

A Workflow for Grant Discovery and Proposal Development Using Large Language Models: Development and Formative Evaluation

Katherine Monsalve Barrientos

`katherine.monsalve@arkangel.ai`

Arkangel AI

Natalia Castano-Villegas

Arkangel AI

Jose Zea

Arkangel AI

Laura Velásquez

Arkangel AI

Article

Keywords: artificial intelligence, natural language processing, research support as topic, research, medical informatics, low- and middle-income countries, workflow

Posted Date: July 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10241865/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. Katherine Monsalve, Natalia Castaño-Villegas, Jose Zea, and Laura Velásquez are employees or co-founders of Arkangel AI, the organization whose internal workflow is described in this manuscript. The paper describes internal organizational methodology and does not evaluate competing products or services.

Abstract

Background

Health-AI organizations in low- and middle-income countries need non-dilutive funding to generate evidence and validate tools in real-world health systems. Grant discovery and proposal development require repeated funder searches, eligibility interpretation, and narrative adaptation. Large language models can support parts of this work, but published guidance emphasizes caution, fact-checking, and human ownership of grant strategy. We did not identify a published description of a governed, repository-based grant development workflow inside a health-AI company.

Objective

To describe a governed, large language model-supported grant development workflow at a Colombian health-AI company, designed to improve throughput and traceability of grant discovery, triage, and proposal drafting while maintaining researcher oversight.

Methods

We describe the workflow architecture using internal repository artifacts, issue records, scouting logs, and scoring documentation. The workflow includes four phases: opportunity discovery, eligibility triage, Win Readiness scoring, and proposal drafting. The system uses repository state, issue labels, deduplication rules, and researcher review checkpoints to convert broad funding searches into prioritized work items.

Results

The workflow covered 20 funding scenarios across foundations, government agencies, pharmaceutical programs, innovation prizes, and bilateral mechanisms. Over 50 search combinations and 200 opportunities were reviewed. A structured scouting log covering 10 runs (2026-05-29 to 2026-06-05) recorded 741 scanned sources, 83 post-filter candidates, 46 net-new issues, and 220 duplicates. The Win Readiness Score was refined as a relative ranking tool across 49 grants. Five applications advanced to interview or finalist stage during the four-month operating period (March to June 2026).

Conclusions

A governed large language model-supported workflow can make grant discovery and proposal preparation more systematic and auditable for health-AI teams in resource-constrained settings. The

evidence supports a formative evaluation, not a causal claim about grant success. Future evaluation should measure proposal quality, researcher effort, and funding secured.

Introduction

Digital health and clinical artificial intelligence initiatives increasingly depend on rigorous evidence, local validation, privacy-preserving deployment, and partnerships with health systems. This is especially true in low- and middle-income countries, where digital health tools may address important access and quality gaps but often face barriers in financing, procurement, regulation, interoperability, and sustainable scaling [1–3]. Recent global health funders have also moved toward evaluation of artificial intelligence tools in real-world settings, including a partnership launched by the Gates Foundation, Wellcome, and Novo Nordisk Foundation to support locally led evaluations in low- and middle-income countries [4].

For early and growth-stage health-AI companies, this creates a practical challenge. Evidence generation requires funding, but funding itself requires a specialized workflow. Teams must identify relevant opportunities, determine eligibility across jurisdictions and entity types, assess fit against changing funder priorities, find academic or health-system partners when direct eligibility is blocked, write funder-specific narratives, and avoid spending scarce researcher time on low-probability applications. This is not a minor administrative problem. In health-AI, the quality of the funding pipeline can determine whether validation studies, implementation partnerships, and regulatory evidence packages happen at all [5].

Large language models are increasingly used to support research development and grant writing, but the scale of this transition brings accountability concerns. A quantitative analysis confirmed a measurable and sustained increase in large language model-modified content across scientific fields [6]. Major publishers have established governance policies addressing accountability, transparency, and content fabrication concerns as a direct response to this trend [7]. In grant writing specifically, Seckel, Stephens, and Rodriguez published ten rules for using large language models in grant development, emphasizing funder policy checks, privacy, fact-checking, plagiarism risk, and the principle that AI should not replace the scientific reasoning, data, voice, or responsibility of the applicant [8]. The relevant implementation question is therefore not whether a model can draft fluent text, but whether a governed workflow can use large language models and structured tools to increase throughput while preserving researcher judgment, confidentiality, traceability, and conservative decision-making.

