AURIFEX HEALTHCARE

Getting hospital patients to the right next place, sooner

Evaluated on 25,000 held-out historical admissions; not randomized and not independently validated. Patients often stay in the hospital after they're medically ready to leave because the next step isn't arranged in time. I co-founded Aurifex to address that inside Epic: AI flags likely inpatient rehab (IRF) candidates and drafts the referral, and a person approves every step. It covers IRF referrals so far, not yet SNF or home health.

ROLE
Chief Product Officer & Co-Founder, AI/ML · 2025 to present

FOCUS
Model, agents, evaluation and safety design

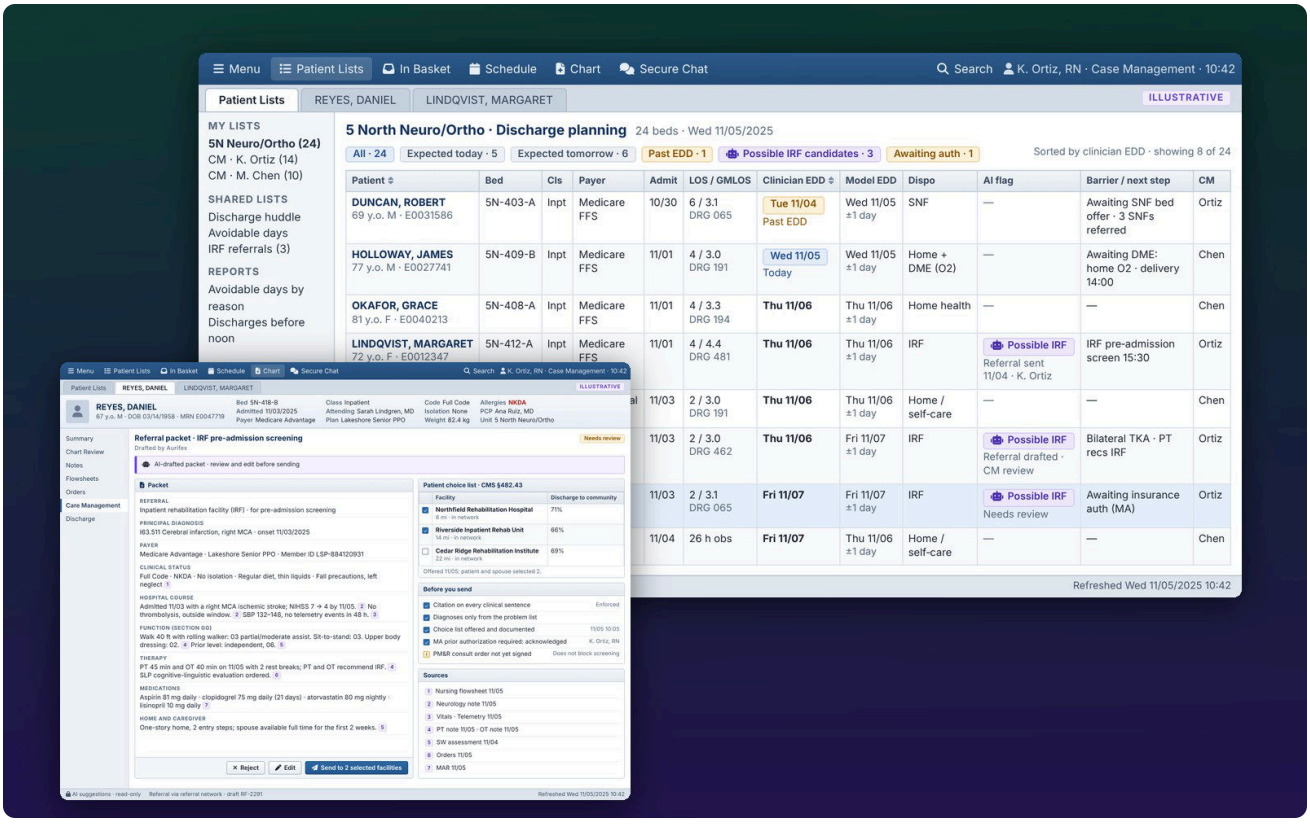
I OWNED
Product vision, model and agent design, evaluation, pricing and go-to-market, from concept to working product

