🧪 The prompt lab kick-off

Bee

Aleš Kalfas
October 25, 2024 v2
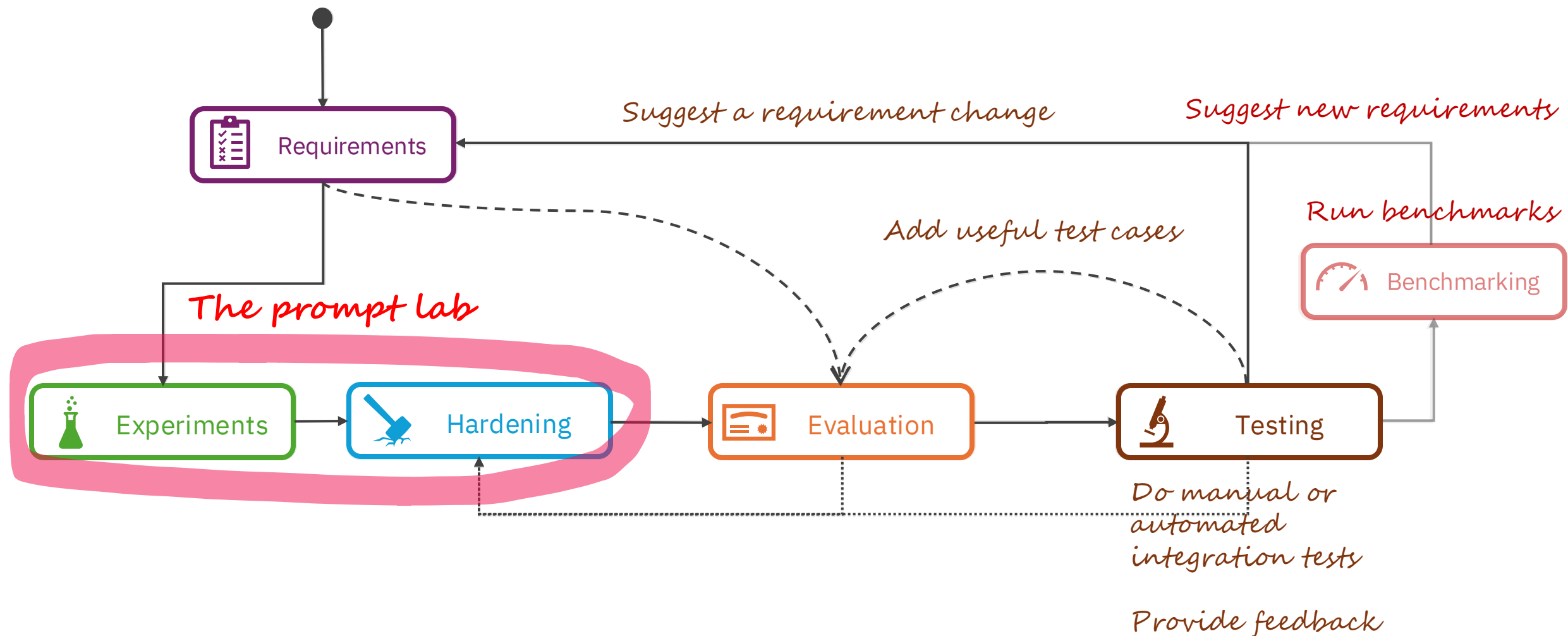
🧪 The prompt lab *context*
The optimization process

