
2026 BUILD FOR NEXTGEN · TAIPEI

AIRO

Can a building explain itself?

Physics-grounded AI for building energy decisions

SEOUL × TAIPEI

Jeong-Uk Kim · Sangmyung University



01

PART ONE

The gap between data and decision

AIRO

Energy Decision Intelligence

Monitoring answers what already happened.

A decision needs why — and what to try next.

AIRO is built for that second half.

Monitoring and decision-making are different jobs

TODAY

A dashboard reports conditions

- Collects meter and sensor data
- Visualises trends and comparisons
- Raises an alarm when a limit is passed
- Answers: what already happened

MISSING

An operator still has to decide

- Which cause explains the increase
- Which intervention to test first
- How much it is worth, and how certain
- Answers: what to do next

TAKEAWAY

A dashboard says a value is high, not which fix to try first.

The building you designed is not the one you run

AT DESIGN

Assumptions are fixed

- Standard occupancy and schedules
- A design weather year
- As-built control settings
- One static energy model

IN OPERATION

Assumptions drift

- Tenants and use patterns change
- Weather distribution shifts over years
- Setpoints are edited locally
- Nobody rebuilds the model

TAKEAWAY

A static design model cannot explain a building that keeps changing.

Every energy decision begins with three questions

Q1

What is happening?

Estimate the current performance state, normalised for type, size, age and climate.

Baseline prediction model

Q2

Why might it happen?

Infer which operating and control configurations could plausibly produce it.

Reverse identification model

Q3

What should we try?

Compare counterfactual scenarios and rank the next action by effect and cost.

Simulation + policy evaluation

TAKEAWAY

Diagnosis is useful only when it ends in a testable intervention.

02

PART TWO

One question, five explicit stages

AIRO

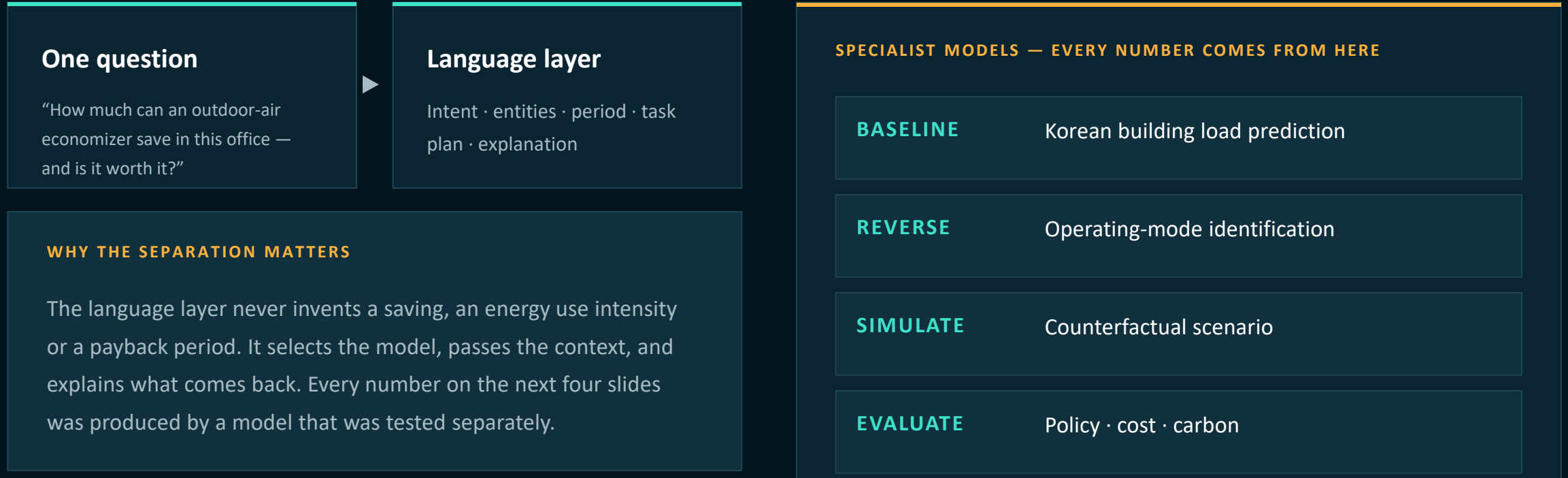
Energy Decision Intelligence

Language interprets and plans the work.

Specialist models produce every number.

Each stage can be inspected on its own.

