

MicroProfile AI minutes

Please Donate 

Directly Via Monero

monero:4B5eqi9Uus4f3ChdBGuvEqdt9G6qXrNfbFQA5oJMyvua27biRrxM3ktNJEiEsoNSv7Uhe7yBxHSmKBGJdDhJW9UeRaYHpAG

Or with other currencies (5\$ USD)

<https://nowpayments.io/payment/?iid=6008069925>

12th September 2025

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Don Bourne (IBM)
- Emily Jiang (IBM)
- Ed Burns (Microsoft)
- Emmanuel HUGonnet (IBM)
- Buhake Sindi (Individual)
- Yann Blazart (Individual)

Minutes:

- Don tried the samples with Liberty against the latest main branch.
 - Ollama model: 3.1 does not have the capability to run the model well.
 - Gemini light model: [gemma](#)
- More documentations
 - To teach end users how to use these libraries
 - A written article should be written to get posted to [jakarta.ee](#) and microprofile.io
- Buhake reported there is an issue with moving up MP versions. Buhake to take a further look and open an issue if it is still a problem.
 - Buhake to do more research next week.
- TechXChange
 - Who's coming to TechXChange? Please send an email to edburns@acm.org to connect. It's Ed's hometown.
 - Collaboration venue for Ed and Emily

- Proposals for where?
 - ~~A topic branch on Ed's fork:~~
<https://github.com/azure-javaee/langchain4j-cdi> -
 - SELECTED: A topic branch on upstream:
<https://github.com/langchain4j/langchain4j-cdi> .
 - Details
 - Branch name **presentations**.
 - Set up git-lfs
 - Per-event directory: **2025-techXchange**
 - Tool that can be used for presentations: <https://revealjs.com/>
 - ~~A purpose-built fresh repo in Ed's personal GitHub.~~
 - ~~Some non-GitHub thing.~~
- Proposals for what
 - SELECTED: New directory in GitHub repo selected above:
 presentations/2025-techXchange.
- Constraints
 - LICENSE is the same as the rest of the repo.
 - No constraint on content as long as the main point is langchain4j-cdi. For example, no problem to have a talk that is about how LangChain4J-cdi works well in <whatever environment>.
 - No constraint on file format.

5th September 2025

Contributors [each attendee must add themself to minutes]

- Arjav Desai (Oracle)
- Ed Burns (Microsoft)
- Emily Jiang (IBM)
- Emmanuel Hugonnet (IBM)
- Buhake Sindi (Individual)
- Don Bourne (IBM)

Minutes:

- Don tried the samples with Liberty against the latest main branch.
 - AZure AI should not be mandatory issue: resolved
 - Ollama works partially. There was an issue with some small LLMs.
 - Serviceability - calls fallback without info on why it calls the fallback. It seems the fallback reason did not log the reason.
- Buhake reported there is an issue with moving up MP versions. Buhake to take a further look and open an issue if it is still a problem.

29th August 2025

Video: [here](#)

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Andrew Rouse (IBM)
- Emmanuel Hugonnet (Red Hat)
- Ed Burns (Microsoft)
- Emily Jiang (IBM)
- Buhake Sindi (Individual)

Minutes:

- Updated PR to be able to release: <https://github.com/langchain4j/langchain4j-cdi/pull/63>
- Test the release workflow:
 - We need to create docs to explain the modules.
 - Might be better to use the stage repo. Work with the langchain4j community to investigate how to set up a stage repo.
 - For the release, use a release tag to do the release.

22th August 2025

Video: [here](#)

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Andrew Rouse (IBM)
- Ed Burns (Microsoft)
- Emmanuel Hugonnet (Red Hat)
- Majid Mostafavi
- Buhake Sindi (Individual)

Minutes:

- Ed is speaking about this work (I only help a little here and there) at IBM JCON@TechXChange <https://reg.tools.ibm.com/flow/ibm/techxchange25/sessioncatalog/page/sessioncatalog/session/1746107434822001TBmB> This is October 6 - 10.
- Reviewed Emmanuel's PR 57 <https://github.com/langchain4j/langchain4j-cdi/pull/57>

15th August 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Emily Jiang (IBM)
- Ed Burns (Microsoft)
- Andrew Rouse (IBM)
- Buhake Sindi (Individual)

Minutes:

- Build is at the good stage.
- Integration tests are to be done. Emmanuel will work on it.
- Review the PRs
 - Comments: when creating PRs, please raise an issue to describe the problems.
- RESOLVED: from now on, all of the contributions should come from a fork not creating branches from the repo. The only active branch should be **main**. All other remote branches would be deleted next week.
- Put the best practice to the [contribution.md](#)
- Discussed a bit on MCP api [smallrye-mcp](#). We can promote the APIs so that they can be adopted by multiple runtimes.

8th August 2025

Video: here

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Emmanuel Hugonnet (RedHat)
- Don Bourne (IBM)
- Phil Chung (Oracle)
- Ed Burns (Microsoft)
- Buhake Sindi (Individual)

Minutes:

- We got the sonatype login and have it set up on repo. Snapshots are being published at <https://central.sonatype.com/repository/maven-snapshots/dev/langchain4j/cdi/langchain4j-cdi-core/1.0.0-SNAPSHOT/maven-metadata.xml> using workflow/action https://github.com/langchain4j/langchain4j-cdi/actions/workflows/snapshot_release.yaml
- gpg signing is not required for snapshot but will be needed for release. Emmanuel can create a key using gmail account: langchain4j.cdi@gmail.com
- As of now sonatype login is in repo setting and used by action only, we should document this in README along with gpg key and gmail account information.
- Use of Ollama for the examples (relates to <https://github.com/langchain4j/langchain4j-cdi/issues/21>)
 - Discussed making Ollama the default for the examples (but not breaking the existing integration with Azure OpenAI)
 - Nobody has concerns on using Ollama as the default

1st August 2025

Video: here

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Emmanuel Hugonnet (RedHat)

- Don Bourne (IBM)
- Ed Burns (Microsoft)
- Emily Jiang (IBM)
- Luís Neto (Payara)
- Buhake Sindi (Individual)
- Phil Chung (Oracle)

Minutes:

- We are still waiting to get publishing account/credentials, see <https://discord.com/channels/1156626270772269217/1202386909410623519/1399377563234865246>
 - Working on the user credentials to get the release going
- Langchain4j-cdi AI Agents
 - Java SDK: <https://github.com/a2aproject/a2a-java>
 - JakartaEE SDK: <https://github.com/wildfly-extras/a2a-java-sdk-server-jakarta>
 - Blog entry: <https://github.com/wildfly/wildfly.org/pull/820>
- MCP servers
 - Historical context of Java MCP
 - The official Java SDK for MCP was written by Christian Tzolov and a team from Broadcom and donated to Anthropic.
 - This SDK has a dependency on Project Reactor
 - Christian Tzolov proposed a remedy, but it includes a dependency on RS Publisher: <https://github.com/orgs/modelcontextprotocol/discussions/246#discussioncomment-12854366>
 - Eventually, someone who had a need to use the official Java SDK for MCP with a technology stack that is not compatible with Project Reactor noticed this and filed this issue: <https://github.com/orgs/modelcontextprotocol/discussions/246> .
 - Related issues
 - <https://github.com/modelcontextprotocol/java-sdk/issues/38> “Major issue - dependency on reactor, slf4j and netty”
 - <https://github.com/modelcontextprotocol/java-sdk/pull/175> “On branch edburns/38-reactor-mcp-sdk Fixes #38, but in a very trivial way.”
 - Ed requests advocacy for getting this merged. At least it will make discussions about the problem more clear.
 - Buhake introduced the Java JDK implementation of MCP client and server, with no dependency of Project Reactor.
 - API implementation: <https://github.com/sinditech/sindi-ai-mcp>

