

Can households talk their way to flexibility?

UCL Energy Institute

16th July 2026

Reda El Makroum

Energy Economics Group, TU Wien, Vienna, Austria

elmakroum@eeg.tuwien.ac.at



TECHNISCHE
UNIVERSITÄT
WIEN



We need flexible homes

A fully decarbonized energy system requires the active participation of people, yet residential demand remains among the hardest to make flexible.

500 GW

global DR capacity
needed by 2030¹

~60%

from buildings and
electric vehicles¹

<20%

voluntary opt-in
for residential DR²

¹ IEA, *Demand Response*, 2023 ² Fowle et al., 2021

Why is it hard?

Research ties this challenge to three barriers:

- Adapting daily routines to grid signals is **complex**, even with incentives.¹
- Handing over control of home devices feels **intrusive**.¹
- Participation in demand response programs tends to **fade over time**.^{2,3}

¹ Parrish et al., 2020 ² Siitonen et al., 2023 ³ Sloat et al., 2023

Can we use AI to deal with this?

Today, we seem to like to **converse** with systems more. Could the same shift make flexibility provision easier?

Search

We now prefer to ask ChatGPT for simple questions.

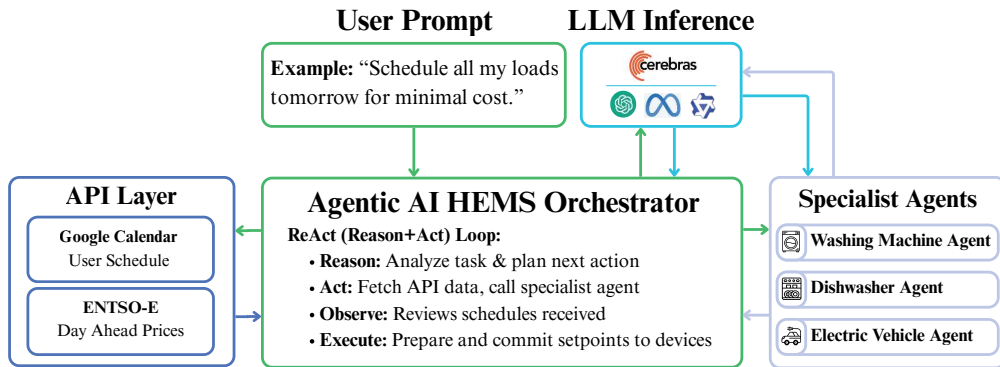
Code

We now talk to a coding agent to do programming.

Agentic AI makes this within reach: these are systems that can go from natural language input to execution of tasks through **understanding, reasoning, planning and tool use**.¹

¹ Chow et al., "From Language to Action: A Review of Large Language Models as Autonomous Agents and Tool Users," 2025.
[arXiv:2508.17281](https://arxiv.org/abs/2508.17281)

What does Agentic AI look like?



The orchestrator toolset

We develop six tools and make them available to the orchestrator, executed in adaptive sequences based on user requests:

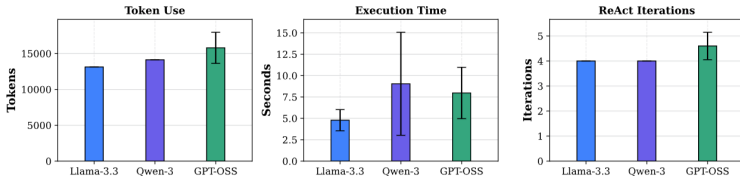
- `get_electricity_prices`: fetches day-ahead prices from ENTSO-E (96 timeslots).
- `get_calendar_constraint`: extracts deadlines from Google Calendar events.
- `calculate_window_sums`: evaluates all valid time windows for cost analysis.
- `call_appliance_agent`: delegates scheduling to a specialist agent.
- `schedule_appliance`: executes the final schedule as a 96-element binary array.
- `finish`: terminates the loop and presents a user-facing summary.

How does it perform?

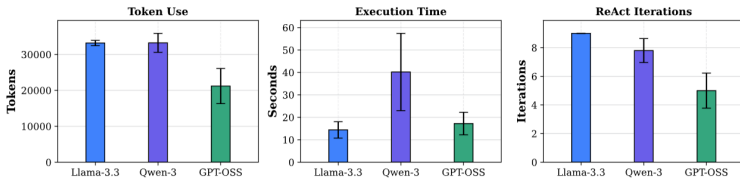
	Was the orchestrator able to schedule the WM only?		Was the orchestrator successful in scheduling all the loads?		
Llama-3.3	✓	✓	✓	✓	✓
Qwen-3	✓	✗	✓	✓	✗
GPT-OSS	✓	✗	✓	N/A	N/A
	Single	Multi	WM	DW	EV

What does it cost to run?

(a) Single-appliance scheduling



(b) Multi-appliance scheduling



Why not just use a traditional optimizer?

