

Guided Agentic Workflows for Pharmacometric Modeling Using MonolixSuite



Authors: Pauline Traynard, Monika Twarogowska, Stephanie Kollmann, Chloe Bracis, Frano Mihaljevic, Géraldine Cellière
Institutions: Simulations Plus, Product & Technology division, France

SimulationsPlus

INTRODUCTION

Agentic AI can substantially accelerate pharmacometric modeling and simulation, yet its **adoption depends on scientific rigor**, transparency, and reproducibility.

We present **Composer**, the Simulations Plus platform, in which **AI-agents organize workflows and delegate computation to the MonolixSuite engine**.

EXAMPLE USE CASES

1 Generation of a Monolix-ready dataset from SDTM

Data-agent – LLM: Opus 4.8 – Cost: \$3.68

Create a PK dataset ready for Monolix using the SDTM files in the /input. We are interested in the plasma GBR12909 concentration.

There are 41 files. Let's explore the files that contain PK conc. and dosing data.

Discovery Summary:

Observations — input/EXTLABS.csv, TESTNAME='Plasma GBR Concentration' (ng/mL): 366 records across 12 subjects.

Dosing — input/INVAGT.csv: once-daily oral dose, Days 1–11 (11 doses). Dose amount from randomization (RDMCODE.csv): 50 mg (arm A1) or 75 mg (arm A2).

Time: COLLDATE (study day) + TIMETEXT (clock HHMM) → continuous hours;

BLQ: RESULT = 0.001 appears 116× as a BLQ placeholder.

Decisions I need to confirm:

BLQ handling: 0.001 is clearly a BLQ placeholder. Options:

- (a) Use LLOQ = a value you specify (do you know the assay LLOQ?)
- (b) Infer LLOQ from the lowest real quantified value

Use LLOQ = 0.1

Now I'll write the **auditable R script**. (scripts/build_pk_monolix.R)

The dataset has been built by running the R script: (output/pk1.csv)

- 406 rows, 12 subjects
- 132 dose events (12 × 11 daily doses ✓)
- 274 observations = 232 quantifiable + 42 BLQ (censored)

Monolix loaded the dataset cleanly with the full column-type mapping.

➤ **Produced files**: Auditable R script + Dataset

2 Autonomous model development

Monolix-agent – LLM: Opus 4.8 – Cost: \$0.33

Develop a popPK model for pk1.csv including structural model, random effects and covariates. Stop only once you found a good model.

[...] The model development is complete. Here's the final summary.

Final population PK model — pk1.csv

Project file: monolix/pk1_05_corrCIQ.mlxtran

Model development path

Run	Model	OFV	BICc
01	1-cpt oral, combined error	687.18	722.70
02	2-cpt oral, combined error	308.11	359.83
03	2-cpt, proportional error	307.60	353.70
04	+ covariate search (COSSAC/BICc)	307.60	353.70
05	+ CI-Q correlation (final)	302.37	350.96

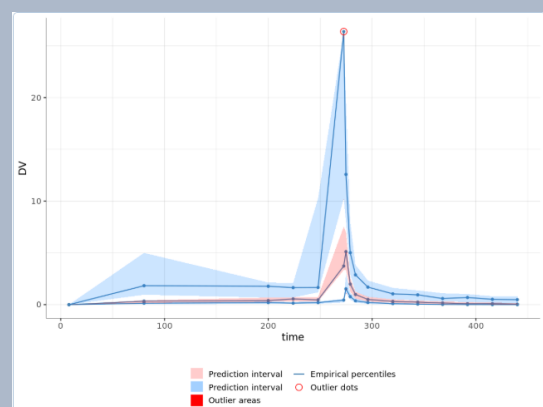
Structural model

2-compartment, first-order absorption, linear elimination (ka, Cl, V1, Q, V2)

Statistical model

- IIV (log-normal) on all five parameters
- CI-Q random-effect correlation
- Proportional residual error
- No covariates (AGE, SEX, DOSE all rejected)

Diagnostics (all acceptable)



➤ **Produced files**: Monolix projects + plots as png images

3 Simulations of various dosing regimens

Simulx-agent – LLM: Opus 4.8 – Cost: \$0.50

Using r09PKPD.mlxtran, I want to simulate several dosing regimens: 200 mg Q8W, 400 mg Q8W, and 800 mg Q12W, and calculate the percentage of individuals which have a reduction in LDLc of at least 50 mg/dL. Import the project in Simulx, setup the simulation, and generate a summary table.

What was built

3 groups, 100 subjects each, 200 mg every 56 days ×7 (Q8W), 400 mg every 56 days ×7 (Q8W), 800 mg every 84 days ×5 (Q12W)

Output: structural LDLc on a daily grid, 0–364 days.