Governance frameworks for AI-assisted knowledge work in health are beginning to appear in the literature [9–11], but we did not identify a published description of a governed, repository-based grant development workflow inside a health-AI company where funder triage, opportunity scoring, and proposal drafting are coordinated through a versioned issue-based state machine. The present paper describes the grant development workflow implemented at Arkangel AI, a Colombian health-AI company. The workflow was designed for a specific operational problem: a small evidence team needed to scan broad funding landscapes, prioritize opportunities across multiple funder categories, draft applications,

and maintain institutional memory without creating generic proposals or losing track of eligibility constraints. We present this as a methodology and formative evaluation paper, not a grant-success evaluation. The aim is to describe a reproducible operating model that health-AI teams can adapt to their own funding environments, particularly in low- and middle-income countries where dedicated research development infrastructure is absent. The workflow is not inherently limited to low-resource contexts; the same governance, triage, scoring, and artifact-preservation logic applies to any organization managing a fragmented funding pipeline.

Methods

Study design

This is a descriptive methodology and workflow paper. It reports the design, implementation, and early operational metrics of a governed grant development pipeline. No patient-level data, protected health information, funder-confidential peer review information, or private applicant records were used for this manuscript. The analysis draws on internal operational artifacts from the organization's grants repository and associated issue records, plus public literature on large language model-assisted grant writing and digital health sustainability.

The manuscript should be interpreted as a process description and preliminary operational evaluation. It does not evaluate causal effects on funding success, reviewer scores, or downstream health outcomes. Metrics are reported using predefined cutoffs: structured scouting metrics cover ten cycles from 2026-05-29 through 2026-06-05; application-history and advancement metrics are reported through 2026-06-09; and repository state labels were exported on 2026-06-10. Subsequent repository activity was not included in this analysis.

Setting

Arkangel AI is a health-AI company operating from Colombia with products and evidence workstreams related to automated clinical record review, clinical coding and billing audit [12], clinical search [13], hospital intelligence, and pharmaceutical evidence generation. The grants pipeline is operated as part of the organization's evidence and research-development function, staffed by a cross-functional team of four members combining research development, product, and leadership responsibilities, without dedicated grant operations personnel. The pipeline supports opportunities across multiple funding classes, including global health foundations, government research agencies, pharmaceutical and medtech innovation programs, payer-related opportunities, compute credits, innovation prizes, bilateral and multilateral programs, antimicrobial resistance programs, US federal opportunities, US state opportunities, Hispanic health programs, and partner-mediated academic routes.

System architecture

The workflow uses commercially available large language models through scheduled agent-based jobs for discovery search expansion, opportunity summarization, eligibility pre-screening, and first-pass section drafting. Interactive review, eligibility interpretation, partner strategy, and submission readiness decisions were performed by the research team in a human-in-the-loop role.

All generated artifacts were version-controlled using Git with the repository hosted on GitHub (arkangelai/grants, private). Researcher review was required before any proposal was submitted or any scoring decision was acted upon. Technical details of the model configuration, including provider, model identifier, runtime, and reasoning settings at the time of execution, are provided in Additional file 1.

Skill modules and automation

The agent behavior was not implemented as a single prompt but as a set of versioned, reusable skill modules, each encapsulating the instructions, conventions, and guardrails for one stage of the pipeline. A skill is a self-contained specification that the language model loads on demand, so that the same task is performed consistently across runs and operators. Skills are stored and versioned alongside the repository, which means that changes to the workflow are reviewed and traceable in the same way as any other artifact. Two activation mechanisms were used. First, scheduled (cron) triggers launched time-based skills automatically; for example, the daily scouting skill runs each morning and the pipeline-reporting skill runs weekly. Second, label-based routing activated the appropriate skill in response to a state change on a repository issue: when a researcher applies a disposition label to an opportunity, the corresponding downstream skill is invoked. This design keeps the human researcher in control of state transitions while delegating the repetitive drafting and bookkeeping work to the skills.