Adding a natural language interface to an existing optimizer only improves **input accessibility**. The agentic approach enables capabilities that go **beyond scheduling**:

- Answering household energy queries (“How much did I spend this week?”)
- Explaining scheduling decisions in plain language
- Identifying consumption inefficiencies and recommending adjustments
- Adapting to qualitative preferences that resist formalization

These advisory and analytical functions require **reasoning capabilities** that optimization backends cannot provide, regardless of their interface layer.

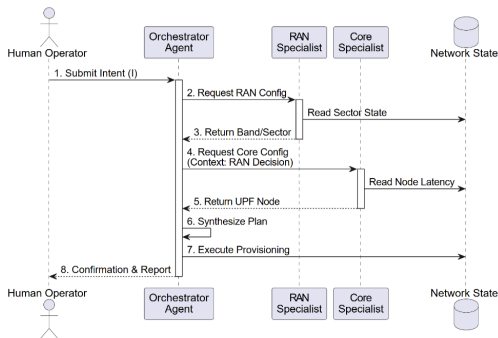
So could this lower the barriers?

- **Complexity:** talking to the system widens who can take part, including anyone for whom apps and screens are a barrier, whether through age, physical impairment, or digital literacy.
- **Intrusiveness:** nothing runs unless the user agrees to it, which can build trust in the system.
- **Disengagement:** every event reports what it did and what it earned, and the user can ask for details whenever they want.

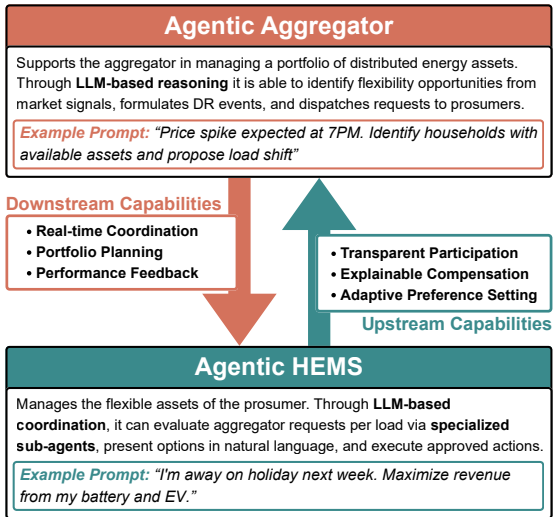
Whether this enables more simplicity, trust, and engagement is worth investigating through real-life pilots.

The pattern generalizes

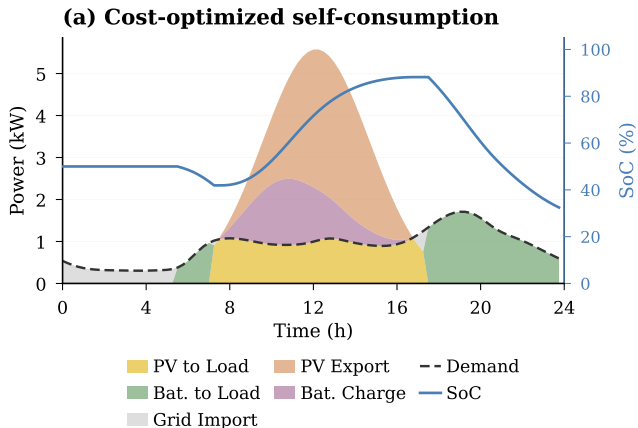
The same orchestrator-with-specialist-agents pattern has been reused beyond the home, here for intent-based 6G network provisioning.



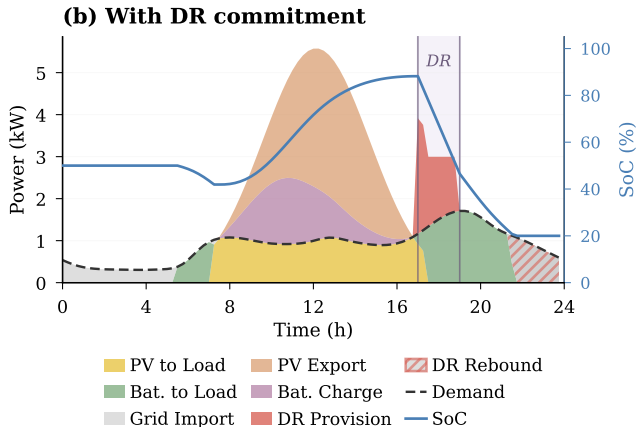
We introduce Conversational Demand Response



The Case Study: business-as-usual



The Case Study: under a DR commitment



Downstream: Aggregator-Initiated DR Dispatch

 Aggregator

We need 3 kW of flexibility from the portfolio between 17:00 and 19:00. What can households offer?

 Agentic Aggregator

Dispatched a request to HH-001. Status: pending. I will monitor and report the household's decision as soon as it is available.

