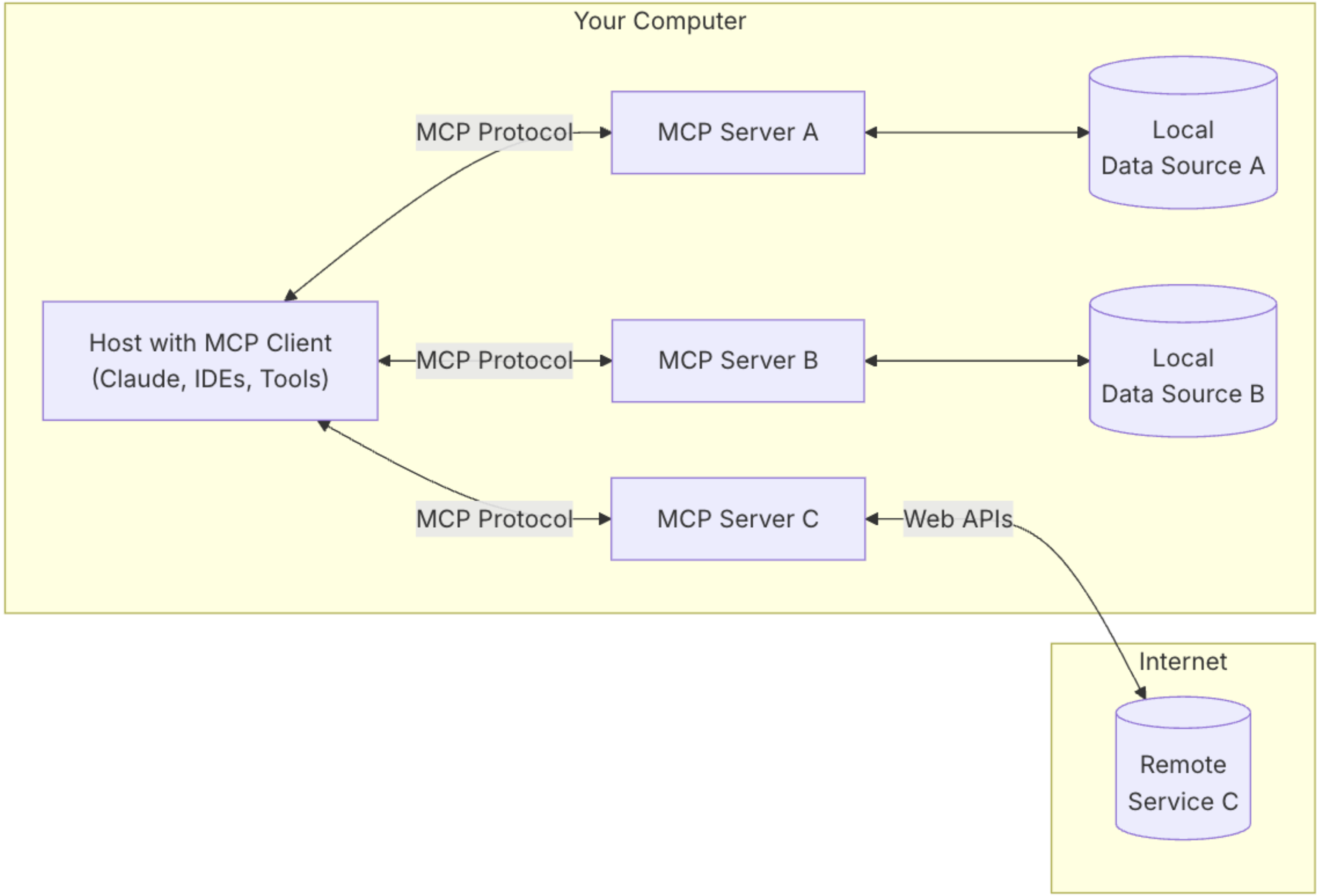
But what is MCP?

Model Context Protocol

https://modelcontextprotocol.io/

But what is MCP?