- MCP server implementation (requires API implementation, mentioned above):
<https://github.com/sinditech/sindi-ai-mcp-server>
- MCP clients to connect MCP tools
- Not provide Tool
 - MCP SDK: spring Java Spring Reactor

25th July 2025

Video: here

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Emmanuel Hugonnet (RedHat)
- Buhake Sindi (Individual)

Minutes:

- Arjav: We have merged couple of more PRs to fix snapshot publishing and now waiting to get publishing account/credentials, see <https://discord.com/channels/1156626270772269217/1202386909410623519/1398231865046863902>
- Emmanuel: We also need to talk about gpg signing keys. Langchain4j main/core is signing only for release and not for snapshots.
- Emmanuel: Langchain4j main/core is using deploy and not release in maven. So we have to talk about how to do tagging after release.
- Emmanuel mentioned there is some A2A work/PoC going on in Langchain4j. We talked about having MCP client injection, which Emmanuel/Buhake said should already work.
- Buhake has written a library <https://github.com/sinditech/sindi-ai-mcp-server> which enables services to be exposed as MCP Server. It's Java only without the project reactor and using CompletableFuture only.

18th July 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Ed Burns (Microsoft)
- Buhake Sindi (Individual)

Minutes:

- We reviewed and merged <https://github.com/langchain4j/langchain4j-cdi/pull/37> to add a snapshot release.
- Buhake joined but after we closed the call.

11th July 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Don Bourne (IBM)
- Arjav Desai (Oracle)
- Emily Jiang (IBM)
- Ed Burns (Microsoft)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)

Minutes:

- Release: Emmanuel is investigating how to release
 - LC4J had beta releases because of lack of staging ability
- There is no PR build or daily build in LC4J-CDI.

- We should be able to use/copy/follow <https://github.com/langchain4j/langchain4j/blob/main/.github/workflows/snapshot-release.yaml>
- Please drop your committer information here
 - Don Bourne
 - Gh: donbourne
 - Ed Burns
 - GitHub: edburns
 - Email: edburns@microsoft.com
 - Arjav Desai
 - GitHub: arjav-desai
 - Email: arjav.desai@oracle.com
 - Phil Chung
 - GitHub: philchung
 - Email: phil.chung@oracle.com

4th July 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Don Bourne (IBM)
- Clement
- Emmanuel
- Buhake Sindi (Individual)

Minutes:

- (Don) evaluated the example for Liberty. Potential issues:
 - Azure Open AI is used for Liberty example (we shouldn't require people to have an Azure cloud account to run our example)
 - Hugging face dependency is in the pom.xml (not used - should remove it)
 - There is no easy way to use Ollama (should perhaps use that as the default for the pom.xml and microprofile-config.properties)
 - Looking at the log-request output, the fault tolerance code seems to have a problem - retries should not repeat the user message. Per Clement, can't use plain fault tolerance with LC4J - need to manipulate the messages when doing a retry (may need to de-duplicate)

```
- }, {
-   "role" : "user",
```

```

-   "content" : "hello, what's up?"
- }, {
-   "role" : "user",
-   "content" : "hello, what's up?"
- }, {
-   "role" : "user",
-   "content" : "hello, what's up?"
- } ],

```

- After making changes to the pom.xml / microprofile-config.properties to make the example use Ollama, it does not return a valid response.
 - Using the current content of the request generated by the langchain4j-cdi code:

```

curl -X POST http://localhost:11434/api/chat -H
'Content-Type: application/json' -d '
{
  "messages" : [ {
    "role" : "system",
    "content" : "You are a customer support agent of a car
rental company named Miles of Smiles. Before providing
information about booking or canceling a booking, you MUST
always check: booking number, customer name and surname.
You should not answer to any request not related to car
booking or Miles of Smiles company general information.
When a customer wants to cancel a booking, you must check
his name and the Miles of Smiles cancellation policy first.
Any cancelation request must comply with cancellation
policy both for the delay and the duration. Today is
2025-07-04. "
  }, {
    "role" : "user",
    "content" : "hello, whats up?"
  }, {
    "role" : "user",
    "content" : "hello, whats up?"
  } ],
  "options" : {
    "stop" : [ ]
  },

```

```
    "stream" : false
  }
  '
```

- returns: {"error": "model is required"}
- Hand editing to add the "model" key/value:

```
curl -X POST http://localhost:11434/api/chat -H
'Content-Type: application/json' -d '
{
  "model" : "llama3.2",
  "messages" : [ {
    "role" : "system",
    "content" : "You are a customer support agent of a car
rental company named Miles of Smiles. Before providing
information about booking or canceling a booking, you MUST
always check: booking number, customer name and surname.
You should not answer to any request not related to car
booking or Miles of Smiles company general information.
When a customer wants to cancel a booking, you must check
his name and the Miles of Smiles cancellation policy first.
Any cancelation request must comply with cancellation
policy both for the delay and the duration. Today is
2025-07-04. "
  }, {
    "role" : "user",
    "content" : "hello, whats up?"
  }, {
    "role" : "user",
    "content" : "hello, whats up?"
  } ],
  "options" : {
    "stop" : [ ]
  },
  "stream" : false
}
'
```

- returns:

```
{ "model": "llama3.2", "created_at": "2025-07-04T16:08:41.3829Z", "message": { "role": "assistant", "content": "Hello! I'm here to help you with any questions or concerns you have about your car rental experience with Miles of Smiles. What can I assist you with today? Are you looking to book a new rental, or perhaps need assistance with an existing booking?" }, "done_reason": "stop", "done": true, "total_duration": 785736791, "load_duration": 69680916, "prompt_eval_count": 149, "prompt_eval_duration": 99196625, "eval_count": 53, "eval_duration": 61594666 }
```

27th June 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Don Bourne (IBM)
- Luis
- Emmanuel
- Buhake
- Arjav Desai (Oracle)

Minutes:

- From Clement (on PTO today, plans to join next week):
 - About the Maven Central publication, I will check with Dmytro to add one of the admins (probably Emmanuel Hugonnet) to have the right to publish to Central. Now, I would keep the version number below 1.0, as there are still so many movement that it should not be considered stable (agentic is changing a lot of internals)
 - Regarding the retirement of Smallrye LLM, yes, we should either delete the repository or make it read-only.
- Emmanuel - suggest we archive the smallrye LLM repo
- Don - trying to run the example for Liberty - encountered some problems. Will do some debug and see if a PR is needed.
- Emmanuel - running the integration testing - it's working
- Buhake - for issue related to fault tolerance, going to bring it to jakara ee group.
 - Should aim to fix it in langchain4j-cdi (not smallrye repo)

- Emmanuel - goal is to release langchain-cdi as soon as possible, so stop evolving smallrye llm
 - Not sure who has the permission to make the repo read only
 - Will ask Roberto Cortez about it
- Luis - can help run the examples in the new repo and verify if everything is working
- Arjav - had joined for the first few minutes and then dropped.

20th June 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Don Bourne (IBM)
- Luís Neto (Payara)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)

Minutes:

- Closed PR 3. Merged PR 4
- Langchain cdi is now in langchain
- Someone needs to try out the newly merged code
 - Anyone can try out the new code (clone and try examples from <https://github.com/langchain4j/langchain4j-cdi/tree/main/examples>)
- Buhake looking at <https://github.com/langchain4j/langchain4j-cdi/issues/15>
- Should next publish packages to maven central
 - Perhaps start with a SNAPSHOT
 - Need to talk to Clement perhaps to figure out how to publish that
 - Don to ping Clement to see if he can join call next week
- Consider retiring smallrye llm eventually
 - This would also be good conversation with Clement
- Reza also trying to set up an AI framework in jakarta space
- Buhake started discussion on fault tolerance - Buhake to re-ping in the thread to try to call to people that could help resolve the issue.

