// WHITE PAPER · PUBLISHED 25 SEPTEMBER 2026

MARA: A Medical Agentic Resident Assistant That Learns From Every Encounter and Never Makes the Decision

MAX pointed at healthcare: an assistant for every person in a patient's care that follows outcomes across a lifetime and a whole population, so care can cost less, work better and reach more people, while a person makes every decision.

Jason Newell · MAX Research Collective · Report WPR-HLC-001 · Published 25 September 2026
Licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/
Original: https://www.maxresearchcollective.com/articles/mara-on-max.html
Cite as: Newell, J. (2026). MARA: A Medical Agentic Resident Assistant That Learns From Every Encounter and Never Makes the Decision. MAX Research Collective, WPR-HLC-001. https://www.maxresearchcollective.com/articles/mara-on-max.html

Healthcare already knows more than it can use. Every visit, prescription, test, claim and recovery is written down somewhere, and almost none of it comes back to help the next patient. MARA is a design for changing that. The name stands for Medical Agentic Resident Assistant. It is the healthcare form of MAX, the self-learning, curious, graph-based assistant this collective builds, and it is designed around a single rule: MARA provides insight and recommendations drawn from all the data it can see, and a person makes every decision.

The hope is large and it is specific. Care that costs less, because nothing is asked, keyed, imaged or denied twice. Outcomes that are better, because every recommendation is graded by what actually happened to the patient rather than by whether someone clicked accept. Care that reaches more people, because the nurses, doctors and staff who give it spend their hours on patients instead of on re-work. And a record that follows a person from birth to the end of life, so that each person's outcomes, followed across a lifetime and read across a whole population, become the evidence that improves care for the next person like them.

You can see it today. The project page sets out the design, the goals and the scale of the work: https://www.maxresearchcollective.com/mara.html. The interactive demo walks one fictional patient through every role in a single episode of care, from the phone call to the insurer's decision and the follow-up weeks later: https://www.maxresearchcollective.com/demos/mara-demo.html. The rest of this paper explains what MARA is, what the demo shows, what it is meant to change for patients and for the people who care for them, and what it will take to build.

What the name means

Each word in the name is a design commitment.

- Medical: it works in clinical care and in everything that surrounds it, including scheduling, pharmacy, coding, billing, claims and coverage, because a patient's care does not stop at the exam-room door.
- Agentic: it acts on its own initiative to gather, check, prepare and follow up. It pulls a record, runs an eligibility check, drafts a note, notices a missing answer, and asks the question nobody asked.
- Resident: like a resident physician it works beside the people who are accountable, prepares the work, and learns from every case. Unlike a resident, it never signs.
- Assistant: its output is insight and a recommendation with the evidence behind it. The decision, and the signature, belong to a person.

The problem: care arrives in pieces

A single ankle injury touches a triage nurse, a scheduler, an eligibility check, a front desk, an intake nurse, a physician, a radiologist, a specialist, a pharmacist, a coder, a biller, the insurer's membership, utilization-review and claims staff, and the patient. Each works in a different system. Each re-asks, re-keys or re-checks something the last one already knew, and each sees only the slice of the patient their own system holds.

The consequences are familiar. Medication lists drift between the pharmacy and the chart. An X-ray is repeated because the first one sits in another hospital's archive. A claim is denied because a prior authorization was never requested, or because an insurance plan that ended in December is still listed first. And almost nothing learns from what happened afterwards: whether the ankle healed, whether the patient came back, whether the cheaper plan of care worked as well as the expensive one.

What MARA is, and what it is not

MARA is an assistant that sits beside every one of those people. It brings the patient's record, coverage, images, results, prescriptions, claims and history into one view; it evaluates each step medically, procedurally and against federal, state and insurer rules; and it recommends. Every recommendation arrives with its evidence, its source and its confidence, and with three choices for the person receiving it: accept it, change it with a reason, or escalate it.

MARA does not diagnose, prescribe, sign, deny a claim or change a price. That is not a policy bolted on at the end. It follows the way MAX already separates what it may do from what it may only recommend: MAX classifies every capability it has, and high-stakes actions sit in a recommend-only class where a person decides and an attempt to execute raises an error rather than acting. Its root-cause reasoning is built the same way; the diagnoser reasons and never dispatches.

What it hopes to achieve

MARA is judged by what changes for patients, for the people who care for them and for the people who pay. Each goal below comes with the evidence that would show it worked, because a goal nobody can measure is only a wish.

- Lower cost of care. Fewer repeated tests and images, fewer avoidable emergency visits and readmissions, fewer denied claims, appeals and resubmissions, and less time spent re-keying what another system already holds. The evidence is the cost of a whole episode of care for a cohort, before and after, with outcomes held level or better.

- Better outcomes that last. Recommendations graded by recovery, return visits and complications rather than by approval, so the advice that works for patients like this one rises and the advice that does not falls. The evidence is recovery and relapse by cohort, followed for years rather than days.
- More available care. Every minute MARA removes from paperwork, phone tag and re-checking is a minute a nurse, doctor, pharmacist or coder can give back to a patient. Wizard-mode questions let newer staff work safely at the top of their training, remote check-ins reach people between visits, and a sourced, complete chart makes a first visit with a new provider as good as a tenth. The evidence is time to first appointment, time returned to clinicians, and the number of patients a team can safely see.
- Safer care. Interactions, allergies, duplicate therapy, missed red flags and stale coverage caught at the moment they enter the record. The evidence is the number caught before harm, and the number that still got through.
- Fairness that is checked, not assumed. Every change to what MARA recommends is tested for bias across age, sex, race, language and payer before it ships. The evidence is published per release.