🧪 The prompt lab *context*
The optimization process

🧪 The prompt lab *concept*

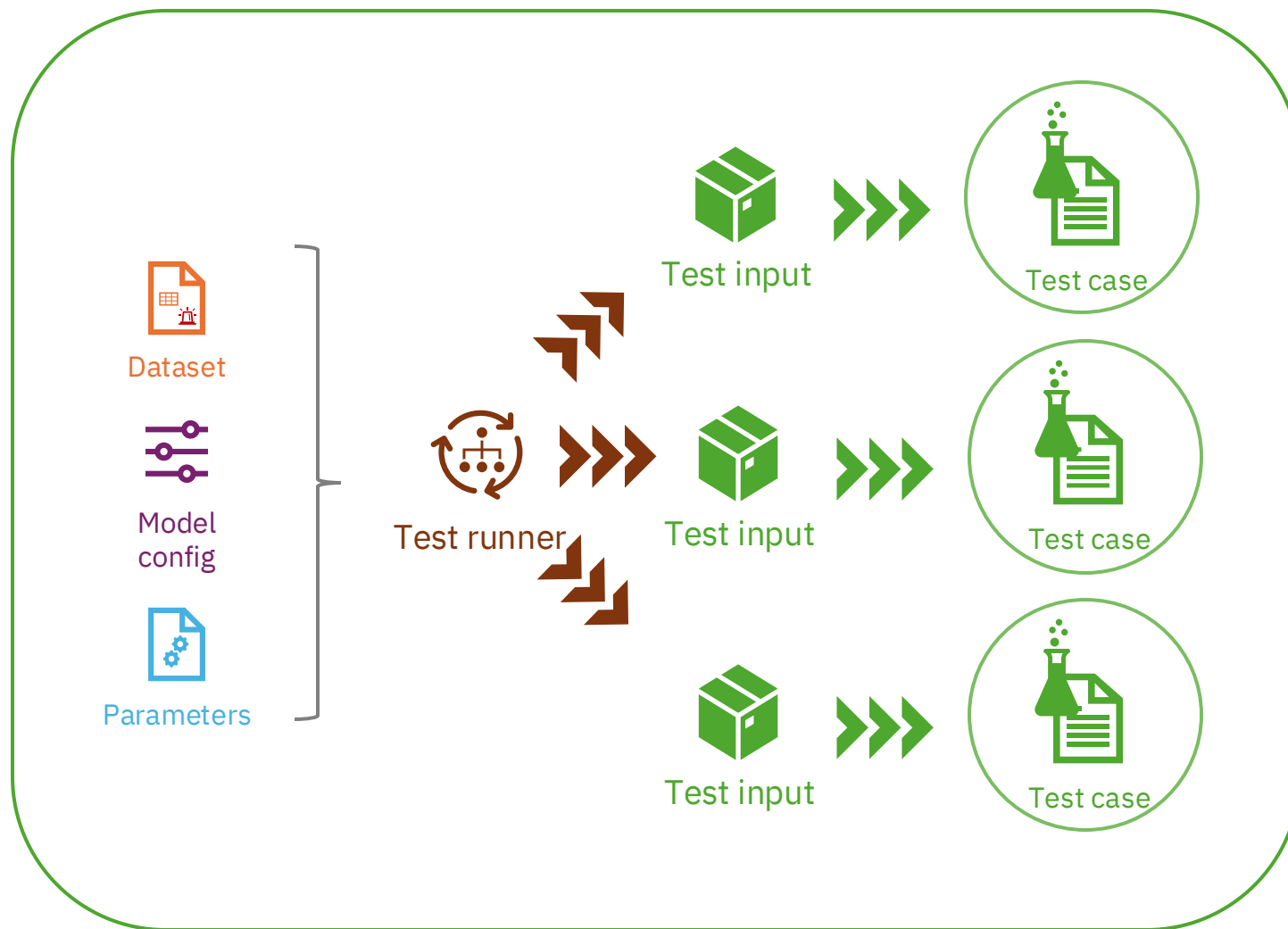Test

🧪 The prompt lab *concept*
Test

Test

It is a code that takes *>dataset*, *>model* config, *>parameters, >test case* and runs them in a test runner.

Test

🧪 The prompt lab *usage*

Configuration

🧪 The prompt lab *usage*
Configuration

Why configure?

The bee project already contains a bunch of tests that amount is multiplied by selected configurations therefore is important to setup them correctly.

The configuration is located inside the bee project **src/projects/bee/base/config.ts**.

```
∨ src
  ∨ projects / bee
    > base
    ∨ tests
      > artifacts / creation
      > awareness
      > base
      > behaviour-control
      > capabilities
      > communication-protocol
      > question-decomposition
  > test-framework
```

There are already around **1,000 tests**!

Running them on a single model with minimalistic configuration takes approximately **4 hours!**

Area for Improvement

🧪 The prompt lab *usage*
Configuration

```
28
29   // ----------------------------------------------------------------
30   // MODELS
31   // ----------------------------------------------------------------
32   export const SUPPORTED_MODELS_CONFIGS = fromObject({
33     "granite-8b": {
34       ...Granite,
35       name: "ibm/granite-8b-instruct-preview-4k",
36       adapter: BAMAdapter,
37     } satisfies ModelConfig<BAMModelConfigTypes>,
38     // "llama-3-1-70b": {
39     //   ...LLama,
40     //   name: "meta-llama/llama-3-1-70b-instruct",
41     //   adapter: IBMVLLMAdapter,
42   💡// },
43     "llama-3-1-70b-new": {        Ales Kalfas, 2 weeks ago • feat: unified test approaches (#8) …
44       ...LLamaNew,
45       name: "meta-llama/llama-3-1-70b-instruct",
46       adapter: IBMVLLMAdapter,
47     } satisfies ModelConfig<IBMVLLMModelConfigTypes>,
48     // "llama-3-1-8b-new": {
49     //   ...OllamaLLamaNew,
50     //   name: "llama3.1:8b",
51     //   adapter: OllamaAdapter,
52     // } satisfies ModelConfig<OllamaModelConfigTypes>,
53   });
54
```

Models

- Currently supported adapters include BAM, IBM vLLM, and Ollama.
- These adapters connect directly to the bee-agent-framework adapters, though this is a subset.
- It's possible to extend this set by adding other framework adapters or even entirely different ones.