The skill modules map onto the four workflow phases as follows. In the discovery phase, a daily scouting skill expands search terms by institutional context, inspects funder pages, and creates one issue per candidate opportunity. In the triage phase, a triage skill applies the structural-eligibility checklist, scores the opportunity against the Win Readiness rubric, deduplicates against prior records, and proposes a disposition label. In the drafting phase, a dispatcher skill routes each approved opportunity to a draft-generation skill selected by grant type (Table 4): a commercial or innovation-prize skill for company-level and challenge submissions, a research-grant skill for investigator-led scientific calls, an outreach skill for partner-route and letter-of-intent contacts, and a pitch-deck skill for opportunities requiring a slide submission. In the tracking phase, a submission-tracking skill and a pipeline-reporting skill update the application history and the weekly pipeline report. Each skill embeds the governance rules described below, so that constraints such as citation verification and confidentiality are enforced at every invocation rather than left to the operator's memory.

Data sources

The workflow description was reconstructed from internal and public sources. The internal sources were held in the organization's version-controlled repositories: the manuscript-writing issue in the papers repository; the issue records and pull requests in the grants repository documenting recent scouting,

triage, and submission activity; the Win Readiness Score rubric and dashboard in shared resources; and the curated scouting-state file. Public sources included literature on the use of large language models in grant writing, funders' evaluations of AI in health, and digital health sustainability in low- and middle-income countries.

Analytical approach

The primary analytical framework for this paper is operational feasibility. Feasibility was defined as the workflow's ability to convert broad funding searches into a managed, prioritized, and traceable opportunity pipeline while preserving researcher review, eligibility gating, decision-aware deduplication, and artifact preservation. Feasibility did not require that any specific grant be won. The operational metrics reported here are throughput and process metrics, not outcome metrics.

The primary implementation outcomes assessed were the number of search combinations tested, opportunities reviewed, and funding scenarios represented. Process throughput was measured as the number of suitable grants identified per scouting cycle. Log-derived metrics included sources scanned, candidates filtered, net-new issues created, and duplicates detected across structured scouting runs. Pipeline state was assessed by the number of grants in ready-to-submit and active pipeline states, and by the presence of explicit scoring, deduplication, and researcher review checkpoints.

Workflow architecture

The workflow has four operational phases (Fig. 1). A governance layer spans all phases: researcher review checkpoints are required at the discovery, triage, scoring, and submission readiness stages. Artifacts generated at each phase are preserved in version-controlled repositories.

Phase 1: Opportunity discovery

The discovery phase searches broadly across funder categories and program types. Search strategies are not limited to generic phrases such as "health AI grant." They include context-based searches anchored in the organization's products, geography, partners, disease areas, funder type, and implementation needs. The purpose is to surface opportunities that may not appear in standard grant databases, including corporate innovation programs, health-system accelerators, government mechanisms, and partner-route funding.

The pipeline explicitly tested more than 50 search combinations and reviewed more than 200 opportunities. This breadth was necessary because the relevant opportunity set for a Colombian health-AI company is fragmented across scientific funding, innovation funding, digital health programs, corporate partnerships, public-sector mechanisms, and regional initiatives. The primary driver of the throughput increase was an operational change termed aggressive search, which introduced deeper funder-page inspection, within-site navigation, and institutional context queries, as reported in the Results section. Discovery ends at the first researcher checkpoint: the research team confirms basic

eligibility before an opportunity is admitted to the pipeline, so that clearly ineligible programs are not carried into triage.