Built on MAX

MARA is not a separate product with its own brain. It is MAX, pointed at healthcare, and it inherits the properties that make MAX what it is.

- The graph is the memory. MAX keeps what it knows as nodes and edges in a graph store rather than as loose documents, with who, what, when, where, why and how recorded as relationships. MARA's patients, conditions, treatments, medications, providers, payers, claims, outcomes and cohorts are nodes; the relationships between them are edges it can traverse and reason over.
- Confidence is first-class. Every fact in MAX carries a confidence score and the source it came from, and what MAX says is shaped by that: it asserts, hedges, or surfaces a disagreement. In MARA a fact without a source is not shown as a fact, and a question nobody answered is shown as unknown, never as no.
- Curiosity is built in. MAX detects its own gaps, such as a low-confidence claim or an unknown entity, stores them, and turns the most important ones into

questions to answer later. In MARA that same behaviour becomes the question nobody asked on a phone call, the outside record nobody retrieved, or the follow-up nobody scheduled.

- It questions before it asserts. MAX's imagination engine proposes parallels and conjectures as hypotheses, not facts, until a source confirms or refutes them. MARA's cohort patterns follow the same discipline.
- It learns from outcomes. MAX's diagnosis memory raises its confidence in a cause when a fix is verified to hold and lowers it when the problem comes back. MARA applies that loop to care: a recommendation that was accepted but turned out wrong loses weight; a correction that proved right gains it.

Some of this exists in MAX today and some is design. The graph store, confidence scoring, gap detection, self-generated questions and outcome-weighted diagnosis memory are part of MAX. The healthcare connectors, the clinical and payer knowledge, and graph-based retrieval across a population of patients are the work that further development has to do.