🧪 The prompt lab *usage*
Configuration

Toolkits

You can configure a custom combination of tools as needed.

```
54
55    // ----------------------------------------------------------
56    // TOOLS
57    // ----------------------------------------------------------
58    export const BASE_TOOLKIT = [duckDuckGo, wikipedia, pythonTool];
59    export const EXTENDED_TOOLKIT = [...BASE_TOOLKIT, openMeteo, arXiv];
60    export const STANDARD_TOOLS_COMBINATIONS = [
61      [],
62      BASE_TOOLKIT,
63      // EXTENDED_TOOLKIT,
64    ];
65
```

```
65
66    // ----------------------------------------------------------
67    // SWITCHES
68    // ----------------------------------------------------------
69    export const SWITCHES = _combine({
70      // 🏛 Use guided parameter in LLM call if it's present in the model config parameters.
71      useGuided: [true, false],
72      // ⏱ Set created at date time to LLM messages metadata, but it's up prompt templates if they use it.
73      useTime: [false],
74      // 🎨 Enable artifacts functionality.
75      useArtifacts: [true],
76    });
77
```

Switches

Switches are special controls to enable or disable various features within the tests.

useGuided Enables or disables constraint decoding on LLM calls.

useTime Enables or disables the inclusion of timestamps in user prompts.

useArtifacts Enables or disables the inclusion of artifacts in prompt parts.

🧪 The prompt lab *usage*
Configuration

```
78   // -----------------------------------------------------------------
79   // RUN CONTROL
80   // -----------------------------------------------------------------
81   export const TESTS_SUITE_RUN_CONTROL: RunControlSuiteConfig = {
82     paths: [/\/artifact-/],
83     // paths: [],
84   };
85
86   export const TESTS_RUN_CONTROL: RunControlTestConfig = {
87     level: RunControlTestLevel.CORE,
88   };
89
```

Filters

Filters limit the final test set across two dimensions:

- *Test suites* Filters tests that match specified regex patterns.
- *Test run* Filters specific tests within suites based on their level. Typically, test suites include CORE level and NORMAL level tests. CORE level tests should always be run when testing new prompt changes, while NORMAL tests should be run when modifying specific features.

🧪 The prompt lab *usage*

Run

🧪 The prompt lab *usage*
Run

Run

Run the current configuration named **my_setup** will produce output to the **./outputs/my_setup/test-output.html**

```
$ pnpm use test:report my_setup
```

Command-line output

The minimalist output version provides a approximate visual indication of test progress.

```
> bee-prompt-lab-poc@1.0.0 test:report /Users/kalfas/Documents/Projects/bee/git/bee-prompt-lab-poc-v2
> tsx src/test-framework/results/json-output/reporter/run-json-report.ts "my_setup"


 RUN  v2.1.2 /Users/kalfas/Documents/Projects/bee/git/bee-prompt-lab-poc-v2

········
(node:47970) [DEP0040] DeprecationWarning: The `punycode` module is deprecated. Please use a userland alternative instead.
·············x········x··············
```