Aurifex website **profio1997.github.io/Aurifex_Healthcare_Website/** · Aurifex on LinkedIn **www.linkedin.com/company/aurifexhealthcare/**

85%+
discharge dates within one day (from day 2), held-out historical admissions

0.91
AUROC for rehab-candidate flags (88% recall, 80% precision)

45 → 10
minutes per IRF referral in workflow testing on simulated cases



Overview

I co-founded Aurifex Healthcare and lead product as Chief Product Officer. Aurifex predicts each patient's expected discharge date, flags likely candidates for acute inpatient rehab (IRF) in the case-management worklist, and drafts the referral packet and the PM&R consult request for clinicians to approve. It's designed to run inside Epic and pairs a gradient-boosted model with GPT-4o agents.

HOW SUCCESS IS MEASURED

Outcomes
Length of stay, excess days and IRF acceptance, each against the unit's own baseline and adjusted for case mix.

Safety
Readmissions and emergency returns as balancing measures, so a shorter stay can't hide a worse outcome.

Real-world accuracy
Flag precision on the unit's own patient mix, and how the worklist is used on a busy morning.

The problem

1

Case managers
They plan every discharge on the unit and write the referrals. A good rehab referral means reading days of notes for evidence that the patient can handle intensive therapy, which took about 45 minutes per referral in workflow testing.

2

Physicians and nurses
They decide when a patient is medically ready, but the discharge plan often starts after that point instead of before it. They need the expected date early enough to act on it.

3

Hospital operations
They carry the cost of every day a patient stays past need, and they need a view of the whole unit, not one patient at a time.

I chose discharge and placement over the many other places AI could help a hospital because the cost of delay is concrete, the people who fix it are a small, identifiable group, and the work that slows them down, reading notes and writing referrals, is what language models do well when someone checks their work. It also has an outcome a hospital already measures: length of stay and excess days, adjusted for how sick the patients were.

What I decided

Agents draft, people decide
Agents review each admission, flag possible IRF candidates and draft the referral packet and the PM&R consult request. A case manager approves, edits or rejects every referral before it's sent, and the attending signs the consult.

Trade-off: slower than sending automatically, and every draft needs a person's time. I took that trade because one wrong referral sent in a clinician's name would cost more trust than the minutes it saved.

Not chosen: auto-sending referrals that pass the automated checks.

Build inside Epic, not beside it
Flags, dates and drafts are designed to appear in the Epic worklist case managers already open every morning, so using Aurifex is part of the job rather than one more login.

Trade-off: slower integration and a dependence on each hospital's Epic analysts. The gain: nothing new to learn on a busy unit.

Ruled out: a separate dashboard, which would have been faster to build and easier to ignore.

