AI Use Case Assessment

Analyst: Lauren Chen, AI Governance Lead
Organization: Crosswind Workforce Partnership
System: ReachFirst

Instructions

A use-case assessment is the structured evaluation of whether and how an AI use case should proceed. It is conducted before any data is touched or model trained. Each of the eight sections below opens with an objective that explains what the section is for and why it belongs in the assessment, followed by the guiding questions to answer. Work top to bottom.

Section 1: Business Problem & Context

Objective: Establish the real-world problem the use case is meant to solve. State the problem in business terms and avoid referencing AI. The goal is to produce a clear-eyed account of what is going wrong today, what the status quo costs, and why acting now is justified. This way, the case rests on a genuine need rather than technology enthusiasm. If the problem cannot be articulated without mentioning the model, that is a signal to pause.

What is going wrong today?

The families who most qualify for Crosswind's programs are the least likely to hear about them. Uptake among eligible households runs below 40 percent for the flagship programs, and the current finding methods, word-of-mouth, walk-ins, and untargeted mailing, favor residents who are already connected to us. The outreach budget is capped by contract, so the finding problem cannot be solved by spending more. It can only be solved by finding better.

What does the status quo cost?

Eligible families going unserved while budgeted outreach reaches households that would never qualify. Under-enrollment that shows up in our performance contract, and a quiet unfairness in the status quo itself: help flows toward the connected, and nobody has ever had to write that down or answer for it.

Why act now?

A funder has underwritten the pilot, the vendor's prototype exists, and the board has conditioned everything on this assessment, which means the governance questions get answered before habits form. Acting now also puts the current system's biases on the record for the first time, which is worth doing even if the model never launches.

SECTION 2: DECISION(S) THE USE CASE SUPPORTS

Objective: Pin down the specific human decision or process the AI system is meant to inform. List who owns the decision and/or process. This section fixes the system's scope and posture: determining whether the system offers decision-support with a human in the loop or automates the decision outright. This section also sets the purpose limitation by drawing the red lines (i.e., the uses the output must never be put to). Everything downstream to the decision or process should inherit these boundaries.

What does the decision change?

The order in which households are contacted with a fixed outreach budget: who we try to find first. Nobody's eligibility is decided by this tool, and no application is touched by it. But because budget runs out before lists do, contact order determines in practice who hears from us at all, which is why this decision deserves the full assessment the board required.

Who is the human decision-maker?

Today, the outreach director and his coaches decide, informed by geography, events, and experience. Under this assessment's recommendation, they continue to decide: the list proposes an order, and the team works it with judgment. No automated contact, no automated exclusion.

Is the AI system decision support, or will it make decisions automatically?

Recommendation: priority ordering only, with every household remaining on the list, published prioritization criteria, and contact-rate reporting by community and group. Not a hard cutoff, which would convert a prediction into a silent denial of outreach. Not neighborhood batching in this phase, because geography carries protected-characteristic history that our reporting is not yet ready to monitor. The asymmetry must be governed explicitly: when budget exhausts before the list does, an ordering behaves like a cutoff for those at the bottom, so the bottom of the list is measured from day one.

What is the AI system NOT to be used for?

ReachFirst must never determine or influence eligibility, and no application may be scored. Scores must never be shared outside the program, never disclosed to employers, landlords, or other agencies, and never shown to residents as a judgment about them. No resident who contacts us is ever deprioritized for service because of a score. The race, sex, and national-origin fields in the extract must not be used as model inputs. Any new use requires a new assessment.

SECTION 3: AI / ML FRAMING

Objective: Translate the business problem into a precise technical problem type. Name the unit of analysis, the target being predicted, and the baseline the model must beat. Then, test honestly whether a simpler non-AI method would do the job. The objective is to keep the team disciplined about whether machine learning is actually warranted, and to make the model's added value (and added risk) an honest comparison rather than an assumption.

What is the technical problem being addressed?

Binary classification at the record level: estimate whether an adult's recorded income falls below the line the vendor uses to approximate eligibility for Crosswind's largest programs, and use the estimates to rank the outreach universe by likely eligibility.

What is the unit, target, or baseline?

The unit of analysis is an adult record, and the dictionary says the label is that person's own income, while our eligibility rules are written for households; this assessment records that gap as an assumption to confirm. The target is income at or below the eligibility-proxy line, and about 24 percent of the extract falls above it. The baseline to beat is the status quo: untargeted mailing plus word-of-mouth and walk-ins. That baseline is unfair in its own quiet way, which makes it beatable, and also makes the comparison mandatory: the model must demonstrably find more eligible families than current practice, across every community we serve.

Could a simpler, non-AI/ML solution solve this problem?

Partially. Saturation outreach in high-need areas and deeper partnerships with community organizations would both improve finding without any model, and both should proceed regardless. What they cannot do is rank the whole outreach universe consistently. A model is a reasonable candidate for that job if its data can be shown to represent today's

residents and its errors can be shown not to concentrate in any community. Those are testable claims, and Section 5 sets the tests.