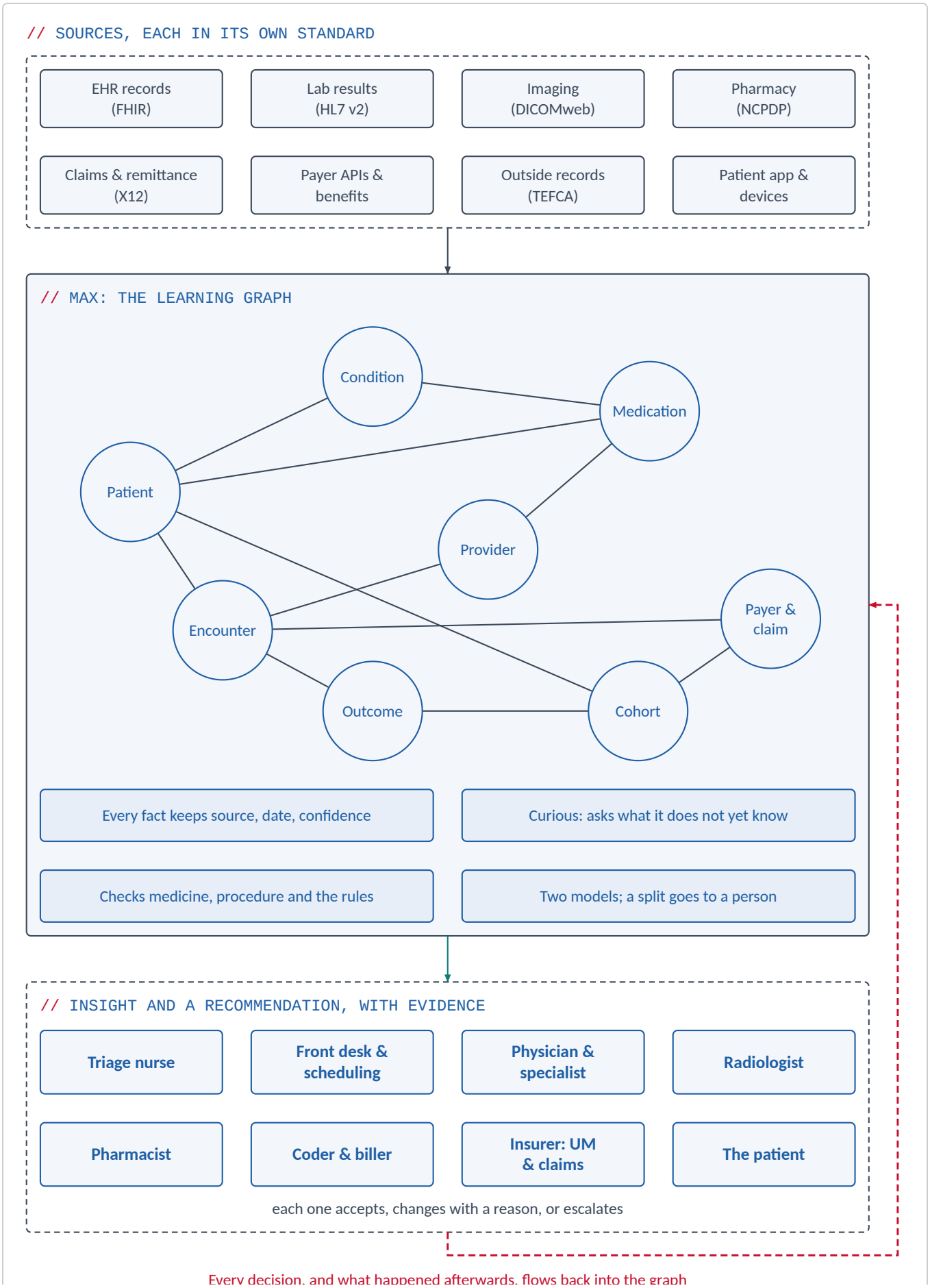

// SOURCES, EACH IN ITS OWN STANDARD
EHR records (FHIR)
Lab results (HL7 v2)
Imaging (DICOMweb)
Pharmacy (NCPDP)
Claims & remittance (X12)
Payer APIs & benefits
Outside records (TEFCA)
Patient app & devices
// MAX: THE LEARNING GRAPH
Condition
Medication
Patient
Provider
Encounter
Payer & claim
Outcome
Cohort
Every fact keeps source, date, confidence
Curious: asks what it does not yet know
Checks medicine, procedure and the rules
Two models; a split goes to a person
// INSIGHT AND A RECOMMENDATION, WITH EVIDENCE
Triage nurse
Front desk & scheduling
Physician & specialist
Radiologist
Pharmacist
Coder & biller
Insurer: UM & claims
The patient
each one accepts, changes with a reason, or escalates
Every decision, and what happened afterwards, flows back into the graph

Figure 1. MARA: every source in, insight out, a person decides.

The patient graph

At the centre of MARA is a graph of the patient. Around them sit their conditions (DX), treatments (TX), medications (RX), the providers who saw them, their payers and plans, every claim and its outcome, their devices and app entries, the people related to them, and the cohorts they belong to. Every node and every edge carries its source, time and confidence.

Because the graph links clinical facts to financial ones and to outcomes, MARA can answer questions no single system can: which treatments for this condition led to recovery without a second visit, for patients like this one, at what cost; which medication combinations preceded an emergency visit; which coverage patterns produced avoidable denials. The same graph, read across many patients and de-identified, is what the learning described below works on.

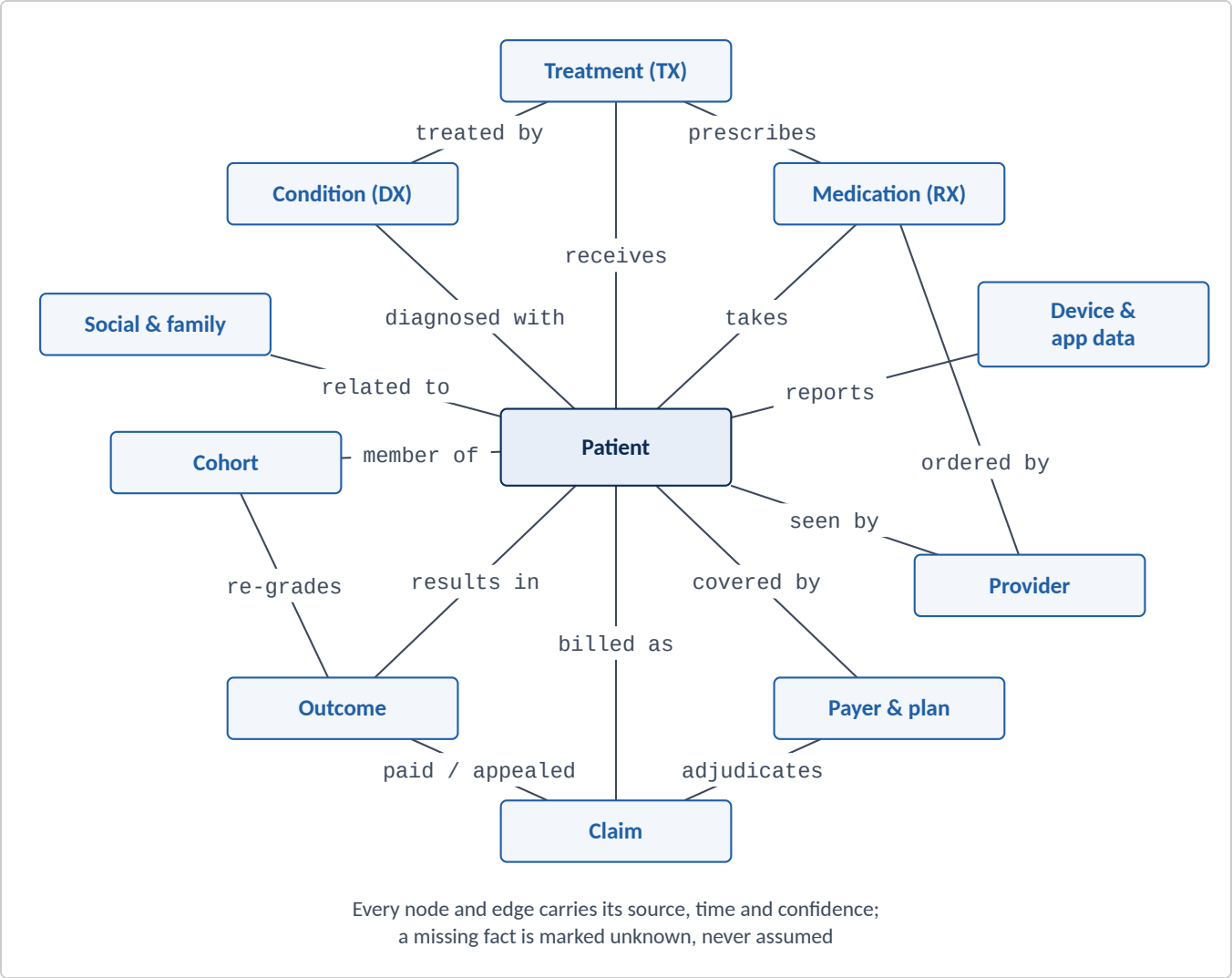


Figure 4. The patient graph: the nodes and edges MARA reasons over.

