FRANCE · COMPUTING · PROJECT 06 · EUCYS 2026, KIEL

What we learned **after** we submitted.

Nothing on this panel existed in the June report. Each block is one observation from real use, what we changed, and what we measured. **Some of it contradicts what we expected.**

01 · THE ANSWER BECAME UNWRITABLE

4 of 6

Our report's mechanisms now run on the phone at this stand.

In June the refusal was an instruction in the prompt. **Since September** it is compiled into the decoder: every spelling of the answer is removed from the grammar the model writes in, so the solution cannot be generated. Four of the six mechanisms of our report run on this phone, each with a switch to turn it off. **Not in the published app.**

02 · THEIR OWN NOTES, READ ON THE DEVICE

102 real pages

On-device handwriting recognition. The photo never leaves the phone.

Asked for by students: feed the tutor their own notes, not only the school feed. **Built:** the phone finds the page, straightens it, runs two recognition passes (with and without language correction, so formulas survive) and keeps the structure: titles, paragraphs, lists. The recognised lesson joins the tutor's index like a school lesson. **Tested** on 102 real notebook pages and revision sheets, cursive included; decorated sheets and small photos still fail.

03 · THE PHONE IS THE LIMIT

2 to **3** GB

On an ordinary phone, it is slow.

Observed: students with older or cheaper phones asked why the tutor took so long. **Measured:** a 2-to-4-billion-parameter model needs 2 to 3 GB of storage and a phone with about 8 GB of RAM; below that, generation is several times slower and the phone warms up. **Changed:** since August the published app sends generation to a server by default, so that phones which cannot run the model still get the tutor. **Next:** our own secured servers in Europe, not built yet. A bridge, not the invention.

04 · IT CAN BE WRONG

45 of 395

Refused rather than invented.

Observed: a small model makes more mistakes than a large one, and in a classroom a confident mistake is worse than silence. **Design:** every turn is **grounded in the student's own lessons and a textbook**, and quotes its source; with no source, the tutor refuses to generate. **Measured:** 45 refusals out of 395 requests in the first week, every one an empty school feed. **Grounding reduces errors. It does not remove them.**

05 · EFFORT, REWARDED

Work that feels lighter

Asked for by students, added after launch.

Observed: students used the tutor once and did not come back to a plain study tool. **Changed:** the app now rewards **effort only**, never marks: a homework done, a quiz submitted whatever the score, a focus session with distracting apps blocked, with daily caps. **Hypothesis:** work feels lighter, so students do more of it. **Not measured yet:** whether they learn more.

06 · ONE ADAPTER PER COUNTRY

17 schools in Kiel

Kiel's own school system, WebUntis, answers our connector.

Every school system is mapped onto one shared data model, so adding a country means adding one adapter. Three French systems are in production. **Since mid-September**, WebUntis (Germany, Austria) and ClasseViva (Italy) answer our connector: the transport is verified, **no student account has been connected yet**. Next: Moodle, the one platform used everywhere in Europe.

07 · THREE LIMITS WE HIT, AND DID NOT SOLVE

Where the science is not finished.

Wording, not meaning.
The grammar removes every spelling of the answer, not its meaning. A paraphrase can still leak. We measure the residual at the stand, with the constraint on and off.

Constraints cost fluency.
Decoding inside a grammar makes the model slower and sometimes clumsier; the literature measures the same trade-off (CRANE, ICML 2025). We report the cost next to the benefit.

Heat and battery.
A few minutes of generation warms the phone and slows it down. We pause between runs to measure the model, not the throttling. Battery cost: not measured.

08 · WHAT WOULD CHANGE OUR MINDS

Three experiments we have not run, and what each would prove.

It is the training, not the locality.
Run the identical blind protocol on our tutor hosted remotely. If it ties, locality is a privacy and cost argument, not a pedagogical one, and we say so.

Graders reward style, not pedagogy.
Add a control arm: a cloud model told to always answer with a question. If it also scores about 4.5, our metric measures form, not effect.

Perceived retention is not retention.
A randomised pre/post test with delayed transfer and a control group. This is the study we have not run.

Why this panel exists. A report is a snapshot, and a project is not. Every number here was measured by us; ask us about number 01. · lucid.lucarnepro.fr/eucys

LUCID