

AI Agents as Power Brokers: Coordinating Self-Interested Household Energy Agents under a Shared Transformer Limit, Backtested on French Open Data

Technical report (proof of concept, living document)

Aditya Narayan Sankaran
Télécom SudParis, Institut Polytechnique de Paris

Version of October 22, 2026

Abstract

Residential flexibility in France is already steered by national price and alert signals (the Tempo tariff, off-peak hours, Ecowatt) and, increasingly, by local needs of the distribution network. Each household reacts to these signals alone, while the low-voltage transformer that feeds a street is shared. We study a street in which every home has its own planning agent that minimises its own bill, and a *station agent* representing the transformer that sends local price nudges only when the sum of the plans would overload it. Plans are computed by a mixed-integer optimiser, checked by a deterministic validator, and (in the planned system) negotiated and explained in plain language by large language models. We build an end-to-end, no-look-ahead backtest on public French data: building stock (BDNB), measured load curves (Enedis), demand and carbon intensity (RTE), Tempo colours, archived weather forecasts, regulated tariffs (CRE) and day-ahead prices (ENTSO-E).

For a single synthetic electrically heated home on Tempo, a day-ahead optimiser fed only with information available at 18:00 the evening before lowers the annual bill by a median of €293 relative to an off-peak timer (range €223–€363 over 6 household variants) at equal comfort, capturing 84 % of the perfect-foresight gain in winter. Outside winter there is almost nothing to gain and forecast error makes the optimiser no better than the timer. On the whole street (222 synthetic homes, five population seeds, 400 kVA), such optimisers acting alone *increase* winter transformer overload compared with a timer (2 764 against 1 080 kWh), because they all move into the same cheap hours. A station agent that sends local nudges and accepts only re-plans that lower the planned overload removes 77 % of it with no comfort violation, at a cost of €38 for the whole street over the winter; a naive version that lets everyone react at once makes the peak much worse. A 7B open-weight language model reads residents’ notes (98.5 % of fields correct) and negotiates on their behalf (90–100 % correct decisions, no false accepts) behind deterministic validators. Today’s congestion depends on whether one large commercial site shares the transformer, which only the network operator can confirm.

1 Introduction

French households already receive strong, time-varying signals about when electricity is scarce. The Tempo option prices a red-day peak hour at several times a blue-day hour [14]; off-peak hours are moving from the night towards the sunny early afternoon [13]; and on tense days RTE’s Ecowatt signal asks citizens to shift or reduce use [27]. These signals are *national*: they reflect the state of the whole power system. The network that actually carries a street’s load is *local*: a medium- to low-voltage (HTA/BT) transformer of a few hundred kVA [1] shared by every home and shop on it. When many households respond to the same national signal at the

same moment (every electric vehicle starting at the first off-peak hour, every hot-water tank at 22:00) the local transformer sees a rebound peak that no national signal is designed to prevent.

This report studies a simple division of labour. Each home has an agent that plans its own flexible loads (heating, hot water, electric-vehicle charging, appliances) to minimise its own bill under its own comfort constraints; homes do not share or trade energy, consistent with the restrictions on direct neighbour-to-neighbour sales outside collective self-consumption schemes [10]. A *station agent*, representing the transformer, adds up the plans and, only if the sum would exceed the transformer limit, sends local price nudges so that homes re-plan; it vetoes a plan only as a last resort. An optimiser computes feasible schedules, a deterministic validator checks them, and language models act as the interface: they turn residents’ preferences into constraints, negotiate with the station and explain each plan in one sentence.

The question is whether such coordination is worth it, judged honestly. We therefore build a backtest on real French data with strict validity rules: agents only see what was known at 18:00 the evening before; whole seasons are simulated rather than chosen weeks; strong baselines (an off-peak timer and a perfect-knowledge oracle) bound the result; and household euros come only from official regulated tariffs.

Contributions and status.

1. **A reproducible, no-look-ahead data and backtest pipeline** for one real street, built only on public French data (Sections 4 and 5). *Status: done.*
2. **Single-household evidence** on what day-ahead optimisation can and cannot gain against a timer under Tempo, with the gap to a perfect-knowledge oracle (Section 7.2). *Status: done (one household archetype, six variants).*
3. **Street-level coordination:** whether uncoordinated household optimisers create transformer rebound peaks and how much a station agent removes them, at what cost (Section 7.3). *Status: in progress.*
4. **Language-model agents as interface and negotiator**, compared across an open-weight and a commercial model family, with a validator rejection rate as a reported metric. *Status: planned.*

2 How flexibility is planned and executed in France today

Our design does not invent a new institution; it rearranges mechanisms that already exist. Table 1 summarises the frameworks through which French household flexibility is currently *planned* (who decides when), *signalled* (how the decision reaches the home) and *executed* (what actually switches). We use them as baselines, as sources of realistic timing, and as the institutional hooks a station agent could plug into.

Three observations shape our design. First, every signal a household receives today is either *national* (Tempo, Ecowatt, NEBEF activations) or *fixed in advance* (off-peak slots); none reflects the state of the household’s own transformer. Second, the network operator already acts on household equipment when needed (Linky control of tanks) and already pays for local flexibility (tenders), but only at a scale of hundreds of kW and above. Third, the timing of these signals matches a day-ahead loop: Tempo colours are known before 11:00 and Ecowatt at 17:00 on $D-1$, so an evening planning step at 18:00 can use them without any look-ahead.

Framework	Who decides	Signal and timing	What is executed; relevance here
Tempo option [14, 15, 29]	RTE chooses the colour; regulated price set by the CRE	Colour of day D published on $D-1$ before 11:00; 22 red and 43 white days per season, red only Nov–Mar on weekdays	Households shift or reduce by themselves. National signal, known at our 18:00 planning time; our main household tariff.
Off-peak hours (HP/HC) and the TURPE 7 reform [12, 13, 42]	Enedis assigns each meter its slot within CRE rules	Fixed daily slots; from 2025–2027 at least 5 night hours plus afternoon hours (11:00–17:00) in summer	The meter’s relay switches the water heater. Today’s default “timer”; the reform moves cheap hours towards midday solar.
Linky load control [35, 43]	Enedis, under ministerial order	Meter contact and virtual relays; e.g. midday water-heater start suppressed in winter 2023–24	Direct control of tanks by the network operator: a precedent for station-level actions.
Ecowatt and load-shedding plans [27, 28]	RTE with ADEME; government for shedding	Green/orange/red published the evening before (17:00) and updated in the morning	Voluntary reduction; rotating cuts as last resort. A national alert; a station veto mirrors the last-resort step locally.
Aggregated demand response (effacement, NEBEF) [8, 16, 19]	Aggregators, valued in RTE mechanisms	Short activations at national peaks	Remote 5–15 min cuts of radiators and tanks (e.g. Voltalis). Shows residents accept automated curtailment of heating.
Local flexibility tenders [9, 39]	Enedis buys flexibility where the grid is constrained	Calls for tenders with capacity reservation; activations of at least tens of minutes	Defers network reinforcement; today mostly MW-scale. The station agent is a street-scale version of this purchase.
Collective self-consumption (ACC) [10, 11]	A legal entity (PMO) groups producers and consumers	Allocation keys sent to Enedis; participants within 2 km (derogations up to 20 km)	The only legal route for neighbours to share production; we keep it as a “what if” only.
Smart-charging capability for EV chargers [2]	Regulation (décret IRVE, revised 2021)	Chargers must be controllable	Makes station-requested EV shifts technically feasible.