Responder definition: reduction from each subject's own baseline (= LDL-c at t=0) at trough, flagged as a responder when the reduction reached ≥50 mg/dL

Assumptions (please confirm)

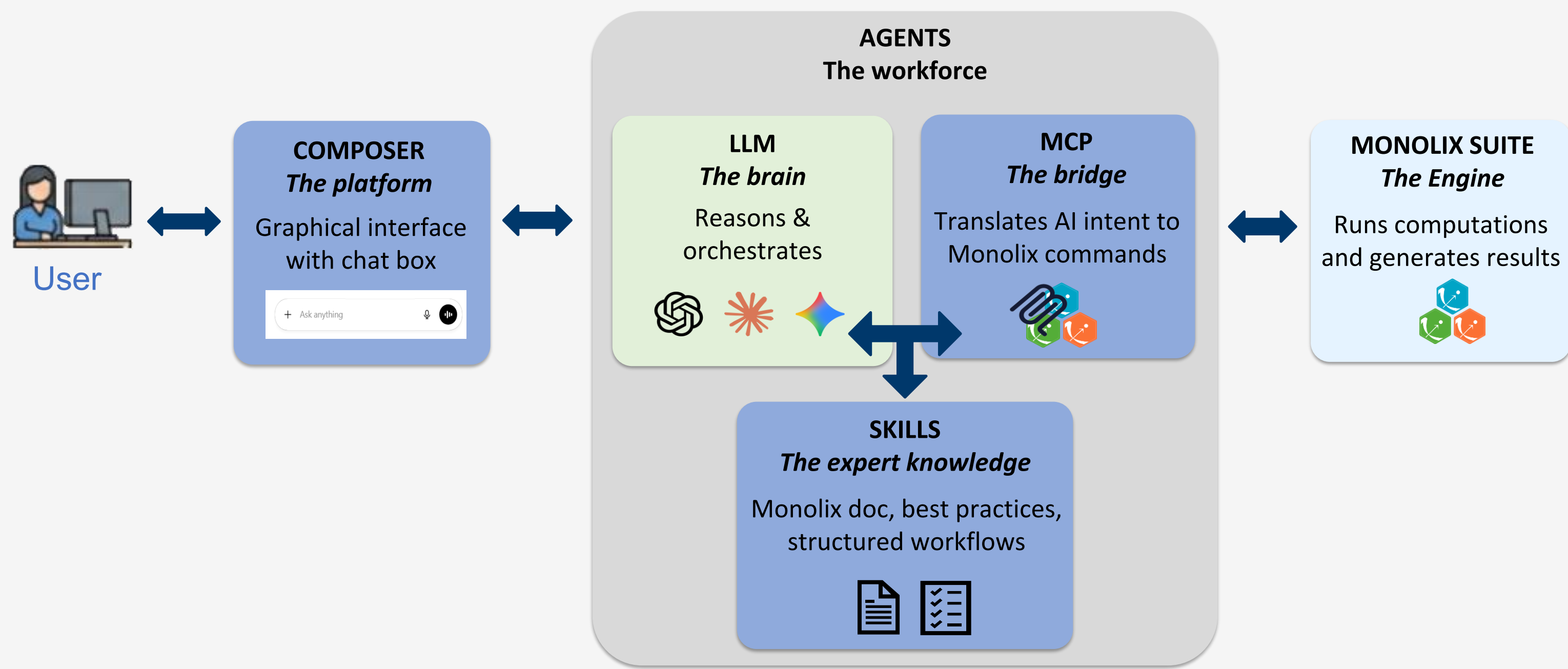
Time unit = days (per the model's TIME_d)