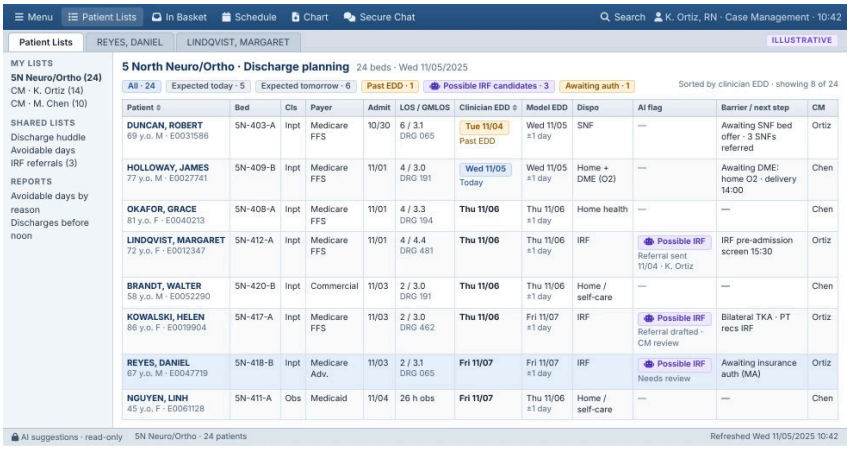
Three smaller calls

- **Solve one decision.** The whole product aims at where a patient goes next and when, which gives Aurifex one buyer, one outcome and one group of users to win over. *Passed on:* a general assistant for the whole care team.
- **The right model for each job.** A gradient-boosted model predicts the discharge date from structured data, where it can be checked against the real date; GPT-4o agents handle the free-text notes. *Instead of:* one language model for everything, including the date.

- **A source for every sentence.** Every statement in a summary or draft must cite the note it came from. *Rejected:* free-form summaries, which are faster to generate and slower to check.

What it does

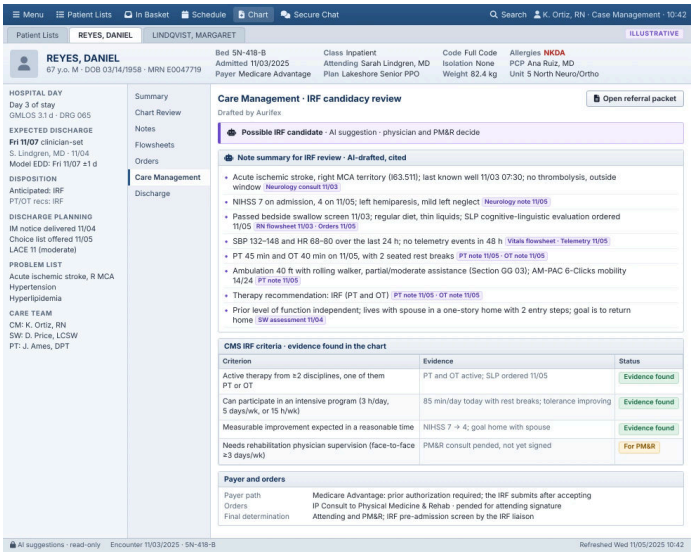
The case manager's day starts in the worklist. Aurifex adds its expected discharge date beside the clinician-set date, which it never overwrites, plus the likely next place of care and a flag on possible IRF candidates, so the morning huddle starts with the patients who need a plan today.

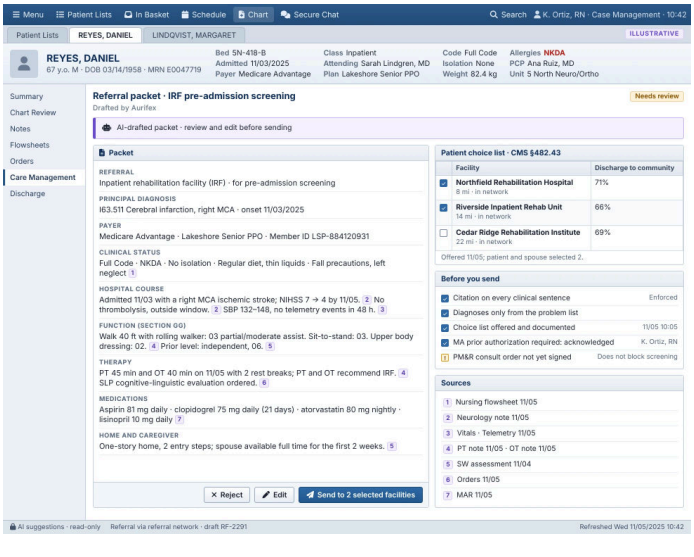


Illustrative.