Table 1: Existing French frameworks that plan, signal or execute residential flexibility. Secondary sources are marked in the bibliography and will be replaced by primary texts.

3 Related work

Language-model agents for home energy management. Recent work uses large language models to schedule appliances in a single home. Jonah et al. [32] report that LLM agents reach close to the savings of an optimisation oracle for one household and recommend a validator; El Makroum et al. [20] use an orchestrator with specialist agents; EnergyBridge [46] adds resident authorisation and an LLM-simulated user; Cremer [17] lets end-users negotiate individual grid-connection contracts with an LLM that calls power-system tools. These studies consider one home or one contract. We study many self-interested homes coupled through a shared physical limit, with a season-long backtest on French data.

Multi-agent demand response benchmarks. CityLearn [45] standardises multi-agent reinforcement learning for demand response in groups of buildings. It is a natural point of comparison for coordination quality, but its buildings, tariffs and data are not French and its agents are learned controllers rather than planners with a natural-language interface.

French household and district simulators. EDF R&D’s SMACH simulates household activities and the resulting appliance load curves with a multi-agent model [6]; CSTB’s DIMOSIM simulates district energy flows, including occupant behaviour and electrical grids [3, 38]. Both are close in spirit, but we found no open release; we use public measured curves (Enedis) instead of simulated occupant behaviour.

Low-voltage grid models. pandapower [44] with the SimBench benchmark grids [36] is the standard open toolchain for low-voltage congestion studies, e.g. transformer overloading under German electrification pathways [4]. We model only the transformer limit; adding feeder voltage limits with pandapower is future work.

4 Data

The street is modelled on the real buildings of one street in the inner suburbs east of Paris; in public material it is called *Rue des Voisins* and real addresses are never shown. Households are synthetic. All data are public and fetched by scripts that save raw responses unchanged with a manifest (Appendix A). Everything is aligned to one hourly grid in Europe/Paris time from 1 September 2024 to 30 September 2025 (9 480 hours, daylight-saving days of 23 and 25 hours).

The street. BDNB lists 89 building groups and about 228 dwellings, mostly old stock (median construction year 1900). Heating energy is known for 34 groups only (26 gas, 8 electric); a domestic hot-water system is known for 34, of which 13 are electric tanks. Enedis meters five residential blocks of at least ten dwellings; one large commercial site uses 623 MWh/year on a single meter. Section 6 explains how gaps are filled.

Source	Content used	Role in the backtest
BDNB [18]	89 building groups on the street: use, dwellings, heating and hot-water systems, DPE label	Builds the synthetic household population
Enedis [22]	Annual consumption per address (blocks of ≥ 10 dwellings, 2018–2025); average half-hourly load per site by tariff profile (≤ 36 kVA and > 36 kVA)	Base load shapes, calibration targets, evidence of Tempo response
RTE eCO2mix [41]	National and Île-de-France demand, production mix, CO ₂ intensity, day-ahead demand forecast	Carbon metric (outcome); demand forecast (known at $D-1$)
Tempo calendar [33]	Daily colour since 2014	Tariff signal, known at $D-1$
Open-Meteo [37]	Forecasts as issued 1 and 2 days ahead; latest run; ERA5	Agents see only forecasts issued ≥ 48 h ahead
PVGIS [25]	Hourly PV output per kWp, 2005–2023	Calibrates a PV model (no 2024–25 data)
ENTSO-E [24]	French day-ahead prices	Wholesale signal; business cost ranges
CRE tariffs [14]	Base, HP/HC and Tempo prices by subscribed power and validity period	The only source of household euros

Table 2: Public data sources and their role. Every column carries a role: *known at 18:00 $D-1$* (may reach agents) or *outcome* (metrics and physics only).

5 Method

5.1 Household model

Each household has optional devices: electric space heating, an electric hot-water tank, an electric vehicle and a shiftable washing machine, plus a non-flexible base load. Space heating follows a first-order thermal model discretised exactly over one-hour steps,

$$T_{t+1} = a T_t + (1 - a) \left(T_t^{\text{out}} + \frac{P_t^{\text{heat}} + G_t}{H} \right), \quad a = e^{-H/C}, \quad (1)$$

where H (kW/K) is the heat-loss coefficient, C (kWh/K) the thermal capacity, and G_t the internal and solar gains (base load plus an effective solar aperture times global horizontal irradiance). The tank is an energy store with standing losses and hourly draws; the EV is a battery that must reach a target state of charge at departure and charges only at home. Each household has its own tariff option (Base, HP/HC or Tempo), subscribed power (kVA) and, for HP/HC, its own off-peak slot.

5.2 Day-ahead optimiser

At 18:00 on day $D-1$ the optimiser chooses hourly heating, tank and EV power and the washing start for day D by solving a mixed-integer linear programme with HiGHS [30]. It minimises the energy bill at the household’s tariff subject to the thermal model, a comfort floor (the same floor every strategy must respect), a pre-heating ceiling, tank and EV constraints, one washing cycle in allowed hours, and the subscribed power. Comfort, reserve and EV constraints carry penalised slack so that every day is feasible; penalties are far above any price so slack is used only when physics leaves no choice. The plan covers 00:00 on D to 07:00 on $D+1$ and only day D is executed (receding horizon): planning only to midnight split the night off-peak window and pushed water heating into peak hours.

5.3 Strategies and one execution engine

Four strategies run through the *same* hourly simulator with actual weather and actual base load: (i) *do nothing*: a thermostat holds the comfort floor, the tank follows the off-peak relay (a plain thermostat on Base), the EV charges on plug-in; (ii) *timer*: as (i) but the EV charges only in off-peak hours; (iii) *optimiser (forecast)*: the day-ahead plan built only from information available at 18:00 $D-1$, starting from a *predicted* midnight state; (iv) *oracle*: the same optimiser with actual weather, base load and start state. The simulator applies the same rules to all: heating is raised if the room would fall below the floor, the tank if it would fall below its reserve, the EV if it would miss its target; above the subscribed power, EV charging is shed first, then the tank, then heating. Deviations from a plan (“overrides”) measure the cost of forecast error.