13th June 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Don Bourne (IBM)
- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)
- Luís Neto (Payara)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)
- Emily Jiang (IBM)

Minutes:

- Don and Phil shared their experiences of talking at Toronto JUG and GardenState JUG respectively.
- PRs: <https://github.com/langchain4j/langchain4j-cdi/pull/4>
- We discussed how to maintain the commit history.
 - Emmanuel and Buhake to try it out.
- Buhake opened the discussion mentioned on June 6, 2025 on the Microprofile Group: <https://groups.google.com/g/microprofile/c/lt05lHWkN1A>

6th June 2025

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Don Bourne (IBM)
- Ed Burns (Microsoft)
- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)
- Luís Neto (Payara)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)

Minutes:

- Updated PR: <https://github.com/langchain4j/langchain4j-cdi/pull/4> - this should supercede PR 3
 - Need to review PR 4
 - Still references smallrye fault tolerance (not MP fault tolerance as was suggested in our discussion last week)

- The issue: When depending on MP fault tolerance, it was necessary to include a dependency on a concrete class (from SmallRye), which makes it so the dependency chain includes non-standard jars.
 - You can't do fault tolerance on something that is a synthetic bean or a proxy bean. There is no way to provide an interceptor binding to the proxy.
 - When we discussed this last week, we resolved to make it so the classes that we were using from SmallRye to satisfy the MP fault tolerance requirement, would **instead** be in our code directly. This would make it so there was no dependency on SmallRye.
 - Emmanuel: it's best to do this as a separate commit, so we can better mitigate it later.
 - PR 4 renames the artifact from langchain4j-core to langchain4j-cdi-core because we don't want to have the same name as langchain4j's artifact.
- Heads up
 - Thoughts on <https://medium.com/@springrod/embabel-a-new-agent-platform-for-the-jvm-1c83402e0014> ?
- ACTION: Buhake: Ensure everything in your PR 3 is represented in Emmanuel's PR 4. Then mark PR 4 closed.

30th May 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Arjav Desai (Oracle)
- Don Bourne (IBM)
- Ed Burns (Microsoft)
- Buhake Sindi (Individual)

Minutes:

- Review [the PR](#)

23rd May 2025

Video Recording

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Arjav Desai (Oracle)
- Luís Neto (Payara)
- Buhake Sindi (Individual)

Minutes:

- Review [the PR](#)

16th May 2025

Video Recording

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Buhake Sindi (Individual)
- Don Bourne (IBM)
- Max Rydahl Andersen (Red Hat)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)

Minutes:

- The current progress of move code from smallrye-llm to langchain4j-cdi
 - Buhake to create a PR from the langchain4j-cdi branch against the langchain4j-cdi
- The project should target MP 7.0 and Jakarta EE 10
 - This basically means CDI api should be 4.0 and Weld impl should be based on that only.
- The project should not depend on smallrye
- The pom.xml should be updated to pull in the current version of the specs or impls

9th May 2025

Video Recording

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Ed Burns (Microsoft)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)
- Don Bourne (IBM)

Minutes:

- Only met for a few minutes. No recording.
- Langchain4j repo structure - ready to be populated
- This blog seems interesting
<https://techcommunity.microsoft.com/blog/appsonazureblog/the-state-of-coding-the-future-with-java-and-ai-%E2%80%93-may-2025/4411210>
- Java version: 17. It is fine.

2nd May 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Don Bourne (IBM)
- Max Rydahl Andersen (Red Hat)
- Arjav Desai (Oracle)
- Phil Chung (Oracle)
- Luis Neto (Payara)
- Buhake Sindi (Individual)

Minutes:

- Langchain4j repo structure
 - Instead of using MicroProfile or Jakarta, Max suggested creating a repo called langchain4j.cdi. Using MicroProfile and Jakarta EE, it could be confusing as people might think this is a standard.
 - Do we create an umbrella repo to contain multiple modules including Fault Tolerance, Telemetry, etc?
 - langchain4j-microprofile-jakarta=>langchain4j-cdi (2025-05-09: Microsoft is happy to see this)
 - langchain4j-microprofile.>langchain4j-*
 - langchain4j-cdi-build-compatible-ext=>langchain4j-build-compatible-ext

- langchain4j-cdi-portable-ext=>langchain4j-portable-ext

22nd April 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM) - only can stay for a few minutes
- Don Bourne (IBM)
- Clement Escoffier (Red Hat)
- Luís Neto (Payara)
- Arjav Desai (Oracle)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)

Minutes:

- Buhake has updated <https://github.com/smallrye/smallrye-llm/tree/langchain4j-microprofile-jakarta> to include last week's GAV changes. It's not pushed to langchain4j repo since we can't make a release from that until we finalize the GAVs. This can take time depending on how Microprofile hosting in Jakarta project discussion progressed.
- How do we make progress and not be blocked while the process is ongoing?
 - Can we push to langchain4j repo but not make a release so that we can continue work there?
 - Emily: Push to the new repo with a PR
 - Alternate is to continue in smallrye repo.
 -
- What do we want to work on next? Buhake suggested MCP support, maybe.
 - Emily: MCP is a good direction
- Agree on a new meeting slot:
 - Use the doodle from Ed Burns to vote (discussion [here](#))
- Langchain4J 1.0 expected in May
 - CR 1 next week
- MCP
 - New version of the protocol
 - SSE transport is deprecated, replaced with the streamable HTTP
 -
 - Declarative MCP - MCP annotation, Red Hat would like to push them to Anthropic as an official SDK

- Annotations are a bit too dependent on CDI right now
 - No capabilities extraction yet
- A2A and ADK
 - Provide a higher level of abstraction to compose agents (which then use A2A to interact)

15th April 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Luís Neto (Payara)
- Buhake Sindi (Individual)
- Ed Burns (Microsoft)
- John Clingan (Red Hat)
- Jan Westerkamp (iJUG)

Minutes:

- Buhake has updated <https://github.com/smallrye/smallrye-llm/tree/langchain4j-microprofile-jakarta> to include last week's GAV changes. It's not pushed to langchain4j repo since we can't make a release from that until we finalize the GAVs. This can take time depending on how Microprofile hosting in Jakarta project discussion progressed.
- How do we make progress and not be blocked while the process is ongoing?
 - Can we push to langchain4j repo but not make a release so that we can continue work there?
 - Alternate is to continue in smallrye repo.
- What do we want to work on next? Buhake suggested MCP support, maybe.

8th April 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Clement Escoffier (Red Hat)

- Don Bourne (IBM)
- Arjav Desai (Oracle)
- Luís Neto (Payara)
- Buhake Sindi (Individual)

Minutes:

- Group ID / Artifact id/
 - Parent
 - Group id: dev.langchain4j.microprofile
 - Artifact id: langchain4j-microprofile-parent
 - Core
 - Artifact id: langchain4j-microprofile-core
 - Config
 - Artifact Id: langchain4j-microprofile-config
 - Fault tolerance
 - Artifact Id: langchain4j-microprofile-fault-tolerance
 - Telemetry
 - Artifact Id: langchain4j-microprofile-telemetry
 - Portable extensions
 - Artifact Id: langchain4j-cdi-portable-ext
 - Build compatible extensions
 - Artifact Id: langchain4j-cdi-build-compatible-ext

Action: Buhake and Emmanuel will use the above ids for their PRs, Before the release, we will decide the final ids.