One encounter, every role

The demo follows Dana Whitfield, 34, who fell from a curb and cannot put weight on her right ankle. Every person, plan and amount in it is fictional. Her episode runs through eleven steps across four lanes, and each step has the same shape: MARA prepares and recommends, the person in that lane decides, and the step is written to an audit ledger.

- Triage: the nurse works through the ankle-injury protocol; MARA hears the answers in the call, flags the red-flag question nobody asked, and works out the protocol's level of care.

- Scheduling, eligibility and check-in: a same-day slot, an electronic coverage check, and a coverage-order check that catches a former employer's plan still listed first.
- Intake and the visit: vitals against her own history, a reconciled medication list, screening questions, the exam and the X-ray order.
- Imaging and the specialist: an AI pre-read with the suspected fracture marked, the radiologist's signed report, and an orthopaedic plan.
- Pharmacy, coding and billing: an interaction check against everything she takes, codes tied to the exact wording of the note, and the claim.
- The insurer: prior authorization, the claim decision with the rule it rests on, the letters and notices, and her history across plans.
- The patient: her results once signed, her bill explained line by line, her privacy choices, corrections to her record, and follow-up questions at two and six weeks.

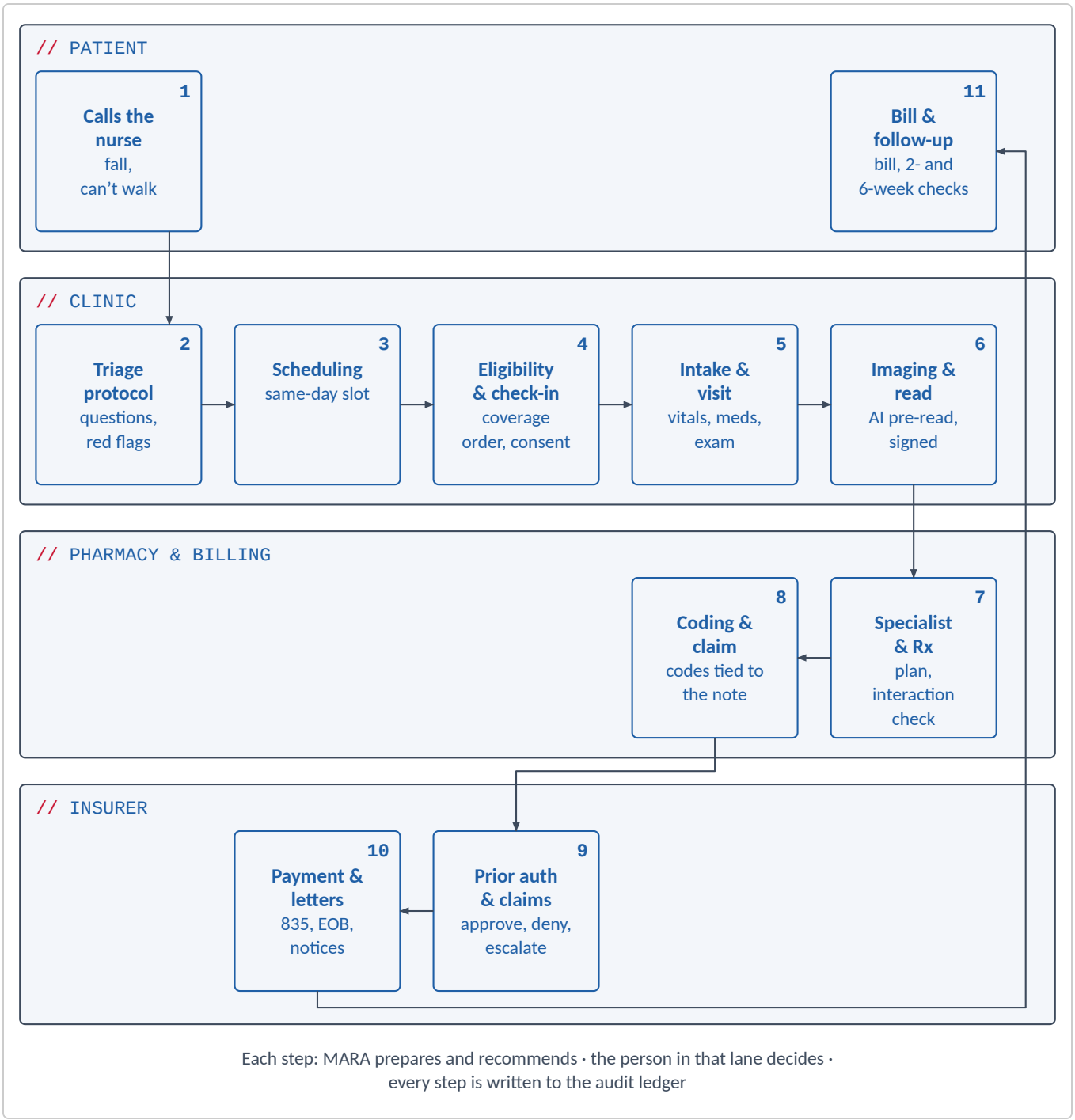


Figure 2. One encounter, every role: the workflow MARA supports.

Asking the right questions

Much of healthcare is asking questions: the triage protocol, the registration questions, the Medicare Secondary Payer questions, the intake screening, the pharmacist's counseling. MARA gives every role that asks questions the same tool, in two modes. Expert mode puts every question on one screen with one-click answers.