5.4 No-look-ahead protocol and self-checks

Every input column is tagged *known at 18:00 $D-1$* (day-ahead prices, Tempo colour, weather forecasts issued at least 48 h ahead, base load forecast from days $D-30$ to $D-3$) or *outcome* (actual weather, actual load, CO₂). Runs fail automatically if the oracle needs any override or shedding (the planning model and the simulator disagree) or if any strategy beats the oracle at equal comfort. A refactor must reproduce a reference run to the cent. These checks caught three modelling errors during development (Section 7.2).

5.5 Street population and calibration

Each residential BDNB building group is expanded into its dwellings (houses or flats). Known attributes always win (for instance, a block with electric heating but solar hot water has no electric tank); unknown heating and hot-water systems are imputed from the street’s own known shares. Per-home parameters (insulation, comfort floor, base-load scale, tank draws, washing days, EV times and trips, tariff option, subscribed power, off-peak slot) are drawn from seeded distributions, with a fixed number of random draws per home so that adoption scenarios change *only* who owns an EV or a heat pump. Commercial groups become fixed loads shaped by Enedis curves and scaled to their metered annual consumption where known. The five blocks metered by Enedis are calibrated on today’s behaviour (*do nothing*): one multiplier per block on insulation (bounded to a factor two either way) and then on the base load, until simulated annual consumption matches the mean of the 2024 and 2025 meter readings; the residual is re-verified on the simulated population and must stay within 3%.

5.6 Street and station agent

The street total is the sum of all household and business loads; the station compares it with the transformer rating in kVA under an assumed power factor of 0.95. At 18:00 on $D-1$ it receives every household’s plan for day D and a business-load forecast (the same hour one week earlier). If the planned total exceeds the rating in some hours, it sends a local price nudge for exactly those hours, proportional to the relative overload. Nudges are signals, not bills: household euros stay at the official tariff.

We compare two coordination rules. In the *naïve* rule every household re-plans against the nudge in each round. In the *accept* rule, batches of the households contributing most to the overload re-plan, and the station accepts a re-plan only if it lowers the planned overload energy; otherwise the household keeps its plan, and the nudge doubles on the next pass. The planned overload can then only decrease. If overload remains after the last pass, the station applies a last-resort cap: flexible power is curtailed proportionally in the overloaded hours, and each capped household may add only an equal share of the remaining headroom elsewhere. A standing check requires that a station run with an unreachable rating reproduce the uncoordinated optimiser exactly. Over the winter at 400 kVA the station sent nudges on a median of 61 [6, 88] days

per population seed and never needed the last-resort cap (0 veto days). In heavier adoption scenarios the cap does fire; it then limits only continuous loads (heating, tank, EV), never a running washing cycle, which an earlier version did and which made some household plans infeasible.

6 Modelling assumptions and why they hold

Where the street’s own data are missing we take the default that public rules or national statistics imply, state why, and sweep it where it matters (Table 3). All are open to correction by a distribution network expert.

Question	Default	Why (source)
Is the 623 MWh/yr commercial site on the street’s transformer?	Yes; “own substation” as sensitivity	Enedis connects consumers above 36 kVA at low voltage up to 250 kVA and at medium voltage (HTA, with their own substation) above [21, 23]. 623 MWh/yr is an average of about 70 kW, so a peak below 250 kVA and a low-voltage connection are plausible.
Transformer rating	Sweep 250, 400, 630 kVA	Ratings are not published; these are standard catalogue sizes for urban HTA/BT substations [1]. Substation locations are public [5].
Off-peak slots (HP/HC)	Pre-reform 8-hour night slots, sampled per meter; “2026 reform” scenario	Slots are set per meter by Enedis and staggered across a street; the TURPE 7 reform (rolled out 2025–2027) keeps at least 5 night hours and adds afternoon hours in summer to use cheap midday solar [13, 31]. Our backtest year precedes the reform.
Tariff option mix	Base 50%, HP/HC 45%, Tempo 5%	National counts reported for 2023: about 10 M Base, 9 M HP/HC, 0.6 M Tempo customers [26]; close to 1 M Tempo by January 2025 [15]. Base customers cannot gain from shifting.
EV adoption	Sweep 3, 10, 25, 50 % of homes	About 3 % of the French vehicle fleet was electric on 1 January 2025 [7], over 1.5 million battery-electric cars [34]; national scenarios foresee steep growth [40]. Congestion is expected only at higher adoption, hence a sweep rather than a forecast.
Heat pumps	Scenario switch (gas homes converted)	Electrification of heating is central to national scenarios [40]; no street-level figure exists.
Conflicting building data	Trust metered energy over modelled attributes	One block is recorded with gas boilers but meters three times the electricity of comparable gas blocks; we keep its heating system and calibrate its non-heating load to the meter.
Diversity between homes	Measured average curves plus seeded per-home variation	Average Enedis curves already contain diversity; identical optimisers would otherwise all choose the same cheapest hour and create an artificial rebound. At least five population seeds.
Power factor	0.95	Assumption for converting kW to kVA.

Table 3: Defaults used where street-specific data are missing.

House 1 (“the Martins”). The single-household study uses an old (~ 1900), $\sim 100 \text{ m}^2$ electrically heated house on Tempo at 12 kVA: $H = 0.25 \text{ kW/K}$, $C = 10 \text{ kWh/K}$ (time constant 40 h),

9 kW radiators, a 200 L tank (2.4 kW), an EV (50 kWh, 7 kW, away 07:00–18:00), a washing machine three times a week, and a constant 19 °C comfort floor. Six one-at-a-time variants change insulation ($H = 0.20, 0.30$), the comfort floor (20 °C) and EV use. All values are synthetic and chosen for plausibility, not measured.

7 Results

7.1 What the open data already show

Households respond to Tempo. On winter weekdays at peak hours, the average Tempo customer in Enedis’ national curves draws 0.76 kW on blue days but 0.58 kW on red days (−23 %), whereas HP/HC customers draw more on red days (+32 %) because red days are cold (Table 4). French households therefore already shift load when a strong, day-ahead price signal exists. These are national averages, not our street, and are not corrected for temperature.

Customer profile	Blue (kW)	White (kW)	Red (kW)	Red vs blue
Tempo (RES3)	0.76	0.81	0.58	−23 %
Peak/off-peak (RES2)	1.07	1.25	1.41	+32 %
Base (RES1)	0.29	0.31	0.34	+16 %

Table 4: Average load per site (kW, all subscribed powers ≤ 36 kVA) at peak hours on winter weekdays (December–February 2024–25), by Tempo colour. Source: Enedis national curves [22]; generated by `src/analysis/data_evidence.py`.