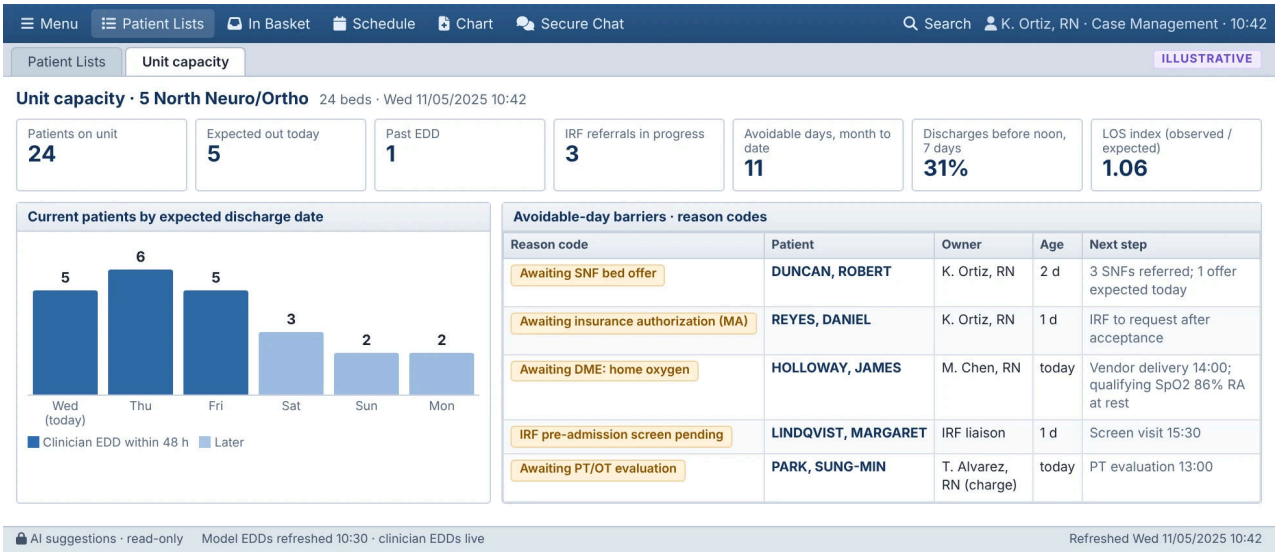
For a flagged patient, the agents have already read the notes and drafted the referral packet with numbered sources. The PM&R consult is a consult request the attending signs, not a PM&R assessment. The case manager reads the packet instead of days of notes, fixes anything wrong, offers the patient a choice of facilities and sends it to the ones they pick.

The rest of the work stays with people. If a patient turns down a facility, the case manager updates the choice list and resends. Medicare Advantage authorizations show as a barrier with their age. If an IRF declines, the packet goes to the next facility the patient chose. Denials, peer-to-peer reviews and appeals stay with the hospital's utilization review team.





Illustrative.



Illustrative.

Evaluation results

The test set is 25,000 de-identified historical admissions held out from training, plus expert-labeled reference cases and simulated referral work.