31st March 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emmanuel Hugonnet (Red Hat)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)
- John Clingan (Red Hat)

Minutes:

- Emmanuel suggested that we can provide examples (similar to Langchain4J Spring Boot example) similarly for SmallRye-LLM.
 - Possible concern: Helidon 4.2?

- Integration with Helidon Fault Tolerance (package private interceptor binding)
- Buhake suggests possible repo name change:
 - Project groupId: dev.langchain4j.microprofile.jakarta
 - Artifact ID: langchain4j-cdi
 - So, for MP FT: langchain4j-cdi-microprofile-fault-tolerance
 - Clement suggested that CDI is a well-known name of Jakarta so rather use that and it's easier for developers to understand.
 - Once the groupId is defined on Maven Central, it's difficult to change (also poses security concerns)
 - More discussion on MP WorkingGroup Slack group.:
<https://microprofile-wg.slack.com/archives/C07HLJBQ36Y/p1743330727930879>
 - Possibility to put this on pause while the discussion of project name is underway and also the possibility of merging MP into Jakarta EE?

24th March 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Don Bourne (IBM)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)
- Luís Neto (Payara)

Minutes:

- Our repo has been created [here](#)
 - Populate the repo
 - Group ID dev.langchain4j.microprofile.jakarta
 - Artifact ID langchain4j-microprofile-jakarta-parent
 - Update the cdi integration
 - Langchain4j-cdi-buildcompatible-extension
 - Langchain4j-cdi-portable-extension
 - Revisit the name at a later time
 - Build infrastructure - deploy and release to maven central
 - Emmanuel and Clement would be able to get the infrastructure done with the help from [Dmytro Liubarskyi](#)
 - Emmanuel will set up OSSH username and password for the Github actions

- New meeting time based on [this](#) doodle poll

17th March 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Arjav Desai (Oracle) - arjav.desai@oracle.com
- John Clingan (Red Hat)
- Don Bourne (IBM)
- Buhake Sindi (Individual)
- Emmanuel Hugonnet (Red Hat)
- Phil Chung (Oracle) - phil.chung@oracle.com
- Luís Neto (Payara)

Minutes:

- Our repo has been created [here](#)
- Committers:
- License: Apache license v2
(<https://github.com/langchain4j/langchain4j/blob/main/LICENSE>)
- Package name: prefix?
 - <https://github.com/smallrye/smallrye-llm/tree/langchain4j-microprofile-jakarta>
 - In the current work in progress branch, we need to fix copyrights and smallrye parent/testing references.
- Release process: need access to publish to maven central esp. for dev.langchain group.
- Demo code: <https://github.com/ehsavoie/car-booking/>

10th March 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Arjav Desai (Oracle)
- Jan Westerkamp (iJUG)
- Luís Neto (Payara)
- Phil Chung (Oracle)
- John Clingan (Red Hat)
- Ed Burns (Microsoft)
- Emmanuel Hugonnet (Red Hat)

Minutes:

- Repo will be created this week.
- Agentic AI
 - <https://quarkus.io/blog/agent-ai-with-quarkus/>
 - <https://quarkus.io/blog/introducing-mcp-servers/>
 - <https://dev.to/bsorrentino/ai-agent-in-java-with-langgraph4j-3c8>
 - Emmanuel working on working on wildfly example using MCP servers
 - <https://github.com/ehsavoie/car-booking>
 - <https://github.com/anthropics/anthropic-sdk-java> however the api was not needed
 - <https://github.com/wildfly-extras/wildfly-ai-feature-pack> - wildfly ai feature pack for providing a mcp server
 - RAG improvement - can use langchain4j-neo4j model for embedding retrieval

3rd March 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Ed Burns (Microsoft)
- Yann Blazart
- Phil Chung (Oracle)

- Buhake Sindi (Individual)
- Luís Neto (Payara)
- Alfonso Valdez (Payara)
- John Clingan (Red Hat)

Minutes:

- Repo name voting: <https://groups.google.com/g/microprofile/c/83xLomtGtIM>
 - The repo name langchain4j-microprofile-jakarta is the winning vote.
- Agentic AI
 - <https://quarkus.io/blog/agentic-ai-with-quarkus/>
 - <https://quarkus.io/blog/introducing-mcp-servers/>
 - <https://dev.to/bsorrentino/ai-agent-in-java-with-langgraph4j-3c8>
 - Emmanuel to produce a PR for Agentic AI

24th Feb 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Buhake Sindi (Individual)
- Ed Burns (Microsoft)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)
- Ryan Cuprak

Minutes:

- Repo name voting: <https://groups.google.com/g/microprofile/c/83xLomtGtIM>
- A repo move plan
 - Maybe a clone and put smallrye-llm in maintenance model.
 - Proposal: Create a new branch forLangchain4j
 - Rename the package names: replace io.smallrye.llm with dev.reponame while the reponame could be either langchain4j-microprofile-jakarta or langchain4j-jakarta-microprofile (pending vote)

- Then fork the branch to contribute back to langchain4j repo
- Agentic AI
 - <https://quarkus.io/blog/agentic-ai-with-quarkus/>
 - <https://quarkus.io/blog/introducing-mcp-servers/>
 - <https://dev.to/bsorrentino/ai-agent-in-java-with-langgraph4j-3c8>
 -
-

17th Feb 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Yann Blazart
- John Clingan (Red Hat)
- Buhake Sindi (Individual)
- Victor Lu

Minutes:

- Contribute the integration of langchain4j to langchain4j repo
 - Finalise the repo name: **langchain4j-microprofile-jakarta**
 - consensus: John and Emily prefer langchain4j-microprofile-jakarta; Yann and Buhake prefer langchain4j-jakarta-microprofile
 - Emmanuel, Victor don't have preference.
 - [Yann]: I prefer **langchain4j-microprofile or langchain4j-jakarta-microprofile** - sorry, Zoom does not let me in to join the conversation unfortunately
 - John: If the argument is that we put jakarta first due to name recognition or CDI integration (CDI is a Jakarta project), then all new MP specs will have to put Jakarta first. The AI effort began in MicroProfile and MicroProfile should remain first, or only, in the name. IMHO, we are already conceding ground by adding Jakarta. Even when MicroProfile AI started, Jakarta folks agreed that AI

should be done in MicroProfile due to it being "early" in AI's evolution.

- The repo name will be determined via a vote among committers via MicroProfile googlegroup and microprofile-wg.
- A repo move plan
 - Maybe a clone and put smallrye-llm in maintenance model.
- AI Agent?

10th Feb 2025

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- John Clingan (Red Hat)
- Ed Burns (Microsoft)
- Emmanuel Hugonnet (Red Hat)
- Jan Westerkamp (iJUG)
- Buhake Sindi (Individual)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)

Minutes:

- Contribute the integration of langchain4j to langchain4j repo
 - Name the repo to be **langchain4j-microprofile-jakarta**
 - Check with Eclipse Foundation to see whether it is okay to use the name of MicroProfile and Jakarta. Emily to check with Eclipse foundation
 - If there is no concerns, I will contact langchain4j community to have this repo created
- We talked about how microprofile only/specific integration or features should be added in a new project. The idea is to use a sub-directory/module for that and not have it in core/main so that jakarta projects can work without microprofile dependencies.
 - One of the issues may be microprofile config since it's used in core/main. Emily mentioned that config specification is being discussed in both communities.