National and local signals diverge outside winter. Wholesale day-ahead prices agree with Tempo in winter: red-day peak hours average €154/MWh against €54/MWh on blue days. Outside winter, however, Tempo’s night off-peak window misses the cheapest hours: on blue days off-peak hours average €56/MWh against €54/MWh for peak hours, because sunny middays clear at negative prices (523 hours in the year, 133 in May alone) [24]. This is the same mismatch the TURPE 7 reform of off-peak hours addresses [13].

The hero week. During the cold spell of 13–19 January 2025 (Tempo red Monday to Friday), national demand peaked at 88 GW on 14 January 2025 and the carbon intensity of French production averaged 47 g/kWh on red days against 18 g/kWh over the season [41].

7.2 One household: what day-ahead optimisation can and cannot gain

Table 5 reports the full-window backtest of House 1 (base variant). The forecast-driven optimiser lowers the annual bill from €3 803 (timer) to €3 511, against €3 427 for the perfect-knowledge oracle and €4 056 for doing nothing, with the same comfort floor. Most of the gain comes from moving heating out of red-day peak hours: energy drawn in those hours falls from 1 357 kWh to 840 kWh. The optimiser uses slightly *more* energy than the timer (it pre-heats into cheap hours) but emits slightly less CO₂, because cheap hours are also cleaner.

Across six household variants (Table 6), the optimiser saves a median €293 per year against the timer (range €223–€363), and €303 in winter alone, where it captures 84 % of the timer-to-oracle gap (78–88 %). Outside winter the picture reverses: even the oracle gains only €16 over the timer, and forecast error leaves the optimiser at €−9 (range €−19 to €5). The value of day-ahead planning for this household is therefore almost entirely a winter, red-day effect; the gap-captured ratio outside winter is unstable because its denominator is small.

Two further findings matter for the street. A plain off-peak timer overloads the 12 kVA subscription on cold nights, when the EV, tank and heating all start at 22:00: the simulator

Strategy	Bill (€)	Energy (kWh)	Red-HP (kWh)	Peak (kW)	CO ₂ (kg)	$\bar{T}_{\text{day}}$ (°C)	Overrides (h)
<i>Full window</i>							
Do nothing	4 056	20 300	1 609	11.9	489	20.6	0
Timer	3 803	20 300	1 357	12.0	493	20.6	0
Optimiser (forecast)	3 511	21 040	840	12.0	480	20.9	820
Oracle	3 427	20 656	803	12.0	472	20.7	0
<i>Winter</i>							
Do nothing	3 101	13 678	1 609	11.9	409	19.0	0
Timer	2 908	13 678	1 357	12.0	407	19.0	0
Optimiser (forecast)	2 606	14 191	840	12.0	403	19.6	634
Oracle	2 547	13 977	803	12.0	397	19.3	0
<i>Rest of year</i>							
Do nothing	954	6 621	0	9.0	80	21.7	0
Timer	895	6 621	0	12.0	86	21.7	0
Optimiser (forecast)	904	6 849	0	12.0	78	21.9	186
Oracle	879	6 679	0	12.0	75	21.8	0

Table 5: House 1 (base variant), 29 Sep 2024–30 Sep 2025, CRE Tempo TTC prices. All strategies run through the same simulator with actual weather and hold the same 19°C floor. Red-HP: energy in red-day peak hours. Overrides: hours in which the simulator corrected the plan (forecast error); the oracle’s zero is a self-check. Winter: 27 Oct–30 Mar.

Period	Optimiser vs timer (€)	Oracle vs timer (€)	Gap captured (%)
Full window	293 [223, 363]	376 [326, 456]	78 [68, 84]
Winter	303 [241, 358]	361 [310, 423]	84 [78, 88]
Rest of year	−9 [−19, 5]	16 [13, 33]	−56 [−115, 16]

Table 6: Savings against the timer, median [min, max] over six one-at-a-time household variants (insulation, comfort floor, EV use). Gap captured = (timer − optimiser) / (timer − oracle).

has to shed up to 325 kWh per year in the leakiest variant, whereas the optimiser plans around the limit. And the self-checks earned their keep: they exposed a comfort penalty set too low (the oracle preferred a slightly cold room to red-day prices), an overheating penalty equal to the comfort penalty, and a planning horizon that cut the night off-peak window at midnight.

7.3 The street: does coordination pay?

Setup. The street has 222 synthetic dwellings and fourteen commercial loads, including one large site metered at 623 MWh per year, assumed to share the transformer (Section 6). Results are medians with [min, max] over five population seeds, at today’s EV adoption unless stated. The transformer rating is not public; we report 400 kVA as the reference. A 250 kVA unit would be undersized for this street (its winter load exceeds 250 kVA for thousands of hours a year whatever the strategy, which no shifting can fix), and 630 kVA is never reached.

Strategy	Peak (kVA)	h > 250 kVA	h > 400 kVA	h > 630 kVA	Energy (MWh)
Do nothing	481 [468, 485]	2 450 [2 295, 2 639]	94 [70, 125]	0 [0, 0]	1 730 [1 690, 1 768]
Timer	458 [451, 467]	2 502 [2 258, 2 665]	60 [39, 111]	0 [0, 0]	1 730 [1 690, 1 768]
Optimiser (forecast)	484 [440, 520]	2 251 [2 163, 2 543]	132 [24, 208]	0 [0, 0]	1 743 [1 703, 1 781]
Oracle	500 [453, 534]	2 187 [2 100, 2 464]	122 [26, 206]	0 [0, 0]	1 734 [1 694, 1 771]

Table 7: Uncoordinated street over the full window (29 Sep 2024–30 Sep 2025), five population seeds, median [min, max]. Every hour above 400 kVA falls in winter.

Uncoordinated optimisers create a rebound. Every household optimising its own bill against the same national signals does *not* relieve the transformer: over the winter, hours above 400 kVA rise from 60 with a plain timer to 132 with day-ahead optimisers, and overload energy from 1 080 to 2 764 kVAh (Tables 7 and 8). Perfect foresight does not help (122 hours): the cause is synchronisation, not forecast error. Homes that individually save money all move into the same cheap hours.

