A System Overview

In Table S1, a description of each agent, along with the LLM they are based on, can be found. Empirically, we found that most models performed significantly better when using gpt-5, compared to older models (gpt-4o, gpt-4.1).

Table S1: Overview of agents used in this work and models

Agent Name	Description	LLM Model
Supervisor	Responsible for understanding the user request, and developing a plan of actions needed to be performed to set up a simulation. Delegates to the various agents according to its plan.	gpt-5
Structure Expert	Finds the appropriate (placeholder) structure, and places a copy in the simulation folder.	gpt-5-mini
Force Field Expert	Given a request, decides what the appropriate force fields are. If necessary, it combines them, and writes new force field files (e.g., `force_field.def`, `pseudo_atoms.def`)	gpt-5
Simulation Input Expert	Creates a `simulation.input` file, depending on the requirements of the simulation. Needs to decide which keywords are appropriate, and where to fill in numbers and to template.	gpt-5
Coding Expert	Writes code to replicate the template folder for each necessary run. Needs to understand how the simulation template is set up, and which fields need to be filled in and how.	gpt-5
Evaluator	Evaluates the task performance of each agent by inspecting files created during their execution and flags any potential mistakes it finds.	gpt-5
Paper Search Agent	Uses Semantic Scholar search to find appropriate research papers, and downloads them.	gpt-5-mini
Paper Extraction Agent	Reads downloaded papers and extracts any relevant information from them (e.g., molecule definitions, interaction parameters).	gpt-5-mini
Force Field Writer	Reads the findings produced by the paper extraction agent and transforms them into RASPA force field files.	gpt-5

An overview of the tools can be found in Table S2. In addition to the tools enumerated here, agents have access to various tools for reading, writing, and copying files.

Table S2: Overview of tools available to various agents.

Tool Name	Description
list_all_example_simulation_inputs	Gives the names and description of example input files from the RASPA manual
read_atoms_in_file	Returns the set of atoms present in a `framework.cif` or `molecule.def` file.
count_atom_type_in_cif	Counts how often a given atom type occurs in a CIF file.
get_unit_cell_size	Returns the lattice parameters of the unit cell defined in a CIF file.
get_all_force_field_descriptions	Lists the available force fields and their metadata.
get_atoms_in_ff_file	Lists the set of atoms for which parameters are defined in a force field file (`force_field.def`, `force_field_mixing_rules.def`, `pseudo_atoms.def`).
semantic_scholar_search	Performs a search query using the Semantic Scholar.
download_paper	Downloads a paper given a DOI.
read_paper_headers	Lists the section headers of a paper after it has been parsed.
read_paper_section	Returns the content of the section of a paper.

B Run Details

In Table S3, a more detailed explanation can be found on issues and mistakes during the simulation setup process. Overall, most mistakes are minor and could be fixed in a more robust, future version of the system.

Table S3: Notes on simulation setups with mistakes/intricacies

Task	Strucutre	Adsorbate	Notes
Isotherm	1	3	In multiple runs, all adsorbate files were copied into each simulation folder. Only the correct adsorbate was used in `simulation.input`, resulting in correct simulations.
HOA	500	1	In one of the runs, all possible moves were defined (with 0 probability, except widom insertions) for the adsorbate. In another run, no moves were specified for the adsorbate, resulting in an incorrect simulation where the adsorbate doesn't move.
HOA	500	3	In one of the runs, the structure CIF files were not copied, resulting in failed simulations. In multiple runs, redundant/unused files were left as part of the force field.
Isotherm	500	1	In one of the runs, a minimum amount of unit cells was enforced, resulting in many structures with more unit cells than necessary, resulting in slower simulations.
Isotherm	500	3	Rather than generating separate simulations for all three adsorbates, a mixture isotherm was calculated in one of the simulations. While this is a valid simulation, this was not requested in the task. In another run, a cutoff of 24 Å was selected, resulting in twice as many unit cells.