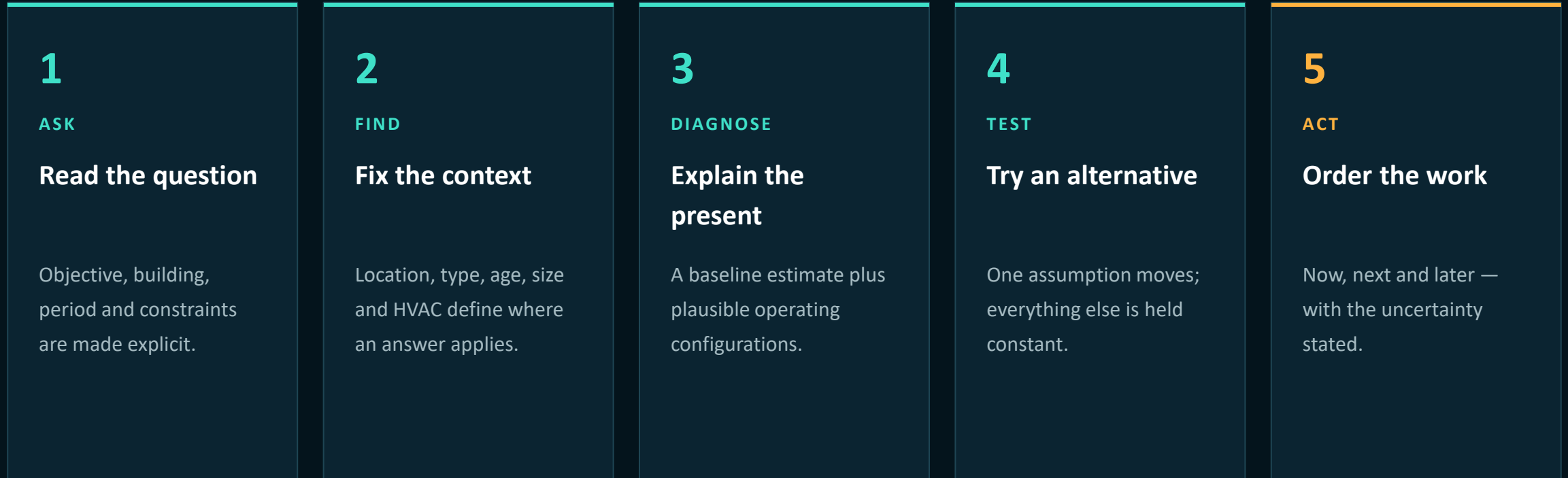
One entry point, several specialist models



TAKEAWAY

One conversation on the surface. Validated tasks underneath.

One question, five separately checkable stages



TAKEAWAY

Every stage has a defined job — so a defined failure mode.

FIND — 7.29 million buildings, then the right few



3D basemap: NGII precision road map · VWorld (MOLIT) · KOGL Type 1

1 Real building stock, in 3D

National precision basemap — 17,589 buildings in this view alone.

2 Colour is a model output

Here, building age band. The same view can carry energy intensity or carbon.

3 Context becomes the evidence trail

Location, type, age, size and HVAC carry forward unchanged into every later stage.

TAKEAWAY

Fixing the context stops the model from inventing the building.

DIAGNOSE — observation to labelled hypothesis

LIVE OUTPUT · 2026-08-08

POST :8030/v1/qa/diagnosis

```
{ "city":"Taipei", "building":"medium_office", "hvac":"A", "cooling_setpoint":26, "lang":"en"
}
```

The medium office in Taipei is operating with a 26°C cooling setpoint, which is relatively high for a tropical climate. The building is equipped with a VAV system that includes chillers and boilers, which is a common but energy-intensive HVAC strategy for medium-sized office buildings in

Taipei.

model exaone_insight_lat_2026-07-10 · 3.85 s · no external LLM

Estimated, not measured

Hidden states — setpoint discipline, control logic, sensor health — are inferred, and the interface says so.

Narrow the search

A useful diagnosis reduces what an engineer must inspect from everything to three things.

Something to check, not a conclusion

AIRO did not open the building. It read the data and named the most likely cause. An engineer still has to confirm it on site.

TAKEAWAY

It tells you where to look first. It does not tell you what is true.

TEST — the alternative, before anyone commits

MACC lowest-cost measure = Recommissioning (net saving (profitable))

Recommissioning **Marginal cost -¥224K/tCO₂ · Payback 1.1 yr** Capex ¥20M · Annual saving ¥18M · NPV ¥59M

LED lighting **Marginal cost -¥193K/tCO₂ · Payback 4.1 yr** Capex ¥75M · Annual saving ¥18M · NPV ¥154M

BEMS optimal control **Marginal cost -¥175K/tCO₂ · Payback 3.7 yr** Capex ¥30M · Annual saving ¥8.0M · NPV ¥41M

Air-sealing **Marginal cost -¥114K/tCO₂ · Payback 6.9 yr** Capex ¥50M · Annual saving ¥7.2M · NPV ¥25M



CHEAPEST FIRST MOVE

Recommissioning

1.1 yr

payback · -¥224K per tCO₂

Change one thing

One control, schedule or retrofit assumption moves.