Group size 100, 1 replicate, default seed

Regimen	N	% Responders
200 mg Q8W	100	9%
400 mg Q8W	100	48%
800 mg Q12W	100	34%

➤ **Produced files**: Simulx projects + summary .csv table

PLATFORM DESIGN

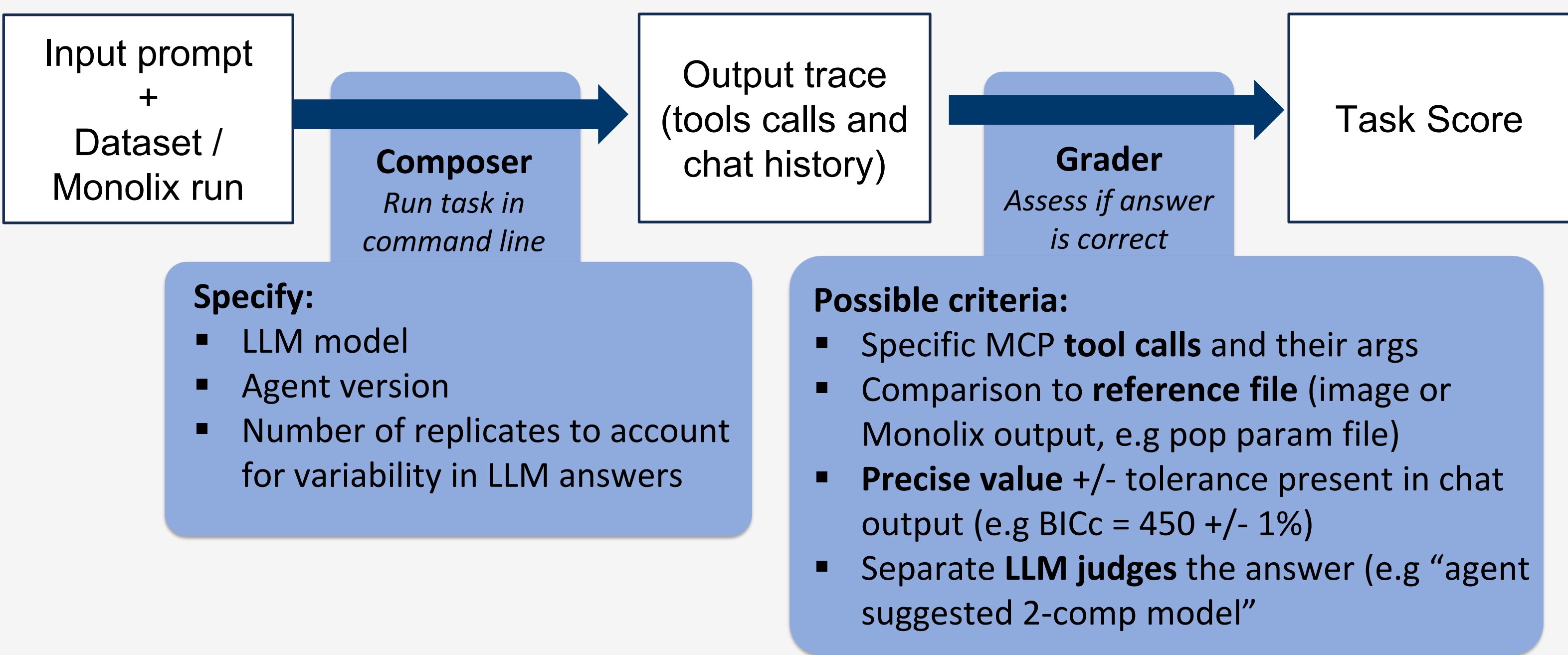


- **Computation-in-the-loop**: every calculation and result is produced by the already validated MonolixSuite engine.
- **Fully reproducible and reviewable**: each prompt creates standard MonolixSuite projects that can be opened, inspected, and refined in the Monolix graphical interface. Chat history showing MCP calls and reasoning can be saved and reviewed.
- **One integrated environment**: the Composer platform unites the chat-driven interface, modeling tools, and agentic capabilities in a single place.
- **Security**: Composer, engines, and MCP tools run in containers on the user's machine. Data is shared with the LLM, and confidentiality depends on your agreement with the LLM provider.

TESTING PROCEDURE

Automated tests allow to:

- **Quantify** the performance of the agents on representative tasks
- **Compare** the performance across different LLM models
- **Assess** the impact of changes in the skills or MCP



We have defined basic requirements for the agents with each requirement being assessed by several test cases (Tasks). **Percentage of test cases with correct agent actions and answers** (passing all defined criteria) is shown below.

	Requirements	MONOLIX			SIMULX		
		# tasks	Sonnet-5	GPT-5.6 Terra	# tasks	Sonnet-5	GPT-5.6 Terra
Capabilities	The agent creates, saves, and reloads projects correctly	8	100%	88%	3	100%	100%
	The agent applies and verifies configuration changes	4	100%	100%	3	100%	100%
	The agent launches every task type and reaches the correct terminal state	6	83%	83%	2	100%	100%
	The agent retrieves results correctly for every task type	5	100%	100%	2	100%	100%
	The agent generates plots that match the requested settings	42	82%	83%	NA	NA	NA
Safety	The agent never invents, approximates or substitutes values	4	100%	87%	4	90%	70%
	The agent never writes to source data files	4	85%	80%	4	75%	80%

Conclusion: Agentic workflows in Composer automate pharmacometric tasks while every result stays reproducible and reviewable in MonolixSuite.

Next steps: Evaluation is ongoing: we are expanding the test suite with new tasks and benchmarking the agents against additional LLMs.