Winter 2024–25, 400 kVA	Peak (kVA)	Hours > 400 kVA	Overload (kVAh)	Street bill vs. uncoord. (€)
Timer	458 [451, 467]	60 [39, 111]	1 080 [654, 2 387]	–
Uncoordinated optimisers	484 [440, 520]	132 [24, 208]	2 764 [363, 7 470]	–
Oracles (perfect foresight, uncoordinated)	500 [453, 534]	122 [26, 206]	2 772 [407, 7 659]	–
Station (accept-only-improving)	437 [431, 440]	66 [23, 99]	624 [258, 965]	38 [1, 110]

Table 8: Winter 2024–25 (27 Oct–30 Mar), 400 kVA, five seeds. The last column is the change in the whole street’s energy bill (official tariffs) relative to uncoordinated optimisers. Generated by `src/analysis/street_report.py`.

The station removes most of the overload at a small cost. With the accept-only-improving station, overload energy falls to 624 kVAh, 77 % below uncoordinated optimisers and below the timer, and the peak falls from 484 to 437 kVA, with no comfort violation. The whole street pays €38 more over the winter than without coordination (median over seeds); 12 of 222 households are worse off, the worst by €36 over the winter. A compensation scheme in the spirit of local flexibility contracts would cover that easily, but we have not modelled one. The remaining overload comes from forecast error, since the station plans exactly to the rating.

Hero week, 400 kVA	Peak (kVA)	Hours > 400	Overload (kVAh)	Bill vs. uncoord. (€)	Worse off
Timer	458 [448, 467]	16 [13, 21]	426 [327, 759]	–	–
Uncoordinated optimisers	460 [437, 506]	23 [9, 25]	631 [126, 1 115]	–	–
Station, naive (everyone re-plans)	728 [439, 790]	27 [9, 51]	1 217 [126, 2 238]	81 [0, 126]	107 [0, 127]
Station, accept-only-improving	433 [429, 439]	19 [9, 21]	222 [126, 251]	8 [0, 13]	7 [0, 16]
Station, accept + 5 % margin	418 [412, 421]	9 [6, 14]	51 [37, 104]	13 [3, 22]	23 [12, 46]

Table 9: Station design ablation, hero week (13–19 January 2025), 400 kVA, five seeds. “Naive”: every household re-plans against the same nudge in every round. “Margin”: the station plans to 95 % of the rating.

How the station asks matters more than whether it asks. If every household re-plans against the same local nudge, all move into the same un-nudged hours: the peak *rises* to 728 kVA and the overload doubles (Table 9), the herding effect familiar from price-responsive demand. Accepting only re-plans that lower the planned overload makes the process monotone and removes the effect. Planning to 95 % of the rating absorbs most forecast error (overload 51 instead of 222 kVAh) at the price of more households asked to move (23 slightly worse off instead of 7).

Electrification makes coordination necessary. Overload grows steeply with adoption (Table 10): at 25 % EV ownership uncoordinated optimisers reach 14 176 kVAh over the winter, roughly twice the timer, and the station brings it down to 1 780 kVAh. At 50 % EVs even the timer overloads by 34 359 kVAh; the station still cuts the optimisers’ 50 300 to 8 553 kVAh, and with heat pumps 22 716 to 4 609 kVAh, without any comfort violation. The price rises with adoption and falls unevenly. At 25 % EVs the street pays €158 [59, 380] more over the winter than without coordination, with 56 [38, 109] households worse off (worst €52). At 50 % EVs it pays €974 [333, 1 460], with 126 [119, 135] households worse off (worst €81), the last-resort

Winter, 400 kVA	Seeds	Peak, uncoord. (kVA)	Overload, timer	Overload, uncoord.	Overload, station
Today (3 % EV)	5	484 [440, 520]	1 080 [654, 2 387]	2 764 [363, 7 470]	624 [258, 965]
10 % EV	3	505 [478, 568]	1 930 [1 508, 3 884]	5 316 [1 664, 14 712]	–
25 % EV	3	581 [546, 621]	7 266 [6 679, 9 338]	14 176 [7 183, 29 129]	1 780 [975, 2 835]
50 % EV	3	728 [680, 763]	34 359 [26 847, 34 391]	50 300 [41 150, 68 809]	8 553 [3 383, 11 539]
50 % of gas homes on heat pumps	3	598 [589, 631]	10 571 [6 612, 13 919]	22 716 [22 211, 37 430]	4 609 [3 801, 9 565]

Table 10: Electrification scenarios, winter, 400 kVA (overload in kVAh). EV adoption changes only who owns an EV; heat pumps (COP 3) replace gas heating in half of the non-electric homes. Station at 50 % EVs and with heat pumps: batch b3, with the corrected last-resort cap (see text).

cap is needed on 30 [4, 53] days, and EVs end the winter 17 kWh short in total over all seeds. With heat pumps, 162 [161, 172] households are worse off (worst €87). At high adoption, coordination therefore becomes a question of who pays for the shared limit: a compensation rule, or reinforcement, is needed, and it is outside this study. All high-adoption station runs use a corrected last-resort cap that never covers a running washing cycle (an earlier version made some household plans infeasible); a regression check confirmed that the correction leaves every other result unchanged.

Sensitivities. Two assumptions matter. First, if the large commercial site has its own supply rather than sharing the street’s transformer, the street never exceeds 400 kVA today (peak 365 kVA): today’s congestion, and therefore the value of coordination today, hinges on that connection, which only the network operator can confirm. Second, moving to post-reform off-peak hours, with afternoon hours in summer, lowers the summer bills of optimising households (street total €66 983 to €66 496) but raises their summer peak from 307 to 329 kVA, as cheap afternoon hours concentrate demand; timer households are unaffected.

7.4 Language-model house agents as the interface

Setup. Each electrically heated or EV-owning household of the calibrated street (seed 1, 25 % EV adoption so that enough homes own an EV) receives a resident’s note for tomorrow, generated from its true settings with varied phrasing, omissions (the agent should then abstain) and distractor sentences (a bedroom temperature at night, a partner’s petrol car). The house agent calls a single tool that records preferences *in the resident’s own units*; code converts them (e.g. a percentage of the battery) because an early test showed the model turning “80 % of a 50 kWh battery” into 50 kWh, a value no range check can reject. A deterministic validator rejects invalid output and feeds its message back (at most two retries). Prompts were tuned on a development split of households and frozen; the numbers below are on the held-out test split (77 households) with one open-weight model (`qwen2.5:7b-instruct`, 4-bit, on a single lab GPU, temperature 0). The plan for three cold days (13–15 January 2025) is then computed with each method’s preferences and executed with the household’s *true* physical routine, while thermostat and charger setpoints follow the method, so the simulator cannot silently repair a misreading; harm is scored against the true needs.