Rank by money and carbon

Cost per tonne, payback and NPV on one axis.

State the interaction

Applied together, savings overlap — total < sum.

Live UI capture 2026-08-08 — Energy Decision Studio, English locale — measure economics (F16) on policy evaluation (F14)

TAKEAWAY

This is a simulation-based estimate — not guaranteed field savings.

ACT — advice is an order of work, not an adjective

NOW

Verify before you spend

Schedules, setpoints and data quality. No capital, no procurement, this week.

NEXT

Test the leading cause

One targeted counterfactual scenario for the most plausible explanation.

LATER

Measure it in the building

Implement the selected measure and verify the saving against a baseline.

AND EVERY ANSWER IS CROSS-CHECKED

Pass · Code

Pass · Carbon

Warn · Safety

Pass · Data

Four independent checks run on the answer itself — building code and physics, carbon accounting, operational safety, and data provenance. "Pass" means the answer breaks no rule; it is not a claim that the value was measured.

TAKEAWAY

Useful advice is ordered, testable and honest about uncertainty.

03

PART THREE

What is proven, and by whom

AIRO

Energy Decision Intelligence

Three modules passed accredited testing.

The evidence base is large — and it is simulation.

Boundaries are stated, not implied.

Language plans the work; physics answers it

THE LANGUAGE LAYER DOES

Interpretation and orchestration

- Extract intent, building, period, objective
- Translate the question into model calls
- Explain returned evidence in plain language
- Adapt register — engineer, owner, regulator

THE LANGUAGE LAYER NEVER DOES

Producing a number

- Savings percentage or kWh
- Energy use intensity or peak load
- Payback period, NPV, cost per tonne
- Any figure a reader could act on

TAKEAWAY

Fluent language is not evidence. Numbers come from models.

Four grounding layers tie an answer to its evidence

01 Building context

Where the answer applies

7.29M building records · type, age, area, HVAC, climate zone

02 Simulation & surrogates

What the physics implies

620,000+ simulation cases · learned surrogates for fast evaluation

03 Evaluation & policy logic

What it means for a decision

Cost, carbon, comfort and code turned into decision quantities

04 Traceable evidence path

How the answer was produced

Which model, which context, which assumption — recoverable per answer

TAKEAWAY

A recommendation is credible only if its evidence path is walkable.

Three modules, tested by an accredited laboratory

KOLAS · ILAC-MRA — test report 26-10050 · KS X ISO/IEC 25023 · ISO/IEC TS 4213

MODULE TESTED	METRIC	RESULT	CRITERION	VERDICT
Load prediction Korean building baseline	CVRMSE 955-building public cohort: 12.91%	12.55%	≤ 13.27%	PASS
Operation identification Reverse EMS model	Macro-F1 (active bits) 2,989 held-out simulation-derived pairs	0.85	≥ 0.80	PASS
Natural-language diagnosis Explanation model	Required-field coverage character BLEU 0.22 (≥ 0.15) on 200 Q&A	0.84	≥ 0.80	PASS

TAKEAWAY

Accredited for three modules — not for AIRO as a whole.

Simulation at a scale that allows learning

620,000

+

CUMULATIVE BUILDING-ENERGY SIMULATIONS

Enough controlled cases to train surrogate models and to run a counterfactual repeatedly, rather than once.

Korean building archetypes	17 types
----------------------------	----------

HVAC variants × control modes	7 × 16
-------------------------------	--------

Weather years, incl. future climate	2050 · 2080
-------------------------------------	-------------

Held-out pairs for reverse testing	2,989
------------------------------------	-------

This is the size of the simulation base — not 620,000 measured buildings.
Climate-change scenarios are included in the count.

TAKEAWAY

Scale is what makes the counterfactual cheap enough to ask twice.

Three things building AIRO taught me

01 Fluency is not evidence

The most convincing wrong answer we ever produced was grammatically perfect. Separating language from measurement was not a preference — it was the fix.

02 Localisation is not translation

Climate, archetypes, construction, schedules, tariffs and control practice all move the number. Translating the interface changes none of the physics.

03 Validation cannot be done from one country

We can find where our assumptions break only with local engineers and local buildings. That is the honest limit of what Korea alone can prove.

TAKEAWAY

The next step is localisation and validation — not a bigger model.

Thank you

Jeong-Uk Kim · Sangmyung University, Seoul