Wizard mode takes a newer staff member through one question at a time, with the words to say, why the question matters, and how people usually answer and what each answer means.

Answers can be yes, no, don't know, couldn't answer, declined, or the question's own choices, and every answer records who gave it: the patient, or someone else and their relation. A person can also type what was actually said. MARA reads the words, suggests the protocol answer with a confidence score, and, when the answer is vague, suggests the next question. "A little tingly earlier but fine now" does not become a clean no; it becomes a follow-up: is it there right now?

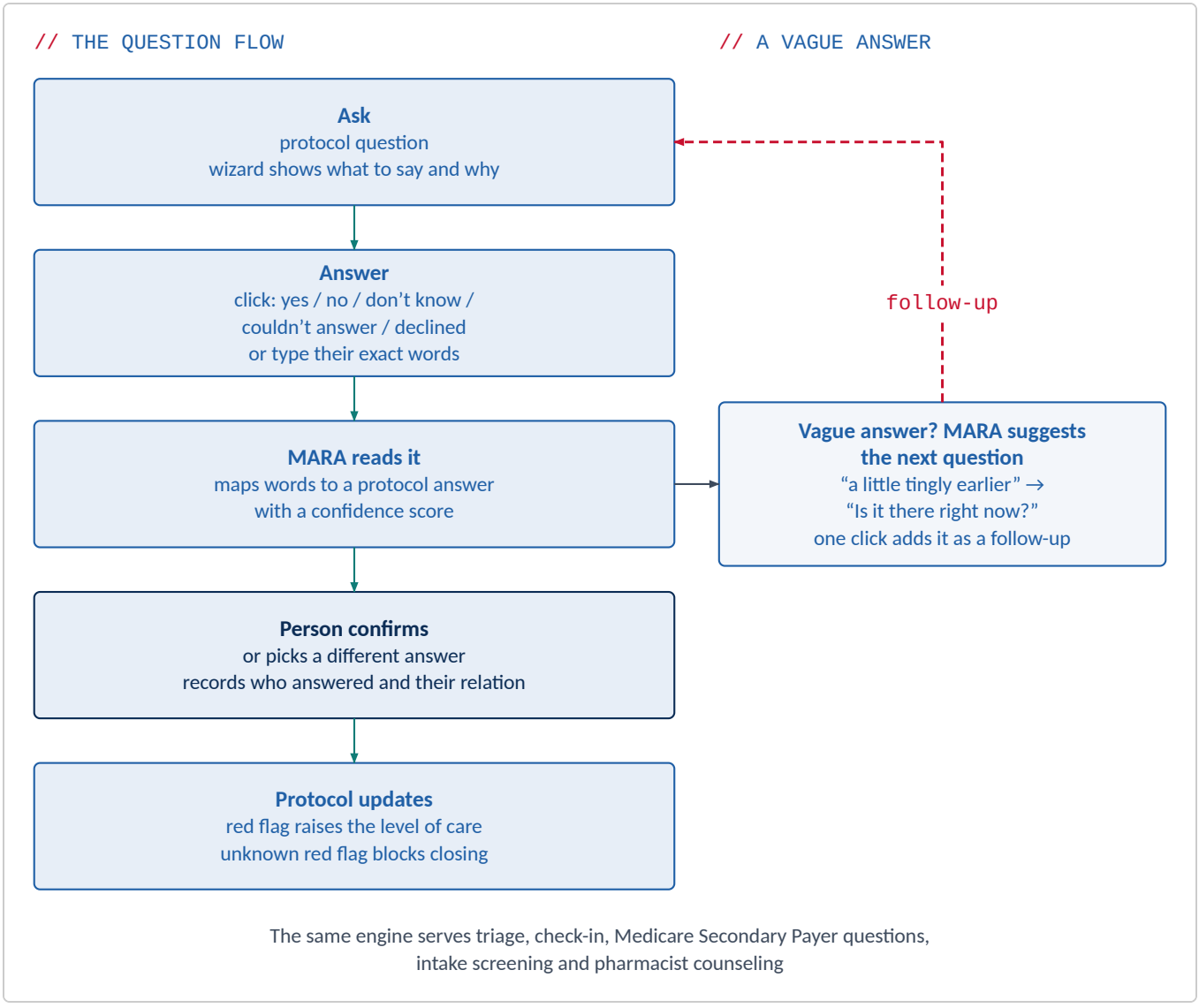


Figure 5. Asking questions: expert or wizard, answered by anyone, confirmed by a person.

Recommendations to every provider on the team

Because MARA sees the whole graph, its recommendations reach everyone caring for the patient, not only the person in front of the screen. The primary care physician sees the orthopaedic plan and the physiotherapy referral. The pharmacist sees the over-the-counter ibuprofen the patient mentioned to the triage nurse. The insurer sees the evidence attached to the prior authorization before anyone has to ask for it. The patient sees what to watch for and when to call.

The aim is four things at once, and the design treats them as one problem rather than four:

- Safer: interactions, allergies, duplicate therapy, missed red flags and stale coverage are caught where they enter, not after they cause harm.
- Less expensive: fewer repeated tests, fewer avoidable emergency visits, fewer denied claims and fewer appeals, and treatment paths priced by what they actually achieve.
- More comprehensive and holistic: the physical injury, the medicines already in the cabinet, the depression screen, the fall history, the home situation and the coverage are part of one picture.
- Better outcomes that last: recommendations continue after the visit, and each outcome teaches the next recommendation.