Results. The agent read 98.5 % of fields correctly (98.1 % of the items actually stated, 98.8 % correct abstentions), against 87.0 % for a regex parser written against the very templates that generated the notes, and 54.1 % for using defaults (Table 11). On twenty hand-written notes, six of them in French, the gap widens: 94.2 % against 73.3 % (stated items: 90.2 % against 37.3 %). End to end, the agent harms almost exactly as many households as a perfect reading of the notes, and far fewer than the parser or the defaults: misreading preferences, not optimisation, is what hurts residents. The validator rejected 10.4 % of first attempts (an EV time for a home without an EV, a comfort temperature out of range); after feedback 1.3 % remained invalid and

Preferences from	Field acc. (%)	Comfort harm	EV shortfall	Wrong laundry
True settings (no reading)	–	0	2	0
Perfect reading of the brief	–	7	2	5
LLM agent	98.5	8	1	7
Regex parser	87.0	11	0	23
Defaults only	54.1	35	12	71

Table 11: House agents on the held-out split: share of preference fields read correctly (including correct abstention), and number of households harmed over three cold days: below the resident’s true comfort floor, EV short of the true target at the true departure, washing on wrong or missed days. “Perfect reading” differs from “true settings” only where the note omitted an item. Model: qwen2.5:7b-instruct. Generated by `src/analysis/report_assets.py`.

fell back to defaults. One-sentence explanations passed a deterministic faithfulness check (every clock time, range, Tempo colour and euro amount must match the plan facts) in 89.6 % of cases; failures were mostly time ranges that overstate when heating runs. Re-running with a different seed reproduced 100.0 % of extractions. Each call took 0.8s on average with about 688 input tokens.

These results concern one small open-weight model on generated notes plus a small hand-written set; they show that the interface layer is feasible and where its errors come from, not that residents would write such notes.

Other models. We repeated both experiments with the frozen prompts on two more local models (Table 12): Qwen2.5-Omni-7B, a multimodal model of the same family whose text part we serve through the same interface, and Llama 3.1 8B Instruct, from a different family. Reading notes is easy for all three (97.8–98.9 % of fields); Llama generalises best to hand-written notes (98.3 % against 94.2 % for Qwen 2.5 and 89.2 % for Omni) and writes the most faithful explanations (96 %), while Omni writes the least faithful (74 %) and, out of the box, ignored the tool-call format and kept inventing the next turns of the conversation. Negotiation separates the models more than reading does, and in opposite directions: both Qwen models err on the cautious side (12 and 18 false declines, no false accept), whereas Llama errs on the permissive side, accepting 17 requests that the resident had ruled out (and skipping the tool call 5 times). All 17 of those re-plans were stopped by the deterministic resident-wish validator (0 reached a plan), and every model left the street with the same overload as perfect understanding. The choice of model therefore changes how often residents are asked or overruled, not the safety of the outcome, which rests on the deterministic gates. Llama 3.1 was downloaded for these runs only and deleted afterwards.

Model	Fields (%)	Hand-written (%)	Rejected first / final (%)	Faithful (%)	Harm: comfort / EV / laundry	Negotiation acc. (%)	False acc. / decl.	s/call
Qwen 2.5 7B Instruct (Ollama, 4-bit)	98.5	94.2	10.4 / 1.3	90	8 / 1 / 7	90–100	0 / 12	0.8
Qwen2.5-Omni-7B, text part (bfl6)	97.8	89.2	5.2 / 0.0	74	7 / 4 / 7	84–100	0 / 18	2.9
Llama 3.1 8B Instruct (Ollama, 4-bit)	98.9	98.3	24.7 / 0.0	96	7 / 2 / 6	70–100	17 / 0	1.6

Table 12: Local language models on the same frozen protocol: field accuracy on held-out generated notes and on hand-written notes, validator rejections (first attempt / after feedback), faithful explanations, households harmed over three cold days (comfort / EV / laundry), negotiation accuracy over held-out populations (seeds 3, 4, 5), false accepts / declines (every false accept was blocked by the resident-wish validator), and seconds per call on one A6000 GPU.

(Pending: same protocol with Claude Haiku and Sonnet (deferred by the author to the end of October 2026).)

Negotiation. In a second experiment each resident’s note also states how flexible they are (anything may move; never the heating; only the car; nothing). On overloaded days the station’s

model drafts a short request to each household it wants to move, the house agent accepts or declines, and two deterministic gates follow: a re-plan that changes a device the resident protected is discarded, and the station keeps only re-plans that lower the planned overload (Table 13). The first version failed instructively: the station’s model sometimes inverted its own request (“use less *outside* 5–6 AM”), and house agents read “move” as “make colder”, so only 43 % of decisions were right and the station had to send 182 requests instead of 54. Passing the request also as structured data, validating the station’s message, and stating that shifting never breaks the resident’s settings raised accuracy to 94 % on the development population. On three held-out populations it reached 90–100 %, with 0 false accepts and 12 false declines in 150 requests, and the street reached the same overload as with perfect understanding. Accepting every request instead would have tried to move a protected device 27 times; the validator caught them all.

Run	Requests	Accuracy (%)	False acc.	False decl.	“Always” false acc.	Overload LLM	Overload truth
Development (seed 1), v1 prompts	182	42.9	0	104	8	287	249
Development (seed 1), v2	54	94.4	0	3	–	249	–
Test (seed 3), v2	97	89.7	0	10	16	265	265
Test (seed 4), v2	26	100.0	0	0	0	222	222
Test (seed 5), v2	27	92.6	0	2	11	210	210

Table 13: Negotiation, hero week, 400 kVA. Accuracy: share of accept/decline decisions matching the resident’s stated flexibility. Overload in kVAh with LLM decisions and with perfect decisions (“truth”). Model: qwen2.5:7b-instruct.

8 Limitations

- **Synthetic households, simulated savings.** Building data are real; household equipment, behaviour and comfort preferences are assumptions. Savings are simulated, not metered.
- **One household archetype so far.** House 1 results come from six one-at-a-time variants, which show sensitivity, not a population distribution.
- **Simple physics.** A first-order thermal model per home, no night setback yet, no hot-water temperature stratification, and no feeder voltage limits (transformer only).
- **Average curves understate individual peaks.** Measured Enedis averages contain diversity; single-home peaks, and therefore shedding against the subscription, are optimistic.
- **Third-party Tempo calendar.** Colours come from a public aggregator, checked against season totals and known red days but not yet against RTE’s own service.
- **Secondary policy sources.** Several regulatory facts are cited from secondary sources; primary texts will replace them.
- **Business euros.** Only grid metrics are reported for shops; euro figures would need network tariffs and supplier margins that are not public.