Downstream: Aggregator-Initiated DR Dispatch

 Agentic HEMS

DR event received. Delegated to battery agent. Fully feasible.
Net benefit: € 1.13. SoC: 30 % → 43 %. No comfort impact.
Would you like me to accept?

 Agentic HEMS

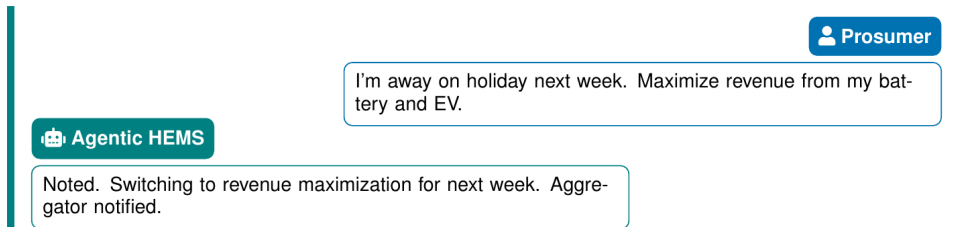
Done. Commitment submitted to aggregator.

 Prosumer

Yes, go ahead.

On approval, the aggregator confirms: *"HH-001 accepted. Portfolio updated."*

Upstream: Prosumer-Initiated Profile Update



Upstream: Prosumer-Initiated Profile Update



Agentic Aggregator

HH-001 will be away next week and has granted full asset access for DR dispatch.
I've re-optimized the portfolio accordingly.

Together, the two directions form the full CDR loop.

Does it work in practice?

Six scenarios spanning the range of interactions, each run five times.

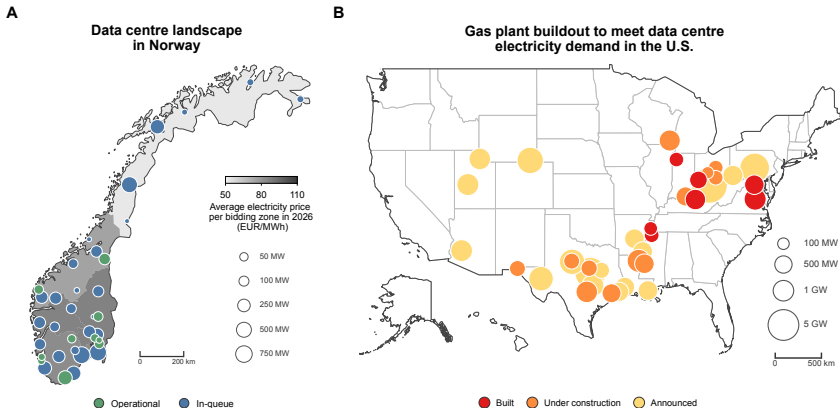
Direction	Scenario	Iters	Tool calls	Tokens (k)	Time (s)
Downstream	Acceptance	3.6	2.6	23.4	8.3
	Rejection	5.0	3.6	34.2	9.8
	High-target	3.4	2.4	21.8	7.8
Upstream	Availability	1	0	1.3	1.3
	Preference	1	0	1.0	1.7
	New Asset	1	0	1.6	1.3

Across all six scenarios, **no interaction exceeds 12 seconds**, with consistent outputs across runs.

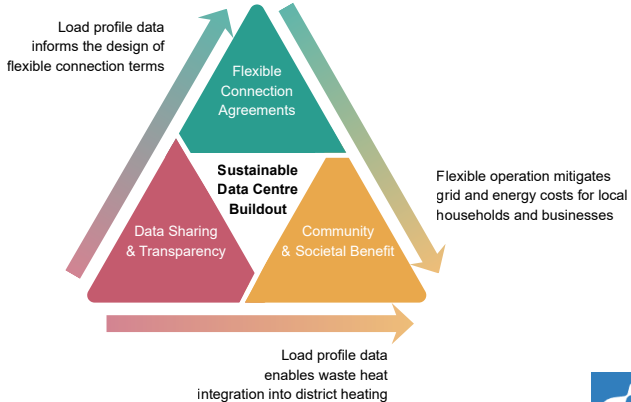
Real-world considerations

- **Ensuring consistent and secure operation:** Guardrails, protection against prompt injection, and data security considerations are needed before agents act on a household's behalf.
- **Fit in the value chain:** We imagine CDR to come as an additional layer to automated demand response.
- **System integration:** Reaching households at scale requires utilities and aggregation services to embed these interfaces into their existing platforms.
- **Toward local inference:** LLM capabilities are evolving fast, and soon we might have capable models that can run locally on-device.
- **Energy Footprint:** More flexibility provided at the cost of higher data center consumption is a tradeoff worth investigating.

What about the AI's own footprint?



What about the AI's own footprint?



Thank You

Reda El Makroum

Energy Economics Group (EEG), Technische Universität Wien

elmakroum@eeg.tuwien.ac.at | redaelmakroum.dev

github.com/RedaElMakroum



TECHNISCHE
UNIVERSITÄT
WIEN