Learning during and after every encounter

MARA learns from every recommendation, but approval is its weakest signal. A person accepting a suggestion in the moment says little about whether the suggestion was right. What says more arrives later: whether the claim was paid or overturned on appeal, whether the ankle healed by six weeks, whether the patient returned with the same injury a year later, and what the whole episode cost.

So each recommendation becomes a record with three parts: what MARA suggested, what the person did and why, and what eventually happened. The last part re-grades the first two. A reasoned change counts for more than a quiet approval, and

an outcome counts for more than either. Patterns found across similar patients become cohort knowledge, and a change to what MARA recommends ships only as a versioned release that people review, test for bias across age, sex, race, language and payer, approve, and can roll back.

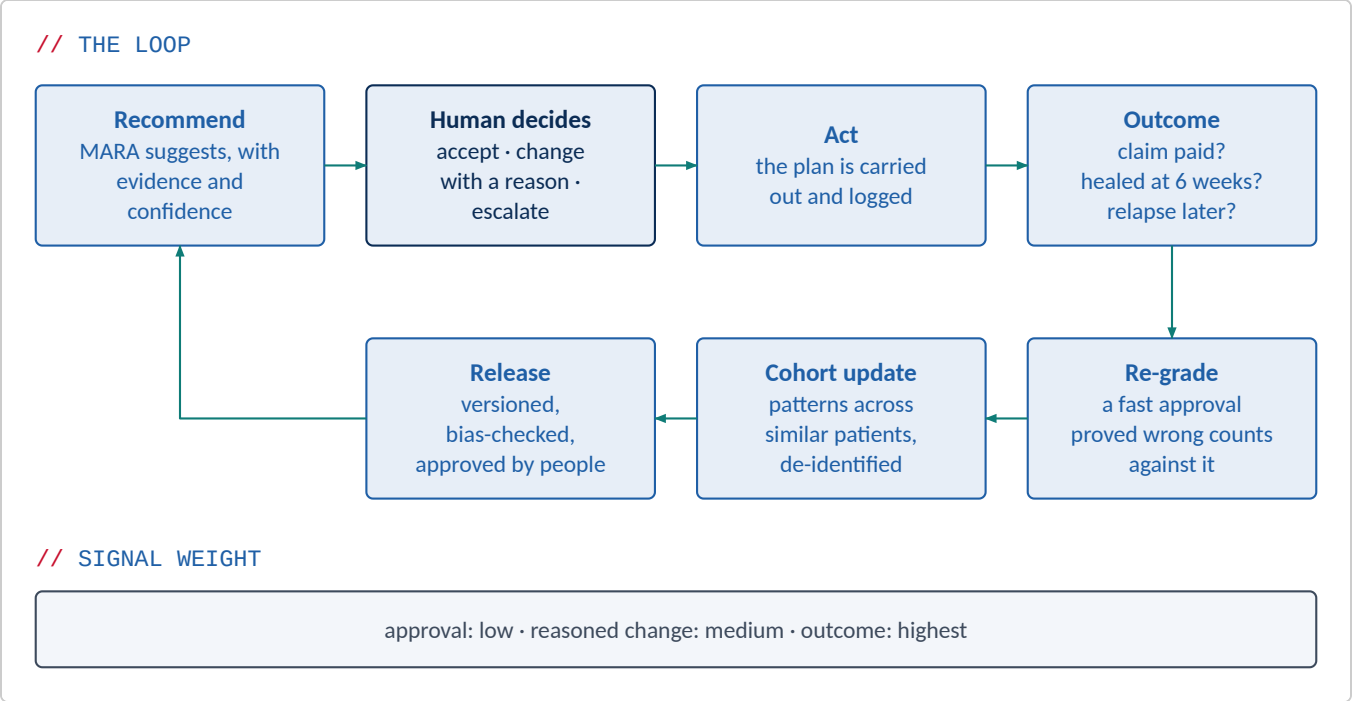


Figure 3. The learning loop: outcomes outweigh approvals.

The learning spans every kind of data the graph holds: patient history and self-reported outcomes, provider decisions and their reasons, prescriptions (RX), diagnoses (DX), treatments (TX), payer rules and decisions, claims and remittances, and cohorts followed across their whole lives, from entry into care through history, prognosis, treatment and recovery or relapse. It uses de-identified data only where consent allows, and requires a minimum cohort size before any pattern can influence a recommendation.

Care that does not stop when the visit does

Most healthcare software forgets a patient the moment the claim is paid. MARA is designed to keep going. Before the visit it knows the care gaps and screenings that are due. After it, it asks the patient how they are doing at two weeks and six weeks, watches for the signs that call for a change of plan, and at a year asks whether the

injury returned. Each answer goes back to the providers who need it and into the learning loop.

Cradle to grave: one record for a lifetime

A person's health is one story told across seventy or eighty years, dozens of clinicians, several insurers and a handful of moves. Today it is kept in fragments, and each fragment forgets the others. MARA is designed to keep the whole story. Birth, childhood vaccinations and growth, the sports injury at sixteen, the pregnancy, the first blood-pressure reading that was a little high, the knee replaced at sixty-eight and the medications that followed, and the choices a person made about the end of their life: each is a node in the same graph, with its source, date and confidence, and with consent deciding who may see it.

A lifetime record changes what a recommendation can know. The ankle that is sprained at thirty-four is not the first injury if there was a fracture at twelve; the blood pressure that is borderline today is not a single reading if it has climbed for ten years; the medicine that is safe on its own may not be safe beside the one another clinic prescribed last spring. Continuity is not a feature added to care. For most of what goes wrong, it is the care.