3rd Feb 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Buhake Sindi (Individual)
- John Clingan (Red Hat)
- Phil Chung (Oracle)
- Ed Burns (Microsoft)
- Reza Rahman (Microsoft)
- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)
- Majid Mostafavi (IJUG)hei

Minutes:

- Discuss the current status
 - Some integrations with CDI, MP Config, Telemetry were done; RAG is in progress
- Should we move this to LangChain4j or MP?
 - Reza: Microsoft would like to commit resources to create a CDI extension within LangChain4j itself that mirrors the support for Spring and Java SE currently baked into LangChain4j. This work would have formal backing from Microsoft. However, Microsoft is seeking to do this work in a collaborative fashion with MicroProfile and Jakarta EE vendors. We would like Red Hat, Oracle, IBM, Payara, etc to also commit resourcing and organizational backing to a reasonable extent.
 - John: extension contributions in Langchain4J, what is the MP contribution towards that?
 - Reza: Use MP AI meetings and infrastructure to drive the collaboration for MicroProfile and Jakarta EE vendors.
 - John: MP Telemetry brings in OpenTelemetry and adds extensions. Does Microsoft feel there is a similar need with MicroProfile AI - a spec to integrate LangChain4J into MicroProfile?
 - Reza: The AI space is still very rapidly evolving. This is not yet the right time to standardise anything. The CDI, MicroProfile, and Jakarta EE integrations should be sorted out in LangChain4j, just as it is for Spring. If in the future the areas are stable enough, we can consider standardisation.
 - Reza: Using MicroProfile AI as a place to have conversations, experiments, and staging is appropriate, but the hands-on work should likely happen in LangChain4j, where the infrastructure may not be ideal for MicroProfile and Jakarta EE stakeholders to effectively collaborate.

27th January 2025

[Video Recording](#)

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)
- Buhake Sindi (Individual)
- John Clingan (Red Hat)
- Phil Chung (Oracle)

Minutes:

- OpenTelemetry integrated
- Fault Tolerance integrated : discussions about separating LangChain4J fault tolerance to be able to use the Microprofile Fault Tolerance retry mechanism
- New release of LangChain4J that might break compatibility
- MCP server, does it make sense ?

9th December 2024

Video Recording [here](#)

Contributors *[each attendee must add themselves to minutes]*

- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)
- Buhake Sindi (Individual)
- Emily Jiang (IBM)

Minutes:

- Buhake has created a branch with code which uses Listeners in LangChain4J to add Tracing and Metrics support, <https://github.com/smallrye/smallrye-llm/tree/22-add-more-cdi-features/smallrye-llm-langchain4j-telemetry/src/main/java/io/smallrye/llm/langchain4j/telemetry> and verified it to work with OpenLiberty.
- Emmanuel is trying to do the same for WildFly and running into a couple of issues, see the thread <https://microprofile-wg.slack.com/archives/C07HLJBQ36Y/p1733760702743009>. He worked with Buhake to get it working after setting properties for OTel in Microprofile.

2nd December 2024

Video Recording here

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Phil Chung (Oracle)
- Emily Jiang (IBM)
- Jan Westerkamp (iJUG)

Minutes:

-

25th November 2024

Video Recording here

Contributors *[each attendee must add themself to minutes]*

- Emmanuel Hugonnet (Red Hat)
- Buhake Sindi (Individual)
- Arjav Desai (Oracle)
- Phil Chung (Oracle)

Minutes:

- Emmanuel mentioned that
 - HTTP transport PR for Langchain4J is still in draft mode and SPI is premature/early for integration.
 - He tried that for Apache HTTPClient as well and had to add more code than Jakarta Rest.
 - Lombok is still being used by some clients on latest 0.36.1
- Buhake mentioned that
 - Fault Tolerance was already there on the AiService interface but was not triggered. He has made changes for now to make it work with Portable Extension.

18th November 2024

Video Recording here

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Yann Blazart
- John Clingan (Red Hat)
- Buhake Sindi (Individual)

Minutes:

- Langchain4J finished the work on externalizing the HTTP transport:
<https://github.com/langchain4j/langchain4j/pull/1103/files>
 - MP AI will need to use the latest release 0.36
 - Emmanuel used Jakarta Rest with Langchain4J and it worked well locally
- Maybe use
<https://github.com/langchain4j/langchain4j/blob/fcc02ee13cdf41ad62ca83d13f0ab275a2758efc/langchain4j-core/src/main/java/dev/langchain4j/model/chat/listener/ChatModelListener.java#L10> to provide MicroProfile Telemetry integration
- Buhake is investigating Fault Tolerance integration
- RAG investigation
 - Buhake: RAG is not working locally for him. CDI injection working with Embedding Model and Embedding Store.

11th November 2024

Video Recording here

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)

Only Emily was there.

4th November 2024

Video Recording [here](#)

Contributors *[each attendee must add themself to minutes]*

- Clement Escoffier (Red Hat)
- Max Andersen (Red Hat)
- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)
- Phil Chung (Oracle)
- John Clingan (Red Hat)

Apologies *[each attendee must add themself to minutes]*

- Emily Jiang (IBM) (traveling to W-JAX)

Minutes:

- Emmanuel mentioned that the PR which added MP specific classes has been reverted with <https://github.com/smallrye/smallrye-llm/pull/68>
- Langchain4J work on externalizing the HTTP transport: <https://github.com/langchain4j/langchain4j/pull/1103/files>
- Maybe use <https://github.com/langchain4j/langchain4j/blob/fcc02ee13cdf41ad62ca83d13f0ab275a2758efc/langchain4j-core/src/main/java/dev/langchain4j/model/chat/listener/ChatModelListener.java#L10> to provide OpenTelemetry integration

28th October 2024

Video Recording [here](#)

Contributors *[each attendee must add themself to minutes]*

- <https://github.com/langchain4j/langchain4j/pull/1103/files> Emily Jiang (IBM)
- Buhake Sindi (Individual)
- Clement Escoffier (Red Hat)

- Max Andersen (Red Hat)
- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)

Minutes:

- Adopt LangChain4J APIs vs new MP AI APIs
 - Buhake: LangChain has many dependencies such as Lombok. If LangChain4J has big changes, MP AI can be protected from these changes. Not sure about the future of LangChain4J. Also integrate with MP Telemetry without impacting LangChain4J. Phase 1: created some facades on top of LangChain4J.
 - Clement: You will have to fork 64 APIs. you will be out of date very fast. What is in the roadmap of LangChain4J v1 (to be released at the end of this year) is the httpclient apis. What about the sdk providing its own clients?
 - Max: work with LangChain4J to detach Lombok from the APIs. What happens between September and October? Creating new APIs caused more confusion on the choices of APIs. Maybe wait for a while to standardize something. Quarkus did some prototype with LangChain4J and contributed some functionalities to LangChain4J.
 - Arjav: If we focus on defining the integration with MP Rest, we showcase LangChain4J.
- Rest apis for LLMs
 - Should we influence the LLMs to use the same pattern?
- For the record: @Tool, @V, @P, @ToolsMemoryId are getting updated in the 1.0 release of Langchain4J (end of 2024)

14th October 2024

Video Recording [here](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Phil Chung (Oracle)

- Buhake Sindi (Individual)
- Yann Blazart (individual)
- Emmanuel Hugonnet (Red Hat)
- Emerson Castaneda
- Arjav Desai (Oracle)
- Antoine Sabot-Durand (individual)

Minutes:

Yann and Buhake gave us an update regarding their work about extraction of the API. There seems to be a misunderstanding about the MP AI goal.

Langchain4j brings a lot of limitations : JDK8, lots of bad dependencies. It was often said in this meeting that it shouldn't be a final goal for an implementation. Obviously this information was not shared or understood.

The idea followed by implementers : Yann, Emmanuel, Buhake was to start with Langchain4j to have a first result but in a second time start working on another solution. Taking inspiration from Spring AI for instance...