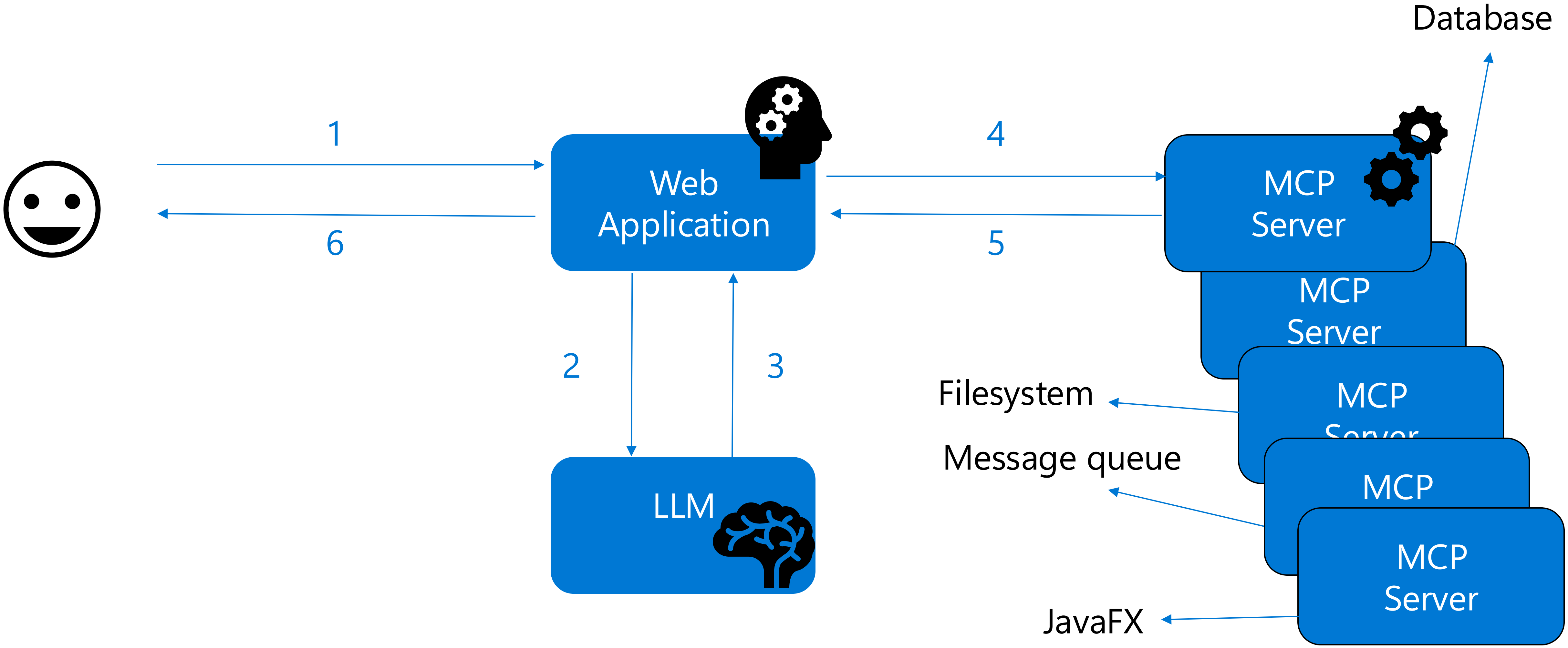
- https://github.com/modelcontextprotocol/servers



Demo

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

MCP tools workflow



Content from Emmanuel Hugonnet http://www.ehsavoie.com/

AI Service with MCP Tools

SimpleAIService.java

```
@RegisterAIService(streamingChatLanguageModelName= "streaming-ollama", toolProviderName = "mcp")
public interface ChatAiService {
    @SystemMessage("""
        You are an AI named Bob answering general question.
        Your response must be polite, use the same language
         as the question, and be relevant to the question.""")
    TokenStream streamingChat(@UserMessage String question);
}
```

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Demo

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Conclusion

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

What we saw

External configuration instead of code

CDI Injection

AI Service support

Cross-vendors: Micronaute, OpenLiberty, Quarkus, WildFly

Enterprise ready : OpenTelemetry / Fault Tolerance / MCP

Makes LangChain4J even easier to use !

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

Resources

- LangChain4J:
  - https://docs.langchain4j.dev/
- Model Context Protocol:
  - https://spec.modelcontextprotocol.io/
- smallrye-llm: a cross vendor LangChain4J -CDI integration library
  - https://github.com/smallrye/smallrye-llm
- WildFly AI Feature Pack:
  - https://github.com/wildfly-extras/wildfly-ai-feature-pack
- WildFly MCP: to interact with WildFly servers through MCP
  - https://github.com/wildfly-extras/wildfly-mcp
- Admin your WildFly with LLM and MCP
  - https://www.youtube.com/watch?v=wg1hAdOoe2w
- Making WildFly Glow with Intelligence:
  - https://www.wildfly.org/news/2025/02/10/Glowing-with-AI/
- Playing with Generative AI with WildFly:
  - https://www.wildfly.org/news/2024/11/04/WildFly-playing-with-generative-ai/

Content from Emmanuel Hugonnet http://www.ehsavoie.com/

- Please fill out this survey
- 5 winners will be chosen to win a $100 digital gift card
- https://aka.ms/edburns/survey2025

Thank you

Ed Burns
@edburns on most of the things

1.What is your key decision making process when it comes to app modernization?

2.What’s your key challenge when modernizing your Java workload?

3.Are you a Java only shop or multiple language shop?

4.What is the tech stack/framework for your Java apps?

5.When migrating, what is your primary destination on cloud? What platform are you choosing? VM? PaaS? Containers? Why?

6.How did you or your team make the above decisions?