From one patient to a whole population, and back

The most hopeful idea in MARA is that every person's care can make everyone else's better. Each recommendation, the decision a person made, and what happened afterwards are recorded together. Read across a whole population of people, de-identified and with consent, those records answer the questions no single visit can: which treatments led to lasting recovery for people like this, which led to relapse five years later, which cost more without doing more, and which groups were served worse than others.

Those answers are measured by cohort and over a lifetime, not by visit. A cohort is a group of people similar enough in condition, history and treatment that their outcomes can inform each other's care: adults with a first non-displaced ankle

fracture, children with asthma in the same climate, older adults on five or more medicines. MARA follows each cohort from entry into care through diagnosis, treatment and recovery or relapse, across years and across insurers, and a pattern only influences a recommendation once the cohort is large enough to trust and the finding has survived review.

Then the knowledge comes back to one person. When the next patient calls with a hurt ankle, the recommendation their nurse sees is informed by everyone like them who came before, weighted by what actually worked for those people over the years that followed. That is the loop the whole design exists to close: population evidence improving individual care, and individual outcomes improving the population evidence.

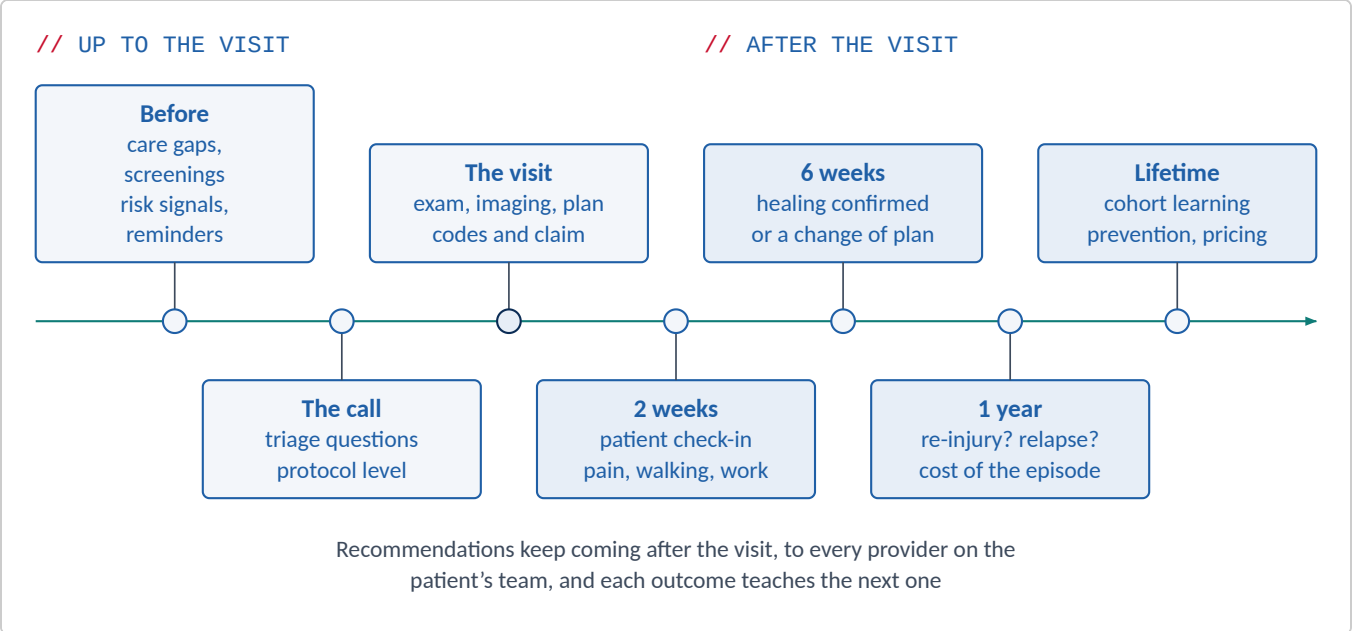


Figure 6. Care that does not stop when the visit does.

Value, cohort pricing and outliers

Because outcomes and payments sit in the same graph, MARA can suggest what treating a group of similar patients should cost, for example adults with a non-displaced ankle fracture managed in primary care with one orthopaedic follow-up. It recommends a bundled price only when outcomes hold or improve, and it flags outliers at both ends: expensive cases that suggest repeated imaging, and cheap

cases with poor outcomes that suggest under-treatment. Contracting staff accept, adjust or reject the price with a reason, a physician reviews the care pathway, and a price changes only through a contract, never automatically and never for an individual patient.

Compliance you can audit

Every event goes to an append-only ledger recording what happened, the rule that applied, whether that rule was federal, state, insurer contract or clinic policy, whether AI was involved, and who decided. The rules modeled include HIPAA's minimum-necessary, access, amendment and accounting-of-disclosures provisions, 42 CFR Part 2, the Cures Act information-blocking rules, Section 1557, the ACA, ERISA, Medicaid and Medicare appeal rules, the No Surprises Act, and coordination of benefits: the NAIC model rules for commercial plans, the Medicare Secondary Payer rules including the secondary payment calculation in 42 CFR 411.33, and Medicaid as the payer of last resort. These describe the design's intent and would be confirmed by regulatory counsel before any real use.