7th October 2024

Video Recording [here](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Arjav Desai (Oracle)
- Phil Chung (Oracle)
- John Clingan (Red Hat)
- Buhake Sindi (Individual)
- Yann Blazart (individual)

Minutes:

- Next step: since we are happy with LangChain4J, we won't invest in creating new alternative APIs to LangChain4J. We will be focusing on integrating with LangChain4J.
 - Provide generic SPIs to work with different app runtimes
 - Provide integration with MP Rest Client etc to plug in other Jakarta EE and MicroProfile technologies.
 - ~~Provide facade APIs to work with LangChain4J:~~
 - Help with LangChain4J to improve the metrics etc.

- The next piece will be the RAG integration with LangChain4J. Issue [here](#)
 - Embedding support
 - In-memory vector store

30th September 2024

Video Recording [here](#)

Contributors *[each attendee must add themselves to minutes]*

- Emmanuel Hugonnet (Red Hat)
- Arjav Desai (Oracle)
- Buhake Sindi (Individual)
- Antonio Goncalves (Individual)
- Phil Chung (Oracle)
- Jan Westerkamp (iJUG)
- John Clingan (Red Hat)
- Max Rydahl Andersen (Red Hat)
- Clement Escoffier (Red Hat)

Minutes:

- Spring AI : <https://github.com/spring-projects/spring-ai>
- Spring support in LangChain4j : <https://github.com/langchain4j/langchain4j-spring>
- Current status of SmallyRye LLM:
 - All examples were merged to the main branch.
 - Implementation for the current phase are complete.
 - Should we consider release as the first version of alpha release?
- Overall agreement that reimplementing a new general GenAI library would be wasted effort - working in langchain4j to have SPI's to enable isolation from vendor specific clients would be good for the broader java system and in smallrye-llm we could focus on extracting a higher level genAI spec (i.e. CDI based beans)
- About lombok
 - Does not prevent native compilation
 - It's something that langchain4j wants to remove in the near future
- HTTP SPI in langchain4j
 - It's something that should be implemented in Langchain4j to allow plug in different HTTP client, including the MP Rest client

23rd September 2024

Video Recording [here](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Emmanuel Hugonnet (Red Hat)
- Arjav Desai (Oracle)
- Buhake Sindi (Individual)
- Phil Chung (Oracle)
- Jan Westerkamp (iJUG)
- John Clingan (Red Hat)

Minutes:

- Native compilation
 - LangChain4J uses Lombok for annotation processing. For native compilation, a maven plugin is needed for build-time compilation.
 - There might be some special process needed for this. Not sure...
 - JDK 23 has the annotation processing disabled by default. Have to enable the lombok specially for the later Java.
- Getting samples to work on Open Liberty
 - Buhake is working on the sample. At the moment, he is debugging a classloading issue.
- [Jan] Links to AI related Apache Camel blog posts:
 - <https://camel.apache.org/blog/2024/09/data-extraction-example/>
 - <https://camel.apache.org/blog/2024/07/data-extraction-first-experiment/>
 - <https://camel.apache.org/blog/2024/09/chunking-dsl/>
 - <https://camel.apache.org/blog/2024/09/camel-ai-examples/>

16th September 2024

Video Recording [here](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily Jiang (IBM)
- Yann Blazart
- Emmanuel Hugonnet (Red Hat)
- Don Bourne (IBM)

- Arjav Desai (Oracle)
- Emerson Castaneda
- John Clingan (Red Hat)
- Buhake Sindi (Individual)
- Phil Chung (Oracle)

Minutes:

1. Yann, Antoine and Emmanuel are working together on the new proposal on the new AI api/spi. Hopefully in the next 2 or 3 weeks, the draft will be available.
 - a. We began with Langchain4J. Due to some technical issues on Langchain4J (on JDK 8, dependency models), it might be a good idea to revisit OpenAI rest api and Spring AI.
 - b. Only target BCE for Quarkus, the other app servers can use portal extensions to provide the integration.
 - c. Langchain4j is not good with native compilation.
 - d. Yann to find out whether Spring AI copes with native compilation.
2. Yann and Emmanuel work on a defect on the CDI extension
3. Emmanuel added the support from Wildfly
4. Buhake to bring the Open Liberty support from a branch to the main
 - a. Emily to work with Buhake on the issues Buhake is facing

9th September 2024

Video Recording [here](#)

Contributors *[each attendee must add themself to minutes]*

- Arjav Desai (Oracle)
- Emily Jiang (IBM)
- Buhake Sindi (Individual)
- Emmanuel Hugonnet (Red Hat)
- Antoine Sabot-Durand

Minutes:

5. Discussed the plan of creating some new APIs
6. Buhake to bring the Open Liberty support from a branch to the main

7. Emmanuel to add the support from Wildfly
8. Discussed the new step for the initiative - produce a creation plan for MP AI

19th August 2024

Video Recording [here](#)

Contributors *[each attendee must add themself to minutes]*

- Emmanuel Hugonnet (Red Hat)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)
- Emily Jiang (IBM)
- Yann Blazart (individual)
- Majid Mostafavi
- Buhake Sindi (Individual)

Current work in

Minutes:

9. Buhake committed some code changes to smallrye-llm on BCE (<https://github.com/sinditech/sindi-langchain4j>). He tested on Open Liberty.
10. Yann - BCE work on Quarkus
11. Buhake to test his changes on Wildfly and/or Payara. He then propose a PR
12. Yann to think about creating Integration tests.

12th August 2024

Video Recording (not available)

Contributors *[each attendee must add themself to minutes]*

- Buhake Sindi (Individual)
- John Clingan (Red Hat)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)
- Emmanuel Hugonnet (Red Hat)

Minutes:

Buhake asked about an issue with Open Liberty and CDI with the current Extension. John suggested raising this as a query in either smallrye github discussion or microprofile google groups.

5th August 2024

Video Recording (not available)

Contributors *[each attendee must add themself to minutes]*

- Buhake Sindi (Individual)
- John Clingan (Red Hat)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)

Minutes:

We closed out the call in 5 mins as there weren't enough participants.

29th July 2024

[Video](#) Recording

Contributors *[each attendee must add themself to minutes]*

- Yann Blazart
- Buhake Sindi (Individual)
- John Clingan (Red Hat)
- Phil Chung (Oracle)
- Majid Mostafavi (IranianJUG)
- Arjav Desai (Oracle)

Minutes:

22th July 2024

Video Recording (not available)

Contributors *[each attendee must add themself to minutes]*

- Yann Blazart
- Buhake Sindi (Individual)
- Emmanuel Hugonnet (Red Hat)
- Phil Chung (Oracle)
- Majid Mostafavi (IranianJUG)
- Arjav Desai (Oracle)

Minutes:

Yann had a conversation with Emmanuel with the possibility of integrating Wildfly config with Microprofile config. Portability of Wildfly, Quarkus and Jakarta EE profile.

Yann proposed to take the direction to take care of Jakarta EE , microprofile and Quarkus Sides. We can have a core module, then a build compatible extension, a portable one, and a SPI to provide config retrieving capabilities when some servers are not able to use mp-config.

It would be also nice to add some examples with Open Liberty and WildFly in the example module.

15th July 2024

[Video Recording](#)

Contributors *[each attendee must add themself to minutes]*

- Yann Blazart
- Buhake Sindi (Individual)
- Emmanuel Hugonnet (Red Hat)

- Phil Chung (Oracle)
- Antoine Sabot-Durand
- Arjav Desai (Oracle)

Minutes:

Yann start the call be describing his PR on SmallRye LLM :

<https://github.com/smallrye/smallrye-llm/pull/8>

Discussion about bean resolution in a dynamic lookup. We can certainly find another solution here (qualifier, list of bean types etc...)