🧪 The prompt lab *usage*

Run

Progress

Errors

Report

```
> bee-prompt-lab-poc@1.0.0 test:report /Users/kalfas/Documents/Projects/bee/git/bee-prompt-lab-poc-v2
> tsx src/test-framework/results/json-output/reporter/run-json-report.ts "my_setup"

 RUN  v2.1.2 /Users/kalfas/Documents/Projects/bee/git/bee-prompt-lab-poc-v2

........
(node:47970) [DEP0040] DeprecationWarning: The `punycode` module is deprecated. Please use a userland alternative instead.
...............x.......x.............

JSON report written to /Users/kalfas/Documents/Projects/bee/git/bee-prompt-lab-poc-v2/outputs/my_setup/test-output.json

 FAIL  src/projects/bee/base/models/llama-3-1-70b-instruct/parse.test.ts > Parser > Valid inputs and schemas for thought and function
Error: Model output couldn't be parsed: Thought: My thinking...
Function: {"name":"DuckDuckGo","parameters":{"query":"Foobar"}}
 ❯ Module.parse src/projects/bee/base/models/llama-3-1-70b-instruct/parse.ts:67:9
     65|      };
     66|    }
     67|    throw new Error(`Model output couldn't be parsed: ${input}`);
       |          ^
     68| };
     69|
 ❯ src/projects/bee/base/models/llama-3-1-70b-instruct/parse.test.ts:81:22
 ❯ src/projects/bee/base/models/llama-3-1-70b-instruct/parse.test.ts:80:13

 FAIL  src/projects/bee/base/models/llama-3-1-70b-instruct-new/parse.test.ts > Parser > Valid inputs and schemas for thought and function
Error: Model output couldn't be parsed: Thought: My thinking...
Function: {"name":"DuckDuckGo","parameters":{"query":"Foobar"}}
 ❯ Module.parse src/projects/bee/base/models/llama-3-1-70b-instruct-new/parse.ts:67:9
     65|      };
     66|    }
     67|    throw new Error(`Model output couldn't be parsed: ${input}`);
       |          ^
     68| };
     69|
 ❯ src/projects/bee/base/models/llama-3-1-70b-instruct-new/parse.test.ts:81:22
 ❯ src/projects/bee/base/models/llama-3-1-70b-instruct-new/parse.test.ts:80:13

 Test Files  2 failed | 5 passed | 39 skipped (46)
      Tests  2 failed | 34 passed | 2320 skipped (2356)
   Start at  14:05:29
   Duration  1.20s (transform 283ms, setup 13ms, collect 889ms, tests 12ms, environment 1ms, prepare 52ms)

Error running tests: Command failed: vitest --run --silent --reporter=dot --reporter=json --outputFile=outputs/my_setup/test-output.json
New temp directory created.
Test report saved to: /Users/kalfas/Documents/Projects/bee/git/bee-prompt-lab-poc-v2/outputs/my_setup/test-output.html
```

🧪 The prompt lab *usage*
Report

🧪 The prompt lab *usage*
Report

One test suite and simple setting

Test suite	Test	Parameters				Models		
		Switches			Tools	llama-3-1-70b-new		
		useGuided	useTime	useArtifacts	ReadFile	default		
		➖	➖	➖	➖	test-output.json		
						Error	Duration	Passed ➖
▼ **Novels Extraction** (16 items)								
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Authors | Extract content…	☑	☑	☐	☑		00:01.4	✔
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Characters | Extract cont…	☑	☑	☐	☑	ASSERT	00:02.9	✖
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Locations | Extract conte…	☑	☑	☐	☑	PARSER	00:13.2	✖
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Story | Extract content d…	☑	☑	☐	☑	PARSER	00:02.4	✖
Novels Extraction	[Multi] Bond_of_Fate.txt | Authors | Extract content data | ReadFil…	☑	☑	☐	☑	PARSER	00:01.8	✖
Novels Extraction	[Multi] Bond_of_Fate.txt | Characters | Extract content data | Read…	☑	☑	☐	☑		00:06.3	✔
Novels Extraction	[Multi] Bond_of_Fate.txt | Locations | Extract content data | ReadF…	☑	☑	☐	☑	PARSER	00:02.4	✖
Novels Extraction	[Multi] Bond_of_Fate.txt | Story | Extract content data | ReadFile …	☑	☑	☐	☑	PARSER	00:02.2	✖
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Authors | Extract conten…	☑	☑	☐	☑		00:01.5	✔
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Characters | Extract con…	☑	☑	☐	☑	ASSERT	00:04.8	✖
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Locations | Extract cont…	☑	☑	☐	☑		00:02.3	✔
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Story | Extract content d…	☑	☑	☐	☑		00:01.9	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Authors | Extract content data | ReadFi…	☑	☑	☐	☑		00:03.2	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Characters | Extract content data | Read…	☑	☑	☐	☑		00:06.2	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Locations | Extract content data | Read…	☑	☑	☐	☑		00:03.2	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Story | Extract content data | ReadFile …	☑	☑	☐	☑	PARSER	00:03.2	✖
						8	00:58.9	50 %