Phase 2: Eligibility and fit triage

After discovery, opportunities enter triage. Triage evaluates legal entity eligibility, geography, sector fit, company stage, entity type, ask size, partner requirements, and strategic relevance. Opportunities are moved into several states: start-draft (approved to begin drafting), vigilar-partner (eligible but blocked until a partner is identified), Letter of Intention (pursued through direct outreach rather than a formal portal), discarded (declined with a documented reason), expired (deadline passed), or no fit (out of scope). These are the repository's actual issue labels (Additional file 3). Opportunities to revisit later are recorded in the scouting-state file as monitor-by-date or do-not-resurface (DNR) entries rather than as labels, preserving institutional memory outside the issue-label system.

A curated scouting state file preserves institutional memory. This prevents the same ineligible or strategically weak opportunities from being repeatedly rediscovered. It also separates structural discards from paused opportunities. For example, a program may be excluded permanently if it is restricted to nonprofits or European entities, while another may remain active if it requires a US academic partner, a European consortium, or a local educational institution.

The issue documentation identified three recurring unblockers: a US national partner for federal opportunities, a European partner for European Union mechanisms, and Colombian educational institutions for domestic or academic routes. These are treated as structural pipeline constraints rather than one-off narrative edits.

Triage is the second researcher checkpoint. The disposition label assigned to each opportunity is a human decision: although the triage skill proposes a label, the research team confirms or overrides it before the opportunity advances, so that eligibility interpretation and partner-route judgments remain under human control.

Phase 3: Win Readiness scoring

The pipeline uses a Win Readiness Score to prioritize opportunities. The score is deliberately not described as a calibrated probability because the organization does not yet have enough outcome data to estimate true win probability. Instead, it serves as a relative ranking and kill-or-go tool.

The score is defined as:

Win Readiness Score = Eligibility_Gate × Fit × Base_Rate_Funder × Stage_Bonus × 100

Variable definitions and operational ranges are provided in Additional file 2. The eligibility gate is binary or borderline. Fit evaluates mission alignment, evidence base, traction, geography and population fit,

team strength, ask size, effort-to-probability ratio, and product fit. Base rate uses public acceptance-rate information when available or category defaults when unavailable. Stage bonus increases priority for funders with prior positive signals, such as interview advancement, substantive feedback, or an active customer or partner relationship.

At the time of manuscript preparation, five opportunities had advanced to interview or finalist stages across a pipeline of 49 scored grants. These advances informed the iterative refinement of the scoring rubric but were not used to estimate the probability of winning a grant.

Scoring constitutes the third researcher checkpoint. The Win Readiness Score orders opportunities, but the decision to commit drafting effort to a given opportunity, or to set it aside, is taken by the research team rather than by the score alone.

Phase 4: Proposal package drafting and artifact preservation

Opportunities that pass triage and prioritization move into drafting. The drafting process uses proposal folders, issue references, source files, official links, outreach drafts, status files, and final submission artifacts. Large language models support summarization, section drafting, narrative adaptation, and consistency checks, but the workflow keeps researcher review as the decision point for scientific claims, confidential information, partner strategy, and submission readiness. Drafting therefore contains the fourth and final researcher checkpoint, a submission-readiness review, at which the research team verifies factual claims and citations, confirms that no confidential material is exposed, and approves the package before it is sent to the funder.

The workflow uses repository commits and pull requests to preserve artifacts. A submitted package, outreach email, partner narrative, or scoring rationale may need to be recovered months later when a funder responds, a similar opportunity reopens, or a partner requests prior material. Version control makes this recovery possible without relying on institutional memory.

Governance rules

The workflow incorporates governance rules derived from grant-writing guidance [8, 14] and operational experience: model output is never treated as authoritative; funder eligibility is verified from official sources; source links and issue rationales are preserved; confidential proposal materials are not exposed unnecessarily; acceptance rates and partner commitments are not invented; and all submitted files are preserved in version control. A researcher review is required before any proposal is submitted. These rules align with published recommendations that large language models can assist grant development but require fact-checking, privacy awareness, and researcher ownership of scientific and strategic content [8].

Ethics

This study did not require informed consent or ethics committee approval. No patient or participant data were collected or analyzed. The workflow description draws exclusively on organizational process artifacts, internal repository records, version histories, and issue-level workflow logs, with no identifiable personal or clinical information. This type of methodological process description is classified as research without risk under Colombian Resolution 8430/1993 [15].