SECTION 4: AFFECTED PARTIES

Objective: Identify everyone the use case touches, not just its users. Identify and list the AI subjects (those being scored or acted on), the decision-makers, and the parties affected downstream of the system's deployment and decision-making. Pay particular attention to those affected who have no say. Surfacing protected groups here is what makes fairness testing possible later. This section is the foundation for the success criteria in §5 and the harm inventory in §7.

Who are the AI subjects?

Residents of the six-county service area whose records are processed and ranked, roughly the adult population of the outreach universe. They are affected whether or not they are ever contacted.

Who are the decision-makers?

The outreach director and his coaches, who decide whom to contact and work the list daily. The Chief AI Officer, as approver of record. The board, whose approval is contingent on this assessment.

Who is affected downstream?

Families contacted earlier and enrolled, the intended beneficiaries. The eligible family never contacted, who appears in no metric unless deliberately counted, and who bears the program's worst failure invisibly. Coaches whose credibility with residents depends on explainable methods. The state contract office reviewing equal-service reports. The funder, whose grant requires a public write-up. The vendor, whose documentation obligations this assessment names.

Which protected groups are present?

The extract contains race, sex, and country of origin as explicit fields, and the household relationship field encodes family structure, entangling sex and marital status. Race, color, and national origin are protected under Title VI; the workforce system's own provision adds sex, age, religion, disability, political affiliation, and citizenship status. Presence in the data is a fact. Exclusion from model inputs, and use for fairness measurement only, is this assessment's recommendation.

Who is affected but has no say?

The never-contacted eligible family, above all. They did not apply, so no file exists. They were not denied, so no appeal exists. They simply never heard from us. If the model under-predicts eligibility for their community, that silence becomes systematic, which is precisely the board's silent-redlining warning. Counting the uncontacted, by community and group, is how this program makes its worst failure visible.

SECTION 5: DEFINITION OF SUCCESS

Objective: State what success means concretely across three lenses: the business outcome, the model metrics that fit the data (not a default like accuracy), and the fairness criterion with its named groups and threshold. Defining failure just as concretely is part of the job. The goal is a target that is measurable and testable rather than aspirational. With this, later stages can verify whether "success" was met.

What is the specific business outcome the solution will address?

Raise enrollment among eligible households above the current sub-40-percent uptake, using the same capped outreach budget, with contact patterns that hold up across every community and group we serve, and with methods a coach can explain at the front desk.

What metrics will be measured to determine success?

Uptake among eligible households, against the historical baseline. Cost per successful enrollment. Contact rates and enrollment rates by community and by demographic group. The share of enrollments originating from model-prioritized contact versus other channels, the honest measure of whether the model finds people the status quo misses. Coach-reported explainability, gathered directly from the people who use the list.

What fairness criterion will be used to identify whether the solution performs equitably?

Per-community and per-group comparison of contact rates, enrollment rates, and the model's error rates, using the extract's demographic fields for measurement only. The working criterion is that prioritization must not skip over any community: contact and error rates should not differ materially across groups at equal estimated need. Measures and thresholds are finalized during evaluation and recorded before first use.

What thresholds will determine the AI system's success?

To be pre-registered before first use and graded at pilot close. Working targets: uptake measurably above the status quo baseline among eligible households. No material adverse gap in contact or error rates across communities or groups without documented justification and correction. The uncontacted population reviewed by community each cycle. Criteria published before first use.

What does failure look like?

A contact pattern that skips any community would be failure regardless of total enrollments, and it would deserve the board's name for it. So would a pilot that raises enrollment among the already-connected while the outreach gap persists, criteria we could not publish, or a list the coaches cannot explain. Failure is defined now, publicly, so nobody redefines it later.

SECTION 6: CONSTRAINTS

Objective: Catalogue the boundaries the use case must operate within before design begins. Naming the legal/regulatory, ethical/fairness, data/privacy, operational, and resource constraints up front prevents late-stage surprises and defines the guardrails the build must respect. Each constraint named here should have a visible home in the design and, where relevant, in the compliance record.

What legal and regulatory considerations inform the AI system's design?

Title VI of the Civil Rights Act (42 U.S.C. 2000d), which reaches how a federally funded program's benefits actually reach people. WIOA Section 188 and 29 CFR Part 38, the workforce system's own nondiscrimination regime, mirrored in our state contract's equal-service reporting. Colorado's Automated Decision-Making Technology Act (C.R.S. 6-1-1701 and following): this assessment takes the position that ReachFirst should be governed as if covered. The classification is genuinely arguable, outreach resembles excluded marketing in some respects, yet an adverse outcome that restricts access to a public-benefits opportunity is close to the statute's core concern, and the asymmetry analysis above shows how an ordering can restrict access in practice. Governing as if covered costs little and protects much; the position is documented for review. The Colorado Privacy Act's principles of purpose limitation and data minimization apply to the extract as design requirements.

What ethical and fairness principles inform the AI system's design?

The front desk test as an operating principle: methods explainable truthfully in one breath. Guardrails that can be published without embarrassment. Equity measured rather than assumed, with the status quo's own bias acknowledged rather than romanticized. The board's community representatives treated as reviewers, not an audience.

What data protection and privacy considerations inform the AI system's design?