🧪 The prompt lab *usage*
Report

One test suite and simple setting

Switches

Tools

Models/Files /profiles

Parameters

Test suites   *Tests*

Test suite	Test	Parameters			Tools	Models		
		Switches			Tools	llama-3-1-70b-new		
		useGuided	useTime	useArtifacts	ReadFile	default		
		▬	▬	▬	▬	test-output.json		
						Error	Duration	Passed
								▬
▼ Novels Extraction (16 items)								
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Authors | Extract content…	☑	☑	☐	☑		00:01.4	✔
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Characters | Extract cont…	☑	☑	☐	☑	ASSERT	00:02.9	✘
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Locations | Extract conte…	☑	☑	☐	☑	PARSER	00:13.2	✘
Novels Extraction	[Multi] A_Tale_of_Unlikely_Friends.txt | Story | Extract content d…	☑	☑	☐	☑	PARSER	00:02.4	✘
Novels Extraction	[Multi] Bond_of_Fate.txt | Authors | Extract content data | ReadFil…	☑	☑	☐	☑	PARSER	00:01.8	✘
Novels Extraction	[Multi] Bond_of_Fate.txt | Characters | Extract content data | Read…	☑	☑	☐	☑		00:06.3	✔
Novels Extraction	[Multi] Bond_of_Fate.txt | Locations | Extract content data | ReadF…	☑	☑	☐	☑	PARSER	00:02.4	✘
Novels Extraction	[Multi] Bond_of_Fate.txt | Story | Extract content data | ReadFile …	☑	☑	☐	☑	PARSER	00:02.2	✘
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Authors | Extract conten…	☑	☑	☐	☑		00:01.5	✔
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Characters | Extract con…	☑	☑	☐	☑	ASSERT	00:04.8	✘
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Locations | Extract cont…	☑	☑	☐	☑		00:02.3	✔
Novels Extraction	[Single] A_Tale_of_Unlikely_Friends.txt | Story | Extract content d…	☑	☑	☐	☑		00:01.9	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Authors | Extract content data | ReadFi…	☑	☑	☐	☑		00:03.2	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Characters | Extract content data | Read…	☑	☑	☐	☑		00:06.2	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Locations | Extract content data | Read…	☑	☑	☐	☑		00:03.2	✔
Novels Extraction	[Single] Bond_of_Fate.txt | Story | Extract content data | ReadFile …	☑	☑	☐	☑	PARSER	00:03.2	✘
						8	00:58.9	50 %

🧪 The prompt lab *usage*
Report

Many test suites and complex config

Test suite ▲	Test ▲	Parameters										Models
		Switches		Tools								llama-3-1-70b-new
		use… ▲ ⊟	use… ▲ ⊟	get… ▲ ⊟	Du… ▲ ⊟	Wi… ▲ ⊟	Pyt… ▲ ⊟	Op… ▲ ⊟	Re… ▲ ⊟	ZblorgColorLookup ▲ ⊟	UserCoordinates ▲ ⊟	default
												test-output-basic.json
												Error ▲ | Du… ▲ | Passed ▲ ⊟
▶ **10k Filings Extraction** (10 items)												
												5 | 02:51.6 | 50 %
▶ **Adhere User Instructions** (1 item)												
												1 | 00:01.8 | 0 %
▶ **Chitchat** (128 items)												
												10 | 03:51.7 | 92 %
▶ **Create Outline Inline** (40 items)												
												0 | 08:03.3 | 100 %
▶ **Custom Tools** (12 items)												
												12 | 00:13.8 | 0 %
▶ **Duckduckgo** (140 items)												
												140 | 04:46.3 | 0 %
▶ **Duckduckgo Vs Wikipedia** (144 items)												
												144 | 11:59.6 | 0 %
▶ **General Knowledge** (8 items)												
												3 | 00:17.5 | 63 %
▶ **Novels Extraction** (8 items)												
												2 | 00:24.4 | 75 %
▶ **Openmeteo** (104 items)												
												104 | 03:11.6 | 0 %