The extension was created with the Build Time Compatible Extension to make sure that the code is working on Quarkus

8th July 2024

[Video recording.](#)

Contributors *[each attendee must add themself to minutes]*

- Emily jiang (IBM)
- Buhake Sindi (Individual)
- Emmanuel Hugonnet (Red Hat)
- Phil Chung (Oracle)
- Majid Mostafavi (IranianJUG)

Minutes:

- LangChain4J
 - Emmanuel did something similar against Wildfly. It can inject embeddings, LLM and vectorstore. Producer for LLM, vector database, embedding.
 - <https://github.com/ehsavoie/ai-testing/blob/main/wildfly-ai/injection/src/main/java/org/wildfly/extension/ai/injection/AiCDIExtension.java>
 - <https://github.com/wildfly/wildfly.org/pull/627>
 - Global producers? One producer per type; builder pattern
 - Put a PR on [smallrye-llm](#) to demonstrate the above idea
 - Embedding APIs? RAG pattern
 - Use CDI extension to create a producer for RAG integration
- Do we need to create our own APIs or wrap LangChain4J
 - <https://github.com/smallrye/smallrye-llm/issues/3>

1st July 2024

[Video recording.](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily jiang (IBM)
- Arjav Desai (Oracle)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)
- John Clingan (Red Hat)
- out (Individual)
- Emmanuel Hugonnet (Red Hat)

Minutes:

- LangChain4J
 - Antonio talked with LangChain4J community and they are happy for us to add a repo under their org.
 - We need to investigate further on the PoC regarding which CDI beans to produce.
 - We can either keep the code on smallrye-llm or contribute to LangChain4J
 - Produce more code to plug in more LLMs, more examples, more docs.
 - Antonio plans to add more LLMs integration.
 - Emmanuel did something similar against Wildfly. It can inject embeddings, LLM and vectorstore. Producer for LLM, vector database, embedding.
 - <https://github.com/ehsavoie/ai-testing/blob/main/wildfly-ai/injection/src/main/java/org/wildfly/extension/ai/injection/AiCDIExtension.java>
 - <https://github.com/wildfly/wildfly.org/pull/627>
 - Global producers? One producer per type
- We now have a GitHub repo: <https://github.com/smallrye/smallrye-llm>
 - Embedding APIs? RAG pattern
 - Use CDI extension to create a producer for RAG integration
 -

- Do we need to create our own APIs or wrap LangChain4J
 - <https://github.com/smallrye/smallrye-llm/issues/3>

24th June 2024

Video recording.

Contributors *[each attendee must add themselves to minutes]*

- Emily jiang (IBM)
- Arjav Desai (Oracle)
- Jan Westerkamp (iJUG)
- Emmanuel Hugonnet (AlpesJUG/Red Hat)
- Phil Chung (Oracle)
- Buhake Sindi (Individual)

Minutes:

- We now have a GitHub repo: <https://github.com/smallrye/smallrye-llm>
 - Embedding APIs? RAG pattern
 - Use CDI extension to create a producer for RAG integration
 -
- Do we need to create our own APIs or wrap LangChain4J
 - <https://github.com/smallrye/smallrye-llm/issues/3>

10th June 2024

[Video recording.](#)

Contributors *[each attendee must add themselves to minutes]*

- Emily jiang (IBM)
- Antonio Goncalves (Individual)
- Buhake Sindi (Individual)
- Phil Chung (Oracle)
- John Clingan (Red Hat)
- Arjav Desai (Oracle)
- JB Roux (Worldline)
- Emerson Castaneda

Minutes:

- We now have a GitHub repo: <https://github.com/smallrye/smallrye-llm>
- Do we need to create our own APIs or wrap LangChain4J
 - <https://github.com/smallrye/smallrye-llm/issues/3>
- [Antoine] at the github repo, Jann experimented.
- LangChain4j Spring Boot Starter repo: <https://github.com/langchain4j/langchain4j-spring>
- Spring exposes a few beans in <https://github.com/langchain4j/langchain4j-spring/tree/main/langchain4j-azure-open-ai-spring-boot-starter/src/main/java/dev/langchain4j/azure/openai/spring>
- We could have a CDI repo under the LangChain4j
- First we should do a PoC with Microprofile Config and CDI
- Yann has a draft PR
- Once merge and experiment, Antonio could talk to the LangChain4j team about having the GitHub repo

3rd June 2024

Video recording.

Contributors *[each attendee must add themselves to minutes]*

- Emily jiang (IBM)
- Antonio Goncalves (Individual)
- Buhake Sindi (Individual)
- Phil Chung (Oracle)
- John Clingan (Red Hat)
- Arjav Desai (Oracle)
- JB Roux (Worldline)

Minutes:

- We now have a GitHub repo: <https://github.com/smallrye/smallrye-llm>

- Do we need to create our own APIs or wrap LangChain4J
 - <https://github.com/smallrye/smallrye-llm/issues/3>
 - LangChain4J has around 30 separate modules. Taking core and adopt different LLMs (we could start with OpenAI).
 -

27th May 2024

Video recording.

Contributors *[each attendee must add themself to minutes]*

- Jan Westerkamp (iJUG)
- Antonio Goncalves (Individual)
- JB Roux (Worldline)
- Buhake Sindi (Individual)
- Emerson Castaneda

Minutes:

- [Yann] is showing an implementation example

13th May 2024

Video recording.

Contributors *[each attendee must add themself to minutes]*

- Antoine Sabot-Durand
- Clement Escoffier (Red Hat)
- Arjav Desai (Oracle)
- Emily Jiang (IBM)
- Phil Chung (Oracle)
- Jan Westerkamp (iJUG)
- Antonio Goncalves (Individual)
- Jean-Francois JAMES (Worldline)

Minutes:

Emily propose a SmallRye LLM project. Clement agreed. Antoine Said that the scope will be to deliver a framework (maybe a LangChain4j wrapper, maybe a REST client.... still not defined yet) this year and think about as specification from this framework in 2025

- Repository: <https://github.com/smallrye/smallrye-llm>
- GitHub contributors: *agoncal, antoinestd, jefrajames, ricken07, yblazart* (*italic means invitation sent*)
- OpenAI4j is a REST Client for OpenAI (it has been developed by the LangChain4j team as an external library <https://github.com/ai-for-java/openai4j>)
- LangChain4j has an SPI to connect to OpenAI, so it's easy to replace OpenAI4j with other things
- What do we implement?
 - Chat Model ? Language Model ? Image Model ? Moderation Model ? Embedding Model ?
 - Chat / Image / Video
 - Have the framework deal with the image / video metadata (there's not much in Java representing an image with Metadata)
 - We could start by generating the Java code from the OpenAI OpenAPI contract so it uses REST Client from MP

06th May 2024

Video recording [here](#).

Contributors [each attendee must add themself to minutes]

- Antoine Sabot-Durand (individual)
- Jan Westerkamp (iJUG)
- Phil Chung (Oracle)
- Arjav Desai (Oracle)
- Don Bourne (IBM)
- Buhake Sindi (Individual)
- Clement Escoffier (Red Hat)
- Scott Stark (Red Hat)
- Ricken Bazolo

Minutes:

Clément explained that most LLM providers are roughly align Open AI API.

A starting point for our vendor neutral MP AI framework could be the implementation of a Rest client able to consume LLM in an abstract approach (Open AI API is a common language at 80% level)

LLM are moving too fast to go for a specification right now. We should wait for 1 year or so to go for it.

- The goal is to find a common model to consume mot LLM around (see hugging face to have an idea of the numbers of available LLMS) :
<https://huggingface.co/models?other=function+calling.text-generation-inference&sort=trending>
- **Jan** and **Clement** agreed on thinking about a dialect concept (like in JPA) to support this diversity
- Ricken proposed to build an abstraction above Open AI. A kind of driver
- Decision was made to start with a RestClient to consume Open AI
- Jan suggested that the name of the initiative should be renamed LLM, Clement agreed

29th April 2024

Video recording [here](#).