Data across the seams

MARA connects to the systems a clinic and an insurer already run through the standards they already speak: FHIR R4, SMART on FHIR and CDS Hooks for the health record; HL7 v2 with LOINC for laboratory results; DICOMweb for images; TEFCA for records from other organizations, with consent; X12 for eligibility, prior authorization, claims and remittance; the payer APIs required by CMS-0057-F; NCPDP SCRIPT for pharmacy; SFTP and AS2 for bulk files; Direct messaging between providers; and the patient app and home devices as sources the clinician verifies.

What further development will add

This is a large undertaking, and it is meant to be. It reaches into the health record, the laboratory, imaging, pharmacy, the revenue cycle and the insurer; it has to speak more than a dozen data standards; it serves every role from the triage nurse to the claims examiner and the patient; and it has to earn trust under federal, state and

contract rules at every step. The demo shows the experience. Turning it into a system people could rely on is a sequence of steps, each of which has to earn the next.

- A read-only data spine: health record, laboratory and eligibility feeds with full provenance, so the first value is simply a complete, sourced chart and verified coverage.
- Assistive drafts with accept-or-change on every one: triage summaries, notes, codes and imaging pre-reads, with every change and its reason captured.
- The revenue-cycle and payer seam: claims, remittances, prior authorization through the payer APIs, denial prevention and appeals support, letters and notices.
- Patient write-back and follow-up: remote check-in, symptom check-ins, corrections, consent-gated outside records and device data.
- Outcome-weighted learning at population scale: graph-based retrieval across de-identified cohorts, re-grading from outcomes, bias monitoring and a formal model release process.
- Value and pricing: episode bundles, outlier detection and contracting workflows with clinician review.

Each step pays for itself before the next begins. A complete, sourced chart and verified coverage are worth having on day one, before any learning exists. Drafts that people accept or correct save time immediately and produce the labelled examples the learning needs. The revenue-cycle seam reduces denials and appeals while it connects outcomes to payments. Only then does population learning begin, on data that has been sourced, consented and checked all the way through.

Beyond those steps sit the wider ambitions the graph makes possible: remote monitoring that tells the care team when a recovery drifts off course, social and home factors that change what a good plan is, medication and genetic interactions checked against the whole list, matching patients to clinical trials they qualify for, population health and public health reporting, and care coordination across organizations that today barely talk to each other. None of it changes the rule. Each

is another source of insight and another recommendation, and a person still decides.

What MARA will never do

- Make a clinical, coverage or payment decision, or sign anything.
- Show a guess as a fact, or an unanswered question as a no.
- Learn from a patient's data without consent, or let a pattern from too few patients drive a recommendation.
- Change a price, a plan or a model silently; every change is versioned and approved by people.
- Show anyone more of a record than their role needs.

How to walk the demo

Open the demo and follow the guided tour, which explains every screen: what the user sees, why it is valuable, what MARA does and what the person decides. The Demo scenario control changes the insurer, the X-ray result and other conditions. Try typing a vague answer to a triage question, reporting a spouse's plan at check-in, rewriting the radiology read, or denying the claim and appealing it, and watch every role's screen follow.

What it could mean

Picture the same ankle injury ten years from now. The nurse on the phone already knows the fracture at twelve and the blood thinner started last year, and the red-flag question is asked because MARA noticed it was missing. The front desk never enters a plan that ended in December. The X-ray from the urgent care across town is on the screen before anyone orders a second one. The claim is paid the first time, because the evidence was attached before anyone asked. Six weeks later the patient answers two questions on their phone, and their answers, with thousands like them, tell the next nurse which advice actually helped.

None of that asks a machine to decide anything. It asks for the information that already exists to arrive where it is needed, for every recommendation to be graded by what happened to real people, and for what is learned from a whole population to come back to each person in it. Care that costs less, works better, reaches more people and follows them for life is not a new ambition in healthcare. What is new is that the pieces to build it now exist, and MARA is a design for putting them together while keeping people in charge of every decision.

Explore the design and the goals on the project page: https://www.maxresearchcollective.com/mara.html. Walk the episode yourself in the interactive demo: https://www.maxresearchcollective.com/demos/mara-demo.html.

Glossary

TERM	DEFINITION
Abductive diagnosis	Reasoning from a symptom back to the most likely cause by asking questions whose answers can be checked.
CDS Hooks	A standard for a health record to ask an outside service for a suggestion at a specific point in a clinician's workflow.
Cohort	A group of patients similar enough in condition, history and treatment that their outcomes can inform each other's care.
Coordination of benefits	The rules that decide which of a person's health plans pays first, second and third.
DX, RX, TX	Shorthand for diagnosis, prescription and treatment.
FHIR	Fast Healthcare Interoperability Resources, the modern standard for exchanging health records over web APIs.
Knowledge graph	Information stored as nodes (things) and edges (relationships), so connections can be followed and reasoned over.
Learning health system	A health system in which the outcomes of care are captured as they happen and fed back to improve the care of the next patient.
Lifetime record	One longitudinal record of a person's health from birth onward, across every provider and insurer, with the source and consent of every entry.
Provenance	Where a fact came from, when, and how confident the system is in it.
TEFCA	The Trusted Exchange Framework and Common Agreement, the national network for exchanging records between organizations.
X12	The family of electronic transactions used for eligibility, prior authorization, claims and payment in US healthcare.

HEALTHCARE WPR-HLC-001

Newell, J. (2026). MAX Research Collective, WPR-HLC-001. CC BY 4.0

// ABOUT THE AUTHOR

Jason Newell leads organizations and people in technology, and is a researcher, deep technologist and builder — from a CRAY to agentic AI. He founded MAX Research Collective to build AI for regulated industries that runs where its owner has control, with healthcare as its strongest focus. About Jason → Connect on LinkedIn →