🧪 The prompt lab *usage*
Report

Many test suites and complex config

												140	04:46.3	0 %
▼ **Duckduckgo Vs Wikipedia** (144 items)														
Duckduckgo Vs Wikipedia	antoinette	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ASSERT	00:03.3	✕
Duckduckgo Vs Wikipedia	antoinette | DuckDuckGo Wikipedia	☐	☐	☐	☑	☑	☐	☐	☐	☐	☐	PARSER	00:01.4	✕
Duckduckgo Vs Wikipedia	antoinette | DuckDuckGo Wikipedia 🚶	☑	☐	☐	☑	☑	☐	☐	☐	☐	☐	PARSER	00:01.3	✕
Duckduckgo Vs Wikipedia	antoinette | DuckDuckGo Wikipedia P…	☐	☐	☐	☑	☑	☑	☐	☐	☐	☐	PARSER	00:01.4	✕
Duckduckgo Vs Wikipedia	antoinette | DuckDuckGo Wikipedia P…	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	PARSER	00:01.3	✕
Duckduckgo Vs Wikipedia	antoinette | DuckDuckGo Wikipedia P…	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	PARSER	00:01.3	✕
Duckduckgo Vs Wikipedia	antoinette | DuckDuckGo Wikipedia P…	☑	☐	☐	☑	☑	☑	☑	☐	☐	☐	PARSER	00:01.1	✕
Duckduckgo Vs Wikipedia	antoinette 🚶	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	ASSERT	00:04.5	✕
Duckduckgo Vs Wikipedia	antoinette mother	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ASSERT	00:01.7	✕
Duckduckgo Vs Wikipedia	antoinette mother | DuckDuckGo Wiki…	☐	☐	☐	☑	☑	☐	☐	☐	☐	☐	PARSER	00:01.2	✕
Duckduckgo Vs Wikipedia	antoinette mother | DuckDuckGo Wiki…	☑	☐	☐	☑	☑	☐	☐	☐	☐	☐	PARSER	00:01.3	✕
Duckduckgo Vs Wikipedia	antoinette mother | DuckDuckGo Wiki…	☐	☐	☐	☑	☑	☑	☐	☐	☐	☐	PARSER	00:01.8	✕
Duckduckgo Vs Wikipedia	antoinette mother | DuckDuckGo Wiki…	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	PARSER	00:01.7	✕
Duckduckgo Vs Wikipedia	antoinette mother | DuckDuckGo Wiki…	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	PARSER	00:01.8	✕
Duckduckgo Vs Wikipedia	antoinette mother | DuckDuckGo Wiki…	☑	☐	☐	☑	☑	☑	☑	☐	☐	☐	PARSER	00:01.8	✕
Duckduckgo Vs Wikipedia	antoinette mother 🚶	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	ASSERT	00:02.2	✕
Duckduckgo Vs Wikipedia	facebook founder	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	ASSERT	00:01.7	✕
Duckduckgo Vs Wikipedia	facebook founder | DuckDuckGo Wiki…	☐	☐	☐	☑	☑	☐	☐	☐	☐	☐	PARSER	00:01.2	✕
Duckduckgo Vs Wikipedia	facebook founder | DuckDuckGo Wiki…	☑	☐	☐	☑	☑	☐	☐	☐	☐	☐	PARSER	00:01.5	✕
Duckduckgo Vs Wikipedia	facebook founder | DuckDuckGo Wiki…	☐	☐	☐	☑	☑	☑	☐	☐	☐	☐	PARSER	00:01.0	✕
Duckduckgo Vs Wikipedia	facebook founder | DuckDuckGo Wiki…	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	PARSER	00:01.2	✕
Duckduckgo Vs Wikipedia	facebook founder | DuckDuckGo Wiki…	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	PARSER	00:01.2	✕

🧪 The prompt lab *usage*
Report

Multiple runs

Files

Rea...		Models																
		llama-3-1-70b-new																
		default																
		test-output-00_serialize_run.json			test-output-01_file_parallelism.json			test-output-02_threads.json			test-output-03_nevim.json			test-output-04_grpc_concurent.json			test-output-05_no_isolate.json		
		Error	Dur...	Passed	Error	Dur...	Passed	Error	Dur...	Passed	Error	Dur...	Passed	Error	Dur...	Passed	Error	Dur...	Passed
☐			00:30.0	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐			00:49.9	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐			00:37.0	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:38.1	✓
☐			00:54.8	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:04.5	✓
☐			00:33.2	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:41.6	✓
☐			00:28.5	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐			00:32.5	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:48.3	✓
☐			00:57.7	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:04.3	✓
☐			00:34.2	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.1	✗		01:41.1	✓
☐			00:39.7	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		00:38.1	✓
☐			00:59.5	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐			00:21.4	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:02.2	✓
☐			00:54.0	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		00:43.6	✓
☐			00:49.0	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:03.0	✓
☐	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐			00:24.5	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	02:00.0	✗
☐			00:20.3	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		01:02.8	✓
☐			00:56.4	✓	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗	TIMEOUT	01:00.0	✗		00:59.2	✓