LIMITS

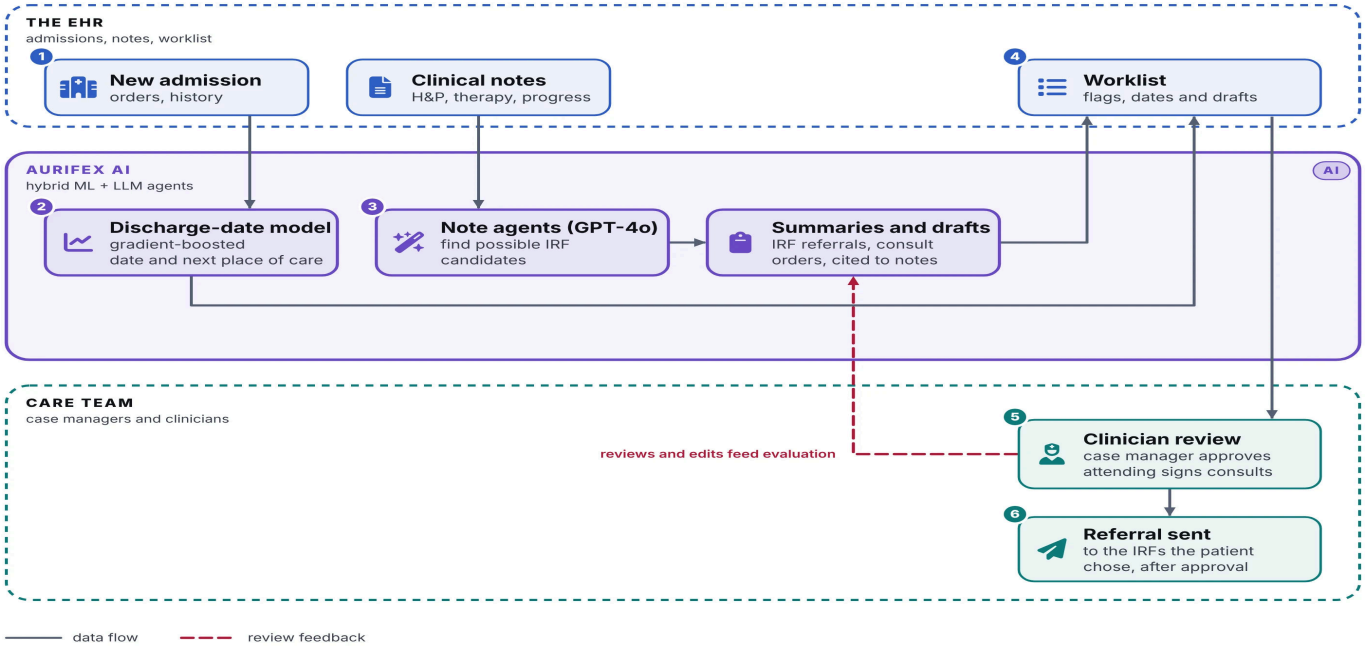
Design
Not randomized and not independently validated.

Still open
Whether discharges before noon and emergency department boarding change, and how much staff time Aurifex takes beyond referral prep.

WHAT WAS TESTED	RESULT	HOW IT WAS MEASURED, AND HOW TO READ IT
Expected discharge date	**85%+**	85.3% (95% CI 84.9–85.7%) within one day of the actual discharge on 25,000 held-out historical admissions, measured from day 2 of the stay. Mean absolute error 0.6 days. Clinician-entered expected dates on the same admissions, where recorded (n = 14,200): 58% within one day.
Next place of care	**82%**	The predicted next place of care matched the actual one (top-1) on held-out admissions.
Rehab-candidate flags	**AUROC 0.91**	On 2,000 reference admissions labeled independently by 3 PM&R physicians and 2 case managers (kappa 0.82; 12% candidates). AUROC 95% CI 0.89–0.93. At the worklist threshold, 88% recall and 80% precision; 95% overall agreement with the labels. Precision depends on how common candidates are, so a unit with a different mix will see a different figure.
Drafted PM&R consults	**85%**	400 drafts rated acceptable as written or with minor edits in blinded review by 6 clinicians (3 PM&R physicians, 3 case managers), two per draft (kappa 0.78); 95% CI 81–88%. Acceptable isn't the same as correct, so the type of each edit is reviewed too.
Hallucination rate	**1.5%**	1.5% of 6,000 drafted statements (95% CI 1.2–1.8%) unsupported by their cited source. Counted per statement, so not comparable to UW Health's per-draft rate. Every statement cites its source note, an automated check runs first, and reviewers then check that the cited text supports the statement.
Referral prep time	**45 → 10 min**	Median time per IRF packet across 60 simulated cases timed with 6 case managers, including the reviewer's check. IRF packets only.
Model cost	**$0.12**	Average model cost per patient, excluding hosting.

WHAT WAS TESTED	RESULT	HOW IT WAS MEASURED, AND HOW TO READ IT
Draft time (sandbox replay)	**< 30 s**	Time to draft a referral packet.
Worklist update (sandbox replay)	**< 5 min**	From an admission event to an updated worklist.