9 Conclusion and next steps

For an electrically heated French home on Tempo, planning tomorrow’s loads with only what is known the evening before captures most of what perfect foresight could gain, and almost all of it on winter red days; outside winter a timer is as good. The open data also show that French households already respond to strong day-ahead signals, and that national signals and local

network needs do not always point the same way. On a street, that mismatch has a physical cost: households that each optimise against the same national prices overload their shared transformer more than simple timers do. A station agent can remove most of that overload cheaply, but only if it asks carefully: a monotone, accept-only-improving rule works where a naive broadcast nudge herds everyone into the next hour. Language models are useful as the interface and as the negotiator, provided deterministic validators keep them from breaking residents' wishes or the station's guarantees. Next steps: confirm the transformer and the large site's connection with the network operator, compare model families (PR-007), and test residents' acceptance with real users.

References

- [1] Fourniture de transformateurs HTA/BT pour cabines préfabriquées (160, 250, 400, 630 et 1000 kVA). Tenders Electronic Daily, notice 153885-2019, 2019. URL <https://ted.europa.eu/en/notice/153885-2019/pdf>.
- [2] Webinaire borne IRVE (Avere-France / ATEE). atee.fr, slides of 6 Mar 2024 (file "06032024 - AVERE - Webinaire Borne IRVE ATEE.pdf"), 2024. Secondary on décret n° 2017-26 as revised 4 May 2021 (smart-charging capability); verify on Légifrance.
- [3] Geographic interface for an energy simulation tool (DIMOSIM). QGIS User Conference 2026 talk, 2026. URL <https://talks.osgeo.org/qgis-uc2026/talk/WLRKDG/>.
- [4] Coupling scenario-based grid simulations with state estimation: Measurement requirements for low-voltage networks under the German energy transition pathway. arXiv:2603.28945, 2026. URL <https://arxiv.org/abs/2603.28945>.
- [5] Agence ORE. Cartographie des réseaux et infrastructures: note méthodologique (v7.4), 2025. URL https://opendata.agenceore.fr/api/datasets/1.0/reseau-aerien-basse-tension-bt/attachments/note_methodologique_cartographie_des_reseaux_et_infrastructurespdf0. Locations of HTA/BT substations and LV lines; capacities not published.
- [6] Édouard Amouroux, Thomas Huraux, François Sempé, Nicolas Sabouret, and Yvon Haradji. Simulating human activities to investigate household energy consumption. In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence (ICAART)*, 2013. URL <https://hal.archives-ouvertes.fr/hal-01852256>.
- [7] ANFA. Baromètre du véhicule électrique 2025, 2025. URL <https://www.anfa-auto.fr/sites/default/files/2025-05/BarometreVE-2025.pdf>. Electric motorisations about 3% of the vehicle fleet on 1 Jan 2025 (SDES).
- [8] Commission de régulation de l'énergie. Délibération relative à la demande de Voltalis (NEBEF), 2014. URL <https://www.cre.fr/fileadmin/Documents/Deliberations/import/140507DemandeVoltalis10mars2014.pdf>.
- [9] Commission de régulation de l'énergie. Délibération n° 2018-240 relative aux flexibilités locales (Enedis), 2018. URL https://www.cre.fr/content/download/20076/file/181024_2018-240_FlexibiliteLocalEnedis.pdf.
- [10] Commission de régulation de l'énergie. Délibération n° 2020-130 relative au périmètre de l'autoconsommation collective, 2020. URL https://www.cre.fr/content/download/22464/file/200611_2020-130_Perimetre_Autoconso.pdf.

- [11] Commission de régulation de l'énergie. Avis n° 2023-209 relatif à l'autoconsommation collective, 2023. URL https://www.cre.fr/content/download/27644/file/230726_2023-209_Avis_autoconsommation_collective.pdf.
- [12] Commission de régulation de l'énergie. Annexe au communiqué de presse TURPE 7 – heures pleines / heures creuses, 2025. URL https://www.cre.fr/fileadmin/Documents/Communiques_de_presse/2025/250206_Annexe_CP_TURPE_7_HPHC.pdf.
- [13] Commission de régulation de l'énergie. FAQ – évolution des heures creuses (TURPE 7), 2025. URL https://www.cre.fr/fileadmin/Documents/Communiques_de_presse/2025/250206_FAQ_HP-HC.pdf. 6 Feb 2025.
- [14] Commission de régulation de l'énergie. Historique des tarifs réglementés de vente d'électricité pour les consommateurs résidentiels. data.gouv.fr, n.d. URL <https://www.data.gouv.fr/datasets/historique-des-tarifs-reglementes-de-vente-de-lelectricite-pour-les-consommateurs-residentiels>. Base, HP/HC and Tempo options; retrieved 22 Oct 2026.
- [15] Connaissance des Énergies. En quoi consiste l'option Tempo?, n.d.. URL <https://www.connaissancedesenergies.org/questions-et-reponses-energies/fourniture-de-lelectricite-quest-ce-que-loption-tempo>. Reports nearly one million residential Tempo customers (CRE, January 2025).
- [16] Connaissance des Énergies. Qu'est-ce que l'effacement de l'électricité?, n.d.. URL <https://www.connaissancedesenergies.org/questions-et-reponses-energies/quest-ce-que-leffacement-de-lelectricite>.
- [17] Jochen L. Cremer. Customising electricity contracts at scale with large language models. *IEEE Transactions on Power Systems*, 2026. URL <https://arxiv.org/abs/2505.19551>. arXiv:2505.19551.
- [18] CSTB. Base de données nationale des bâtiments (BDNB). Open API, api.bdnb.io, n.d. URL <https://api.bdnb.io/v1/bdnb/>. Retrieved 21 Oct 2026.
- [19] Effy. Boîtier Voltalis: comment fonctionne ce dispositif gratuit?, n.d. URL <https://www.effy.fr/chauffage/autres-travaux/thermostat-connecte/voltalis>. Secondary; reports short cuts of radiators during national peaks and an ADEME estimate of 7–8% savings.
- [20] Reda El Makroum, Sebastian Zwickl-Bernhard, and Lukas Kranzl. Agentic AI home energy management system: A large language model framework for residential load scheduling. arXiv:2510.26603, 2025. URL <https://arxiv.org/abs/2510.26603>.
- [21] Enedis. Le raccordement au réseau Enedis (diaporama, 30 mars 2023), 2023. URL https://centre.energie-partagee.org/partagefichiers/LIFE/20230330_Diaporama_Raccordement-ENEDIS.pdf. BT up to 250 kVA, HTA above.
- [22] Enedis. Enedis open data: conso-inf36, conso-sup36, consommation-annuelle-residentielle-par-adresse. opendata.enedis.fr, n.d.. URL <https://opendata.enedis.fr/>. Retrieved 21–22 Oct 2026.
- [23] Enedis. Proposition technique et financière – raccordement d'une installation de puissance comprise entre 36 et 250 kVA au réseau public de distribution basse tension, n.d.. URL <https://www.enedis.fr/media/1980/download>.
- [24] ENTSO-E. ENTSO-E transparency platform: Day-ahead prices (A44), bidding zone France, n.d. URL <https://transparency.entsoe.eu/>. Retrieved 22 Oct 2026.