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Antonio Goncalves (Individual)
- Phil Chung (Oracle)
- Antoine Sabot-Durand(individual)
- John Clingan (Red Hat)
- Clement Escoffier (Red Hat)
- Yann Blazart(Individual)
- Arjav Desai (Oracle)
- Buhake Sindi (Individual)
- JB Roux (Worldline)
- Ricken BAZOLO (Individual)

Minutes:

Clément explains the way LangChain4j Quarkus extension is working:

In November 2023 they started building a wrapper around LC4j. But a wrapper wasn't enough (security propagation, tracing, metrics, etc.). Had 6/7 backward breaks due to LC4J movement. There are more and more users for LC4J. How to make things better for the users, dev mode, etc? Passing emoji or videos was problematic (because streaming does not work well with reactive frameworks like Quarkus). They have difficulties in integrating the Google model (Gemini, VertexAI)

- Quarkus LangChain4J challenges
 - Integration into its reactive network stack
 - Antonio: How different (simpler?) would it be if there weren't a reactive stack?
Clement: Not much different. The (Quarkus) API is still blocking. It's not really the reactive stack per se, it's everything else that comes into play (ecosystem integration)
 - Google's model - Integration with rest of stack (metrics for example)
 - Rapid change of LangChain4J
 - Expect this space to be completely different a year from now
 - Rewrite the prompts due to the LLM changes
 - Antonio: The Quarkus LangChain4J wrapper is more complex than expected (thought it would be a few classes, but it is "hundreds"). Clement noted that it's not really *that* complex. Need to wrap models and vector stores, etc.

22nd April 2024

Video recording [here](#).

Contributors *[each attendee must add themselves to minutes]*

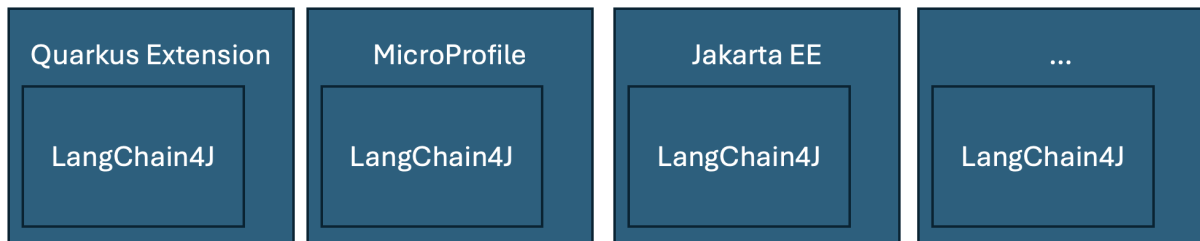
- Antoine Sabot-Durand (individual)
- Antonio Goncalves (Individual)
- Emily Jiang (IBM)
- Luís Neto (Payara)
- Yann Blazart
- Pierre Cesmat
- John Clingan (Red Hat)
- Buhake Sindi (Individual)
- Jean-François James (Worldline)
- JB Roux (Worldline)
- Majid Mostafavi(IJUG)

Minutes:

- Devovx France experience:

- Start a impl framework similar to Quarkus Langchain4J extension
 - Option 1: Ask Quarkus to get the langchain4J integrate and make it vendor neutral, e.g. making it to smallrye-ai
 - John to follow up
 - Option 2: Develop our own LangChain4j integration with MP
 - Option 3: Develop our own implementation (but will end up being very similar to LangChain4j)
-

Option 1



15th April 2024

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Luís Neto (Payara)
- Antonio Goncalves (Individual)
- Antoine Sabot-Durand
- Don Bourne (IBM)
- Marie-Alice Blete
- Emerson Castaneda
- Phil Chung (Oracle)
- JB Roux (Worldline)
- JF James (Worldline)
- Buhake Sindi

Minutes:

- Proposal: [here](#)
- Sandbox: [here](#)
 - Create annotation APIs together with specs - Anotine and Jean-Francios, Ricken
 - TCKs - Buhake

- LangChain4J is related to LangChain (the Python project) so some terminology/architecture might not be totally “Java compliant” (eg. [LangChain4J uses the term “AIMessage”](#) while most of the AI tools would use the term “AssistantMessage” but “AIMessage” is used in the Python version)
- Do we want to add our own MP API on top of LangChain4J, or just expose the LangChain4J interface?
 - LangChain4J isn’t the cleanest API, but it may be early in maturity terms to build one API that interfaces consistently with all LLMs
- Have a look at the Quarkus approach to see if there are things we’d want to do to improve

8 April 2024

Contributors *[each attendee must add themselves to minutes]*

- Antoine Sabot-Durand (SCIAM)
- Antonio Goncalves (Individual)
- Don Bourne (IBM)
- JF James (Worldline)

Minutes:

We started by sharing the existing doc and discussion.

Doc opened 2 weeks ago had a lot of contribution:

https://docs.google.com/document/d/1Nnj3LHJPEMNF_nQDXnyAO_A-0MMOZ73CA9ONHAA5DCQ/edit?usp=sharing

Discussion about the target for V1:

- Antonio mention that RAG is moving too fast for V1
- What about having something very basic like EasyRag for a V1
- MP has a core specification and then standalone specs. MicroProfile AI should be in the Standalone space. They can be backward incompatible
- We agreed on the difficulty of TCK
- Consider use of [rouge scores](#) for TCK for RAG implementations
-

25th March 2024

Meeting was canceled because only 2 people showed.

But Ricken and Antoine started to work on draft regarding what the goal of the spec could be.
The document is here : [MicroProfile AI Working Doc](#)

To summarize, our main idea is to have an API that abstract LLM consuming and that would help building RAG.

More to discuss next week.

18th March. 2024

Video: [here](#)

Contributors *[each attendee must add themself to minutes]*

- Emily Jiang (IBM)
- Ivar Grimstad (Eclipse Foundation)
- Emerson Castaneda
- Antoine Sabot-Durand (SCIAM)
- Ricken Bazolo (SCIAM)
- Buhake Sindi

Agenda:

Context: more info [here](#)

In the recent MicroProfile Live Hangout (minutes [here](#)), Reza brought up the idea of embracing OpenAI libraries in MicroProfile and the idea was welcomed and discussed briefly. At the call, we decided to have some

offline discussion on this mailinglist to accommodate the ones who could not make the calls and then further discuss at next week's technical call.

As you may know, recently, Spring has provided [Spring AI](#) to create AI applications (see code [here](#)) and it seems that a lot of attention was drawn. The other project [langchain4j](#) is also very popular, which is to simplify the integration of AI/LLM capabilities. Spring has also provided an [example](#) of working with Langchain4j.

I have a few ideas:

1. Provide MicroProfile AI programming model to develop AI applications
2. Provide some APIs if needed to connect with LLMs, which might use Langchain4j
3. Integrate with some other popular OpenAI libraries Context

- Spring has more interactions with LLMs.
- Ricken: Mistral IE. Provide a simple way to add architecture. Provide lib. Start transformer; use vector database to bring forward LLMs to Java apps such as Postgres to perform RAG
 - Antoine and Ricken work on the similar projects.
- Buhake:
 - Langchain4J - integrated abstract for LLMs
 - Unified model for different LLMs
 - New version of LLMs have changes surfacing up to the app
 - Developers only need to work with interfaces and runtime handles the code changes.
- Next action:
 - <https://github.com/eclipse/microprofile-sandbox/tree/main/proposals>
 - Ricken and Antoine to start a proposal here
 - Buhake to start a proposal
 - <https://www.eclipse.org/legal/ECA.php> ECA to sign
 - Next week, we will bootstrap the proposals