The extract is personal data covering tens of thousands of residents. Purpose is limited to outreach prioritization for this pilot. Access is limited to the program team, retention runs no longer than the pilot requires, and the vendor's data handling is bound by contract. No resident data is shared outside the program.

What operational considerations inform the AI system's design?

The list must fit how outreach actually works: coaches, calls, mailers, and events, inside a capped budget. It must fail safe: if ReachFirst is unavailable, outreach continues under current practice. Coach training comes before first use, and coach feedback is collected as evaluation evidence, not anecdote.

What resource considerations inform the AI system's design?

No in-house data science: the vendor owns the code, so Crosswind's leverage is documentation, contract terms, and the right to walk away. A governance office of two plus this analyst. A capped budget that makes every prioritization decision a real allocation. The pilot clock, with public results due at pilot close.

SECTION 7: INITIAL RISK & HARM INVENTORY

Objective: Produce a first-pass catalogue of the positive and negative harms the use case could cause, including who bears each and the worst/best plausible case. This is not the full risk assessment that comes later. The objective is to see the hazards clearly enough in order to judge whether and how to proceed, and to hand the downstream risk assessment a starting list rather than a blank page.

What negative harms could the AI system cause?

Silent redlining of outreach: under-prediction concentrated in particular communities, producing systematically less contact, invisible to any complaint process. Proxy discrimination through occupation, household structure, or geography even with protected fields excluded. A stale extract misrepresenting today's residents, so the model finds yesterday's eligible families. Patterned missingness in employment fields quietly deprioritizing exactly the households hardest to document. A vendor tool the deployer cannot fully inspect. Erosion of community trust if the method cannot be explained or defended.

What positive harms could the AI system cause?

More eligible families found and enrolled with the same budget. The current system's invisible biases replaced by criteria that are written down, published, and reviewable. A documented equity baseline for outreach that outlives the pilot. A governance file that makes Crosswind's next tool easier to do right.

Of the harms listed, which stakeholders are affected?

Residents bear the contact-pattern harms, with under-predicted communities bearing the most, invisibly. The never-contacted eligible family bears the worst case. Coaches bear the explainability burden at the counter. The organization bears the contract, funder, and trust consequences. The benefits accrue to found families, to coaches with a workable list, and to the communities whose access finally gets measured.

What is the worst plausible case that could emerge from the AI system's development?

The pilot proceeds on the vendor's assurances without documentation, the extract's vintage and gaps unexamined, and the first equal-service report with the model in production shows a contact skew nobody can explain because nobody looked.

What is the best plausible case that could emerge from the AI system's development?

Development becomes the occasion to write down what was never written: how outreach actually allocates attention today, what the data can and cannot support, and what the vendor's tool actually does. The pilot earns the board's trust by the quality of its file before it earns results.

What is the worst plausible case that could emerge from the AI system's deployment?

The model launches, contact quietly skews away from particular neighborhoods, the funder's public write-up surfaces the pattern, and the organization built to widen access becomes the case study in narrowing it. The board's four words, on the record, proven right.

What is the best plausible case that could emerge from the AI system's deployment?

More families found than the old way ever managed, in every community, with published criteria, measured patterns, and a front desk that can explain the whole thing in one breath. The quiet parts of the old system written down and fixed. Public money spent finding the people it was appropriated to find.

SECTION 8: GO / NO-GO DECISION & DOCUMENTED ASSUMPTIONS

Objective: Make and record the gate decision: proceed, proceed with guardrails, or stop. Document the guardrails that condition a "go" and, critically, the key assumptions that would change the answer if the assumptions turn out to be false. This creates an auditable decision trail and gives later stages a set of assumptions to monitor and revisit.

Should we build this?

Proceed, with guardrails. The need is real, the status quo has its own unmeasured biases, and the risks identified here are manageable if the sequence is respected: vendor documentation first, honest data work, published criteria, and measurement built to see the invisible failure. This is a go for the next lifecycle stage, documented data work and review. It is not a deployment decision, and nothing touches real outreach until the conditions in this file are met.

With what guardrails should the AI system be developed?

Obtain full vendor documentation of the tool: training data, intended use, limitations, and known failure modes, before any further development. Document the extract honestly: its vintage and what has changed since, the weighting field's proper handling, the patterned missingness in employment fields, and how the eligibility-proxy line maps to real program

rules. Exclude protected-characteristic fields from model inputs and evaluate proxies. Independent review of vendor performance claims.

With what guardrails should the AI system be deployed?

Priority ordering only, per Section 2, with coaches deciding their week and everyone remaining on the list. Published prioritization criteria before first use. Contact and enrollment reporting by community and group, to the board and in state equal-service reporting. The uncontacted counted from day one. Coach training before first use, and a rollback that returns outreach to current practice.

What assumptions, if false, would change the organization's decision to develop and deploy this AI system?

That the extract can be shown to adequately represent today's service-area population. That the income line, drawn on a person's income rather than a household's, tracks actual program eligibility closely enough to be a fair proxy. That the vendor will deliver real documentation and evidence. That coach capacity matches the contact plan. That the state-law position holds under review. If any of these fails, this assessment returns to the gate.