🧪 The prompt lab *usage*
Report

Multiple models

Models

	Parameters								Models					
	Switches			Tools					granite-8b			llama-3-1-70b-new		
	useGui...	useTime	useArti...	DuckD...	Wikipe...	Python	OpenM...	ArXiv	default			default		
	▬	▬	▬	▬	▬	▬	▬	▬	test-output.json			test-output.json		
									Error	Duration	Passed ▬	Error	Duration	Passed ▬
DuckDuckGo Wikipedia Python ...	☑	☑	☐	☑	☑	☑	☐	☐	ASSERT	00:03.4	✖	PARSER	00:01.1	✖
DuckDuckGo Wikipedia Python ...	☑	☑	☑	☑	☑	☑	☐	☐		00:05.8	✔		00:10.1	✔
DuckDuckGo Wikipedia Python O...	☑	☑	☐	☑	☑	☑	☑	☑	ASSERT	00:05.8	✖	PARSER	00:01.2	✖
DuckDuckGo Wikipedia Python O...	☑	☑	☑	☑	☑	☑	☑	☑		00:05.6	✔	ASSERT	00:11.5	✖
ython 🏃⏱	☑	☑	☐	☐	☐	☑	☐	☐	ASSERT	00:05.6	✖	PARSER	00:01.4	✖
ython 🏃⏱🏺	☑	☑	☑	☐	☐	☑	☐	☐		00:05.8	✔		00:11.6	✔
⏱	☑	☑	☐	☐	☐	☐	☐	☐	ASSERT	00:06.4	✖	ASSERT	00:09.8	✖
⏱🏺	☑	☑	☑	☐	☐	☐	☐	☐		00:05.5	✔		00:12.0	✔
DuckGo Wikipedia Python 🏃⏱	☑	☑	☐	☑	☑	☑	☐	☐	ASSERT	00:04.9	✖	PARSER	00:01.4	✖
DuckGo Wikipedia Python 🏃...	☑	☑	☑	☑	☑	☑	☐	☐	ASSERT	00:06.2	✖		00:07.3	✔
DuckGo Wikipedia Python Open...	☑	☑	☐	☑	☑	☑	☑	☑	ASSERT	00:02.7	✖	PARSER	00:01.6	✖
DuckGo Wikipedia Python Open...	☑	☑	☑	☑	☑	☑	☑	☑	ASSERT	00:04.7	✖		00:09.8	✔
on 🏃⏱	☑	☑	☐	☐	☐	☑	☐	☐	ASSERT	00:04.4	✖	PARSER	00:01.4	✖
on 🏃⏱🏺	☑	☑	☑	☐	☐	☑	☐	☐	ASSERT	00:05.8	✖	ASSERT	00:09.5	✖
	☑	☑	☐	☐	☐	☐	☐	☐	ASSERT	00:06.3	✖	ASSERT	00:02.7	✖
🏺	☑	☑	☑	☐	☐	☐	☐	☐		00:03.7	✔		00:13.6	✔
DuckDuckGo Wikipedia Python 🏃⏱	☑	☑	☐	☑	☑	☑	☐	☐	ASSERT	00:04.5	✖	ASSERT	00:05.2	✖
DuckDuckGo Wikipedia Pyth...	☑	☑	☑	☑	☑	☑	☐	☐	ASSERT	00:04.5	✖	ASSERT	00:04.8	✖

🧪 The prompt lab *tests*
Creation

🧪 The prompt lab *usage*
Creation

```
12    export default dataset(
13      conversation(
14        turn(
15          userPrompt("Hello"),
16          callLLM(
17            "User say hello",
18            RunControlTestLevel.CORE,
19            assert({ parsed: { final_answer: { exists: true } } })
20          )
21        )
22      )
23    );
24
```

🧪 The prompt lab *usage*
Creation


```
12  export default dataset(
13    conversation(
14      turn(
15        userPrompt("Hello"),
16        callLLM(
17          "User say hello",
18          RunControlTestLevel.CORE,
19          assert({ parsed: { final_answer: { exists: true } } })
20        )
21      )
22    )
23  );
24
```

▼ **Example** **(2 items)**	
Example	User say hello ǀ getUserName 🚶 ⏱

🧪 **The prompt lab** *usage*
Creation

```
12  export default dataset(
13    conversation(
14      turn(
15        userPrompt("Hello"),
16        callLLM(
17          "User say hello",
18          RunControlTestLevel.CORE,
19          assert({ parsed: { final_answer: { exists: true } } })
20        ),
21        assistantFinalAnswer({
22          thought: "User is greeting me I have greet him too.",
23          finalAnswer: "Hello, how can I help you?",
24        })
25      ),
26      turn(
27        userPrompt("Do you know my name?"),
28        callLLM(
29          "User asks to his name",
30          RunControlTestLevel.CORE,
31          assert({
32            parsed: {
33              tool_name: { exactMatch: "getUserName" },
34            },
35          })
36        )
37      )
38    )
39  );
40
```

▼ **Example** (2 items)	
Example	User say hello	getUserName 🚶 ⏱
Example	User asks to his name	getUserName 🚶 ⏱