How it works

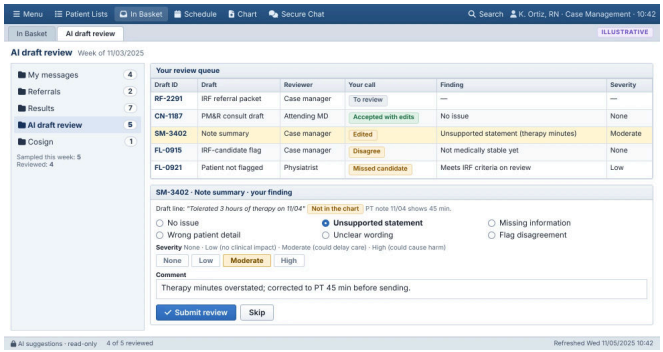


Illustrative.

Two kinds of AI do two jobs. The discharge date comes from a gradient-boosted model trained on historical admissions and designed to be refit on each hospital's own history, because the inputs are structured and the answer can be checked against the real discharge a few days later. Dates are made on admission and revised as notes arrive. The note work uses GPT-4o agents, because the evidence an IRF wants (how far a patient walks, how much therapy they tolerate, who is at home) lives in free text. The agents use the CMS coverage criteria as a guide. The CMS-13 list sets an IRF's compliance threshold, not a patient's eligibility, so patients with other diagnoses can be flagged. Flags sit in worklist columns, with no pop-up alerts at the bedside.

How it's checked

Evaluation covers what a case manager sees, not only the model. Every change must pass a 1,200-case evaluation gate before it merges: accuracy, source checks, 40 prompt-injection tests, and no more than a 3-point accuracy gap across age, sex and race groups.



Illustrative.

Guardrails built into the product

RESPONSIBLE AI

A person approves
Agents read, flag and draft. Referrals wait for a case manager, and consults wait for the attending's signature.

Decision support, not a decision
Clinicians see the basis for every flag and date. Self-assessed against the Cures Act criteria for non-device clinical decision support; legal review comes before any hospital rollout.

Privacy and audit
A BAA is signed before any patient data is shared, data is encrypted in transit and at rest, each agent sees only what its task needs, and agent actions are logged.

Fair across patient groups
Flags and dates are checked by patient group (ACA Section 1557 applies). In testing, the largest accuracy gap across age, sex and race groups was 3 points or less (smallest group 900 admissions), and results go to each hospital's compliance team.

Build and evaluate

Aurifex went from concept to a working product in about 6 months, with 300+ changes merged behind the evaluation gate. It's built with Claude Code, and every change goes through a GitHub pull request with automated tests and the evaluation gate. For go-to-market I also built the company website.

A hospital rollout is designed to run about 20 weeks on one or two units:

1. **Connect** — Epic FHIR access and the ADT feed
2. **Test on history** — The date model is checked on the hospital's own past discharges
3. **Silent mode** — Predictions and drafts run on live data, unseen, and are compared with what happened
4. **Go-live** — On one or two units, once the launch gates are met and the hospital signs off
5. **Weekly review** — Sampled drafts, flags and misses, reviewed with the hospital

What I'd do differently

- **Test in real workflows early.** Simulated cases time referral prep well, but only a busy unit shows how the worklist gets used, so live-workflow measures belong in the plan from the start.
- **Design the balancing measures in from the start.** Readmissions and emergency returns are the check that a shorter stay isn't a worse outcome. I'd define them, and how they're reported, before anything goes live, so a length-of-stay result always arrives with its safety check.
- **Scope skilled nursing and home health earlier.** IRF was the right first referral because it is evidence-heavy and well defined, but SNF and home health are most of a case manager's referrals. I'd have started that design work sooner.
- **What carries over.** The evaluation gate, the cited-draft review pattern and the check for unsupported statements apply to any AI that drafts work a person must approve.