C Combined System Run

In Figures S1 and S2, a full trace is provided, to illustrate the systems behavior. For clarity, some messages have been summarized or omitted, highlighting the key interactions and decision points while omitting routine or repetitive exchanges.

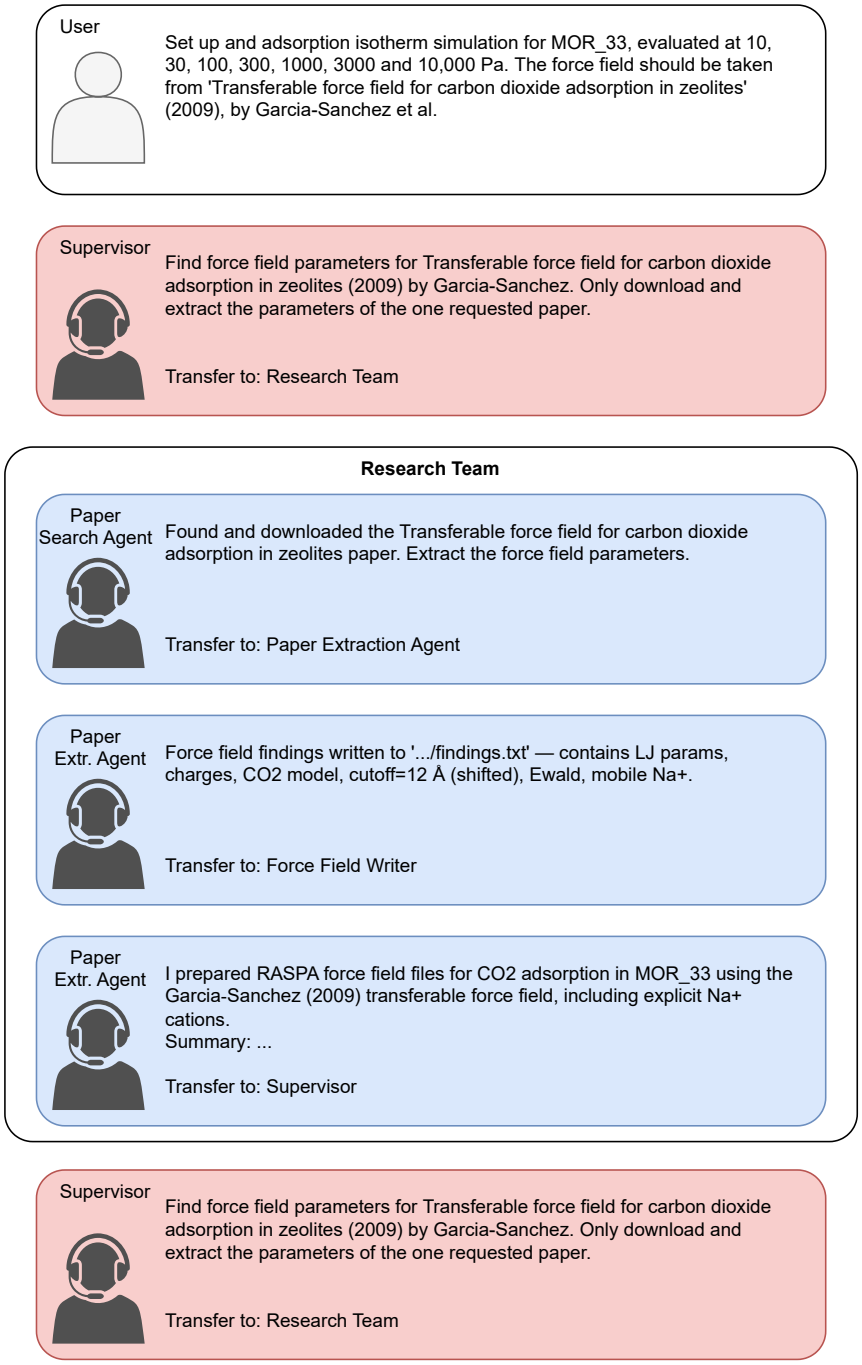

Figure S1: First part of the trace of the full system run.