Results

Across 10 recorded scouting cycles between 2026-05-29 and 2026-06-05, the pipeline tested over 50 search combinations across 20 funding scenarios (Table 1), processing 741 sources and identifying 352 pre-filter candidates. Of these, 83 were retained as post-filter candidates after applying eligibility and fit criteria, generating 46 net-new pipeline issues. A total of 220 duplicates were detected across cycles (Fig. 2). Opportunities spanned heterogeneous funder categories, including global health foundations, government agencies, pharmaceutical and medtech programs, payer-related opportunities, compute-related programs, innovation prizes, bilateral mechanisms, federal and state-level US opportunities, Hispanic health programs, corporate innovation programs, and partner-mediated academic routes. The largest yield occurred on 2026-06-01, when 180 sources generated 76 pre-filter candidates, 43 post-filter candidates, and 30 net-new issues. Later runs showed lower net-new yield and higher duplication, consistent with saturation of the immediate search space.

Table 1
Operational metrics from the structured scouting log and pipeline state.

Metric	Value	Interpretation
Search combinations tested	50+	Breadth of search strategy
Opportunities reviewed	200+	Broad initial opportunity surface
Funding scenarios	20	Heterogeneous funder categories
Suitable grants before search strategy change	~ 5–6 per day (pre-change rate)	Baseline operational throughput per team report
Suitable grants after search strategy change	~ 18 per scouting cycle	Post-change reported throughput per team report
Active near-term submission candidates	7	Near-term submission portfolio
Grants in issue pipeline	11	Active managed pipeline
Structured scouting runs analyzed	10	Log-derived process evidence
Sources scanned	741	Search scale
Pre-filter candidates	352	Candidate capture
Post-filter candidates	83	Triage narrowing
Net-new issues created	46	New actionable opportunities
Duplicates detected	220	Saturation and deduplication burden
Applications advanced to next stage	5	Interview or finalist stage as of manuscript preparation
Note: Repository activity after the analytic cutoff date (2026-06-09) is not reflected in the metrics above.		

Scouting throughput

The issue protocol reported a change in suitable-grant throughput after three search strategy changes: deeper inspection of funder pages, navigation within funder websites, and context-based institutional search. The team reported identifying approximately 18 suitable grants per scouting cycle after those changes, compared to approximately 5 to 6 per day under the prior approach. These figures use different units of measurement (scouting cycle versus daily rate) and are based on internal operational records; they should not be interpreted as a controlled before-after comparison, but as the team’s operational characterization of a change in search strategy

This breadth matters because a health-AI company in Colombia rarely fits a single funding pathway. Some opportunities are open to for-profit applicants, some require US or European academic leads,

some are innovation prizes with low acceptance rates, and some are effectively business-development routes rather than grants. A single undifferentiated grant list would obscure those differences. The implemented workflow classified opportunities by route and decision state rather than treating every funder as equally actionable (Table 2).

Table 2
Core workflow components.

Component	Purpose	Main artifact
Context-based discovery	Search beyond generic grant terms	Opportunity issues and scouting logs
Eligibility triage	Prevent work on structurally impossible grants	Labels, issue comments, do-not-resurface state
Partner-route assessment	Identify opportunities requiring US, European, Colombian, or clinical partners	Partner-needed list and issue state
Win Readiness Score	Rank opportunities by relative readiness	Scoring rubric and dashboard
Proposal drafting	Convert prioritized opportunity into package	Proposal folder, drafts, official links
Researcher review	Preserve scientific and strategic accountability	Issue approvals, pull requests, final submission files
Artifact preservation	Make prior work reusable and auditable	Repository commits, pull requests, status files

Table 3 summarizes the source artifact, date range, denominator, and verification status for each metric reported in this study. Scouting throughput was verified directly from the structured log across 10 recorded runs. Opportunity pipeline state and Win Readiness Scores were exported from the scoring dashboard on 2026-06-10 and verified against repository records. Application history was reconstructed retrospectively from the submission tracker with a cutoff date of 2026-06-09. Advancement to next stage was team-reported for 5 opportunities tracked through issue comments.