🧪 The prompt lab *usage*
Creation


```
bee-prompt-lab-poc-v2 > src > projects > bee > tests > example > TS example.test.ts > 🔷 <function>
 3    import { testRunner } from "@/projects/bee/base/test-runner.js";
 4    import { getTestFile } from "@/test-framework/core/index.js";
 5    import { overrideFns } from "@/test-framework/core/parameters.js";
 6    import { describe } from "vitest";
 7
 8    import dataset from "./dataset.js";
 9    import { zodToJsonSchema } from "zod-to-json-schema";
10    import { z } from "zod";
11
12    const testFile = getTestFile(import.meta);
13    describe.runIf(runControlSuite({ testFile }))(testFile.readableFileName, () =>
14      testRunner({
15        testFile,
16        dataset,
17        config: {
18          tools: overrideFns.addToEach([
19            {
20              name: "getUserName",
21              description: "Use to get user name.",
22              schema: JSON.stringify(zodToJsonSchema(z.object({}))),
23            },
24          ]),
25        },
26      })
27    );
28
```

🧪 The prompt lab *usage*
Creation / Configuration

```js
26      wikipedia,
27    } from "./tools.js";
28
29    // --------------------------------------------------
30    // MODELS
31    // --------------------------------------------------
32    export const SUPPORTED_MODELS_CONFIGS = fromObject({
33      "granite-8b": {
34        ...Granite,
35        name: "ibm/granite-8b-instruct-preview-4k",
36        adapter: BAMAdapter,
37      } satisfies ModelConfig<BAMModelConfigTypes>,
38      // "llama-3-1-70b": {
39      //    ...LLama,
40      //    name: "meta-llama/llama-3-1-70b-instruct",
41      //    adapter: IBMVLLMAdapter,
42      // },
43      "llama-3-1-70b-new": {
44        ...LLamaNew,
45        name: "meta-llama/llama-3-1-70b-instruct",
46        adapter: IBMVLLMAdapter,
47      } satisfies ModelConfig<IBMVLLMModelConfigTypes>,
48      // "llama-3-1-8b-new": {
49      //    ...OllamaLLamaNew,
50      //    name: "llama3.1:8b",
51      //    adapter: OllamaAdapter,
52      // } satisfies ModelConfig<OllamaModelConfigTypes>,
53    });
54
55    // --------------------------------------------------
56    // TOOLS
57    // --------------------------------------------------
58    export const BASE_TOOLKIT = [duckDuckGo, wikipedia, pythonTool];
59    export const EXTENDED_TOOLKIT = [...BASE_TOOLKIT, openMeteo, arXiv];
60    export const STANDARD_TOOLS_COMBINATIONS = [
61      [],
62      BASE_TOOLKIT,
63      // EXTENDED_TOOLKIT,
64    ];
65

65
66    // --------------------------------------------------
67    // SWITCHES
68    // --------------------------------------------------
69    export const SWITCHES = _combine({
70      // 🧭 Use guided parameter in LLM call if it's present in the model config parameters.
71      useGuided: [true],
72    // ⏰ Set created at date time to LLM messages metadata, but it's up prompt templates if they use it.
73      useTime: [true],        You, now • Uncommitted changes
74      // 🎭 Enable artifacts functionality.
75      useArtifacts: [false],
76    });
77
78    // --------------------------------------------------
79    // RUN CONTROL
80    // --------------------------------------------------
81    export const TESTS_SUITE_RUN_CONTROL: RunControlSuiteConfig = {
82      paths: [/\/example/],
83      // paths: [],
84    };
85
86    export const TESTS_RUN_CONTROL: RunControlTestConfig = {
87      level: RunControlTestLevel.CORE,
88    };
89
```

🧪 The prompt lab *usage*
Creation / Run

$ pnpm use test:report example

🧪 The prompt lab *usage*
Creation / Report

Test suite ▲	Test ▼	Parameters							Models	
		Switches			Tools				granite-8b	llama-3-1-70b-new
		useGui... ▲	useTime ▲	useArt... ▲	DuckD... ▲	Wikipe... ▲	Python ▲	getUse... ▲	default	default
		▬	▬	▬	▬	▬	▬	▬	test-output.json	test-output.json

									Error ▲	Duration ▲	Passed ▬	Error ▲	Duration ▲	Passed ▬
▼ **Example** (4 items)														
Example	User say hello | getUserName 🧍⏱	☑	☑	☐	☐	☐	☐	☑		00:01.5	✔		00:02.5	✔
Example	User say hello | DuckDuckGo Wikipedia Python getUserName 🧍...	☑	☑	☐	☑	☑	☑	☑		00:00.8	✔		00:01.6	✔
Example	User asks to his name | getUserName 🧍⏱	☑	☑	☐	☐	☐	☐	☑		00:01.2	✔		00:01.3	✔
Example	User asks to his name | DuckDuckGo Wikipedia Python getUserN...	☑	☑	☐	☑	☑	☑	☑		00:01.0	✔		00:01.7	✔
									0	00:04.5	100 %	0	00:07.1	100 %

The End