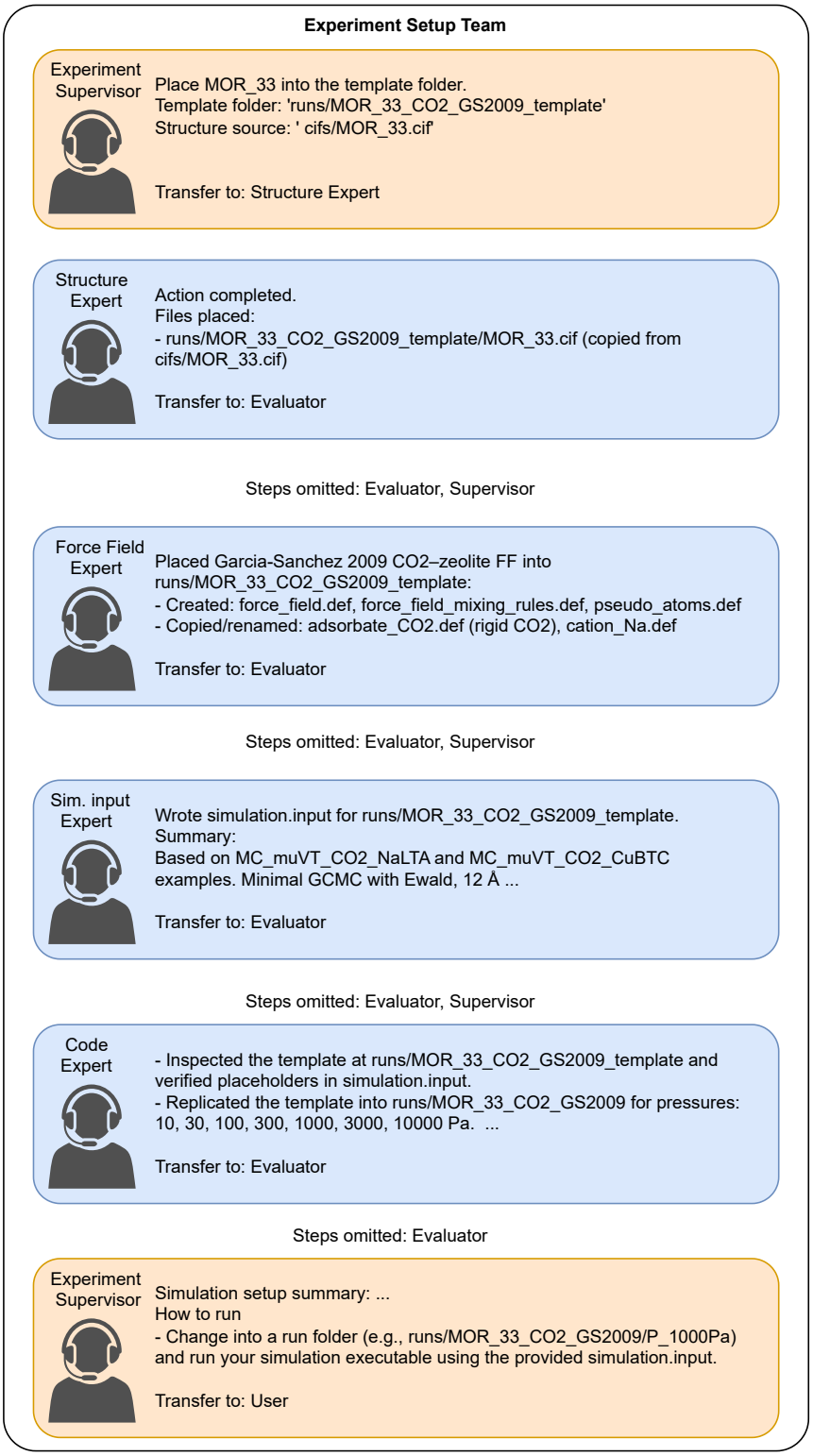

Figure S2: Second part of the trace of the full system run.