- [25] European Commission Joint Research Centre. Photovoltaic geographical information system (PVGIS) 5.3, n.d. URL https://re.jrc.ec.europa.eu/pvg_tools/. Hourly series 2005–2023; retrieved 21 Oct 2026.
- [26] Fournisseur-Energie. La CRE plaide pour des tarifs réglementés plus flexibles, n.d. URL <https://www.fournisseur-energie.com/actualites/cre-tarifs-reglementes-flexibles/>. Secondary: reports CRE counts for 2023 (about 10 M base, 9 M HP/HC, 0.6 M Tempo, 0.3 M EJP).
- [27] Fournisseurs-Electricite.com. Ecowatt: le signal de RTE et de l’ADEME, n.d. URL <https://www.fournisseurs-electricite.com/suivi-conso/ecowatt>. Secondary; primary: monecowatt.fr.
- [28] Gouvernement français. Délestage électrique, n.d. URL <https://www.info.gouv.fr/grand-dossier/reduire-notre-consommation-denergie/delestage-electrique>.
- [29] Hello Watt. Jours rouges Tempo: comment sont-ils calculés?, 2026. URL <https://www.hellowatt.fr/blog/criteres-jours-rouges-tempo/>. Secondary description of RTE’s colour-selection rules.
- [30] Qi Huangfu and J. A. Julian Hall. Parallelizing the dual revised simplex method. *Mathematical Programming Computation*, 10(1):119–142, 2018.
- [31] JeChange. Heures creuses: horaires d’été 2026, une deuxième vague, 2026. URL <https://www.jechange.fr/energie/electricite/news/heures-creuses-horaires-ete-2026-9-3-millions-foyers>. Secondary (press).
- [32] Sokipriala Jonah, Queen Moses, Abiola Babatunde, Michael Ajao-Olarinoye, and Daniel Bammeke. LLMs for agentic home energy management. arXiv:2607.04569, 2026. URL <https://arxiv.org/abs/2607.04569>.
- [33] Jour Tempo. Historique des couleurs Tempo depuis 2014. data.gouv.fr (third-party aggregator), n.d. URL <https://www.data.gouv.fr/datasets/historique-des-couleurs-tempo-depuis-2014-jours-bleus-blancs-et-rouges>. Third-party; to cross-check against the RTE Tempo API.
- [34] Journal de l’Automobile. Une année 2025 record pour la mobilité électrique, 2025. URL <https://journalauto.com/services/une-annee-2025-record-pour-la-mobilite-electrique/>. Avere-France: more than 1.5 million battery-electric vehicles in circulation.
- [35] Journal de l’Économie. Avec Linky, Enedis va automatiquement débrancher le chauffe-eau à midi cet hiver, 2023. URL <https://www.journaldeleconomie.fr/avec-linky-enedis-va-automatiquement-debrancher-le-chauffe-eau-a-midi-cet-hiver/>. Secondary; order of 6 Sep 2023 (verify on Légifrance).
- [36] Steffen Meinecke et al. SimBench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis. *Energies*, 13(12):3290, 2020.
- [37] Open-Meteo. Open-meteo previous runs and historical weather (ERA5) APIs, n.d. URL <https://open-meteo.com/>. Retrieved 21 Oct 2026.
- [38] Nicolas Perez. Contribution for district energy system design: Simulation, optimization and decision support. PhD thesis, La Rochelle Université (theses.fr 2017LAROS017); describes the CSTB DIMOSIM platform, 2017. URL <https://theses.fr/2017LAROS017/document>.

- [39] Photovoltaïque.info. Lancement d’appels d’offres flexibilités locales par Enedis, 2025. URL <https://www.photovoltaïque.info/fr/actualites/detail/lancement-dappels-doffres-flexibilites-locales-par-enedis/>.
- [40] RTE. Bilan prévisionnel 2023–2035: principaux résultats, 2023. URL <https://www.connaissancedesenergies.org/sites/default/files/pdf-actualites/Bilan-previsionnel-2023-principaux-resultats.pdf>.
- [41] RTE. eCO2mix national and regional consolidated data. ODRE open data platform, n.d. URL <https://odre.opendatasoft.com/>. Retrieved 21 Oct 2026.
- [42] Seban Avocats. Heures creuses méridiennes: délibération de la CRE n° 2026-33 du 4 février 2026, 2026. URL <https://www.seban-associes.avocat.fr/?p=127964>. Secondary source for the 2026 midday off-peak decision; verify on cre.fr.
- [43] Selectra. Linky: relais virtuels pour chauffe-eau, borne et pompe à chaleur, n.d. URL <https://selectra.info/energie/actualites/marche/linky-relais-virtuels-chauffe-eau-borne-pac>. Secondary.
- [44] Leon Thurner et al. pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Transactions on Power Systems*, 33(6):6510–6521, 2018.
- [45] José R. Vázquez-Canteli, Sourav Dey, Gregor Henze, and Zoltán Nagy. CityLearn: Standardizing research in multi-agent reinforcement learning for demand response and urban energy management. arXiv:2012.10504, 2020. URL <https://arxiv.org/abs/2012.10504>.
- [46] Xudong Wu, Zeqing Wu, Jiarui Zhang, Xuhao Fan, Ziang Ding, Yuming Zhuang, Mingqi Yuan, Yilun Du, Hongjie Jia, Yunfei Mu, and Jiayu Chen. EnergyBridge: Benchmarking household energy management, user participation, and grid flexibility. arXiv:2608.08691, 2026. URL <https://arxiv.org/abs/2608.08691>.

A Reproducibility

Code, configurations and run outputs live in the project repository (`src/`, `configs/`, `runs/`); experiments run on a lab server in detached sessions with a pinned Python environment (`src/requirements.txt`). Every fetcher saves raw responses unchanged with a manifest; every run writes its configuration, a manifest with the SHA-256 of every source file, and its self-check result.

Result	Run folder
House 1, full window (Tables 5, 6)	<code>runs/20261022T074506Z_martins_season_v2</code>
House 1, hero week	<code>runs/20261022T074504Z_martins_hero_v2</code>
Refactor regression check	<code>runs/20261022T085116Z_martins_regress</code>
Data evidence (Section 7.1)	<code>data/processed</code> (hashes in <code>generated/evidence_macros.tex</code>)

Generated report assets are rebuilt with `python -m src.analysis.report_assets house1 <run>` and `python -m src.analysis.data_evidence`.