Table 3
Data sources, date ranges, and verification status for each reported metric.

Metric domain	Source artifact	Date range	Denominator	Status
Scouting throughput	Structured scouting log	2026-05-29 to 2026-06-05	10 recorded runs	Verified from log
Opportunity pipeline state	Repository issue labels	Exported 2026-06-10	49 scored grants	Verified from repository
Application history	Submission tracker	March to June 2026 (cutoff 2026-06-09)	52 submitted applications	Retrospective operational estimate
Advancement to next stage	Issue comments and status updates	March to June 2026 (cutoff 2026-06-09)	5 opportunities	Team-reported
Win Readiness Score	Scoring dashboard	Exported 2026-06-10	49 scored grants	Verified from dashboard

Triage and deduplication

The curated scouting state file functioned as a memory layer for the pipeline. It captured do-not-resurface decisions, monitor dates, and partner-needed opportunities. This design addresses a common problem in grant discovery: generic searches repeatedly surface the same ineligible or weak-fit programs, consuming reviewer time without improving the pipeline.

The state file preserved the rationale for exclusion. Reasons included nonprofit-only restrictions, European or national-residency requirements, disease-scope mismatch, therapeutics-only requirements, lack of partner route, operational cloud-credit programs that did not belong in a grant pipeline, and equity-only investor programs that should be handled outside research development.

This approach changed deduplication from simple text matching into decision-aware triage. A program could be excluded not because it had been seen before, but because a prior human decision determined that it was structurally ineligible, strategically weak, or not worth resurfacing until a specific date.

Win Readiness Score

The Win Readiness Score introduced a transparent ranking mechanism. The score combined eligibility, institutional fit, funder base rate, and stage bonus. It was explicitly framed as a relative readiness score rather than a true probability. This distinction is important because early outcome counts were insufficient for probability calibration.

The scoring dashboard organized opportunities into actionable categories: advanced to next stage, submitted and awaiting response with score above 5%, submitted with score 3 to 5%, low-score

exploratory submissions, ineligible or discarded, business-development outreach, and historical outcomes. This structure made prioritization visible to the team and reduced reliance on subjective urgency alone. In the repository, these score bands are recorded as the priority labels grade-a through grade-d (and P0 through P3), while disposition categories map to the actual issue labels discarded and Letter of Intention (Additional file 3).

At the time of the rubric, five opportunities had advanced to a next stage, including three at interview stage and two at finalist stage. These advances were used to justify stage bonuses but not to claim a grant-winning rate. Among the submitted applications, one opportunity (MIT Solve) did not advance past the May 2026 cohort selection; the next application cycle opens in September 2026, and the pipeline retains the funder profile and prior application artifacts for reuse. The score therefore served as a disciplined way to allocate drafting effort rather than a guarantee of funding success.

Automation coverage and researcher effort

Although this study did not run a controlled time-and-motion comparison, the operational records support retrospective operational estimates for the skill-based workflow relative to the previous manual process. Under the manual process, a researcher scanning funder portals, newsletters, and referrals identified on the order of five to six eligible opportunities per working day and performed triage, scoring, deduplication, and record-keeping by hand. Under the skill-based workflow, the daily scouting skill surfaced and pre-screened candidate opportunities automatically, and the median time from discovery to a documented go/no-go decision was under one hour, with a higher mean (on the order of eleven hours) reflecting a long tail of opportunities held for partner or eligibility clarification. Time from the start-draft decision to submission ranged from approximately one to twenty-nine days, with a mean of roughly ten to twelve days, depending on form complexity and whether a partner or supporting letter was required. These estimates are derived from internal workflow records (issue timestamps, commit histories, and submission logs) and were not independently validated; they should be interpreted as retrospective operational estimates, not as controlled measurements. The principal effort saving was concentrated in the repetitive discovery, summarization, first-draft, and bookkeeping tasks; researcher time was redirected to the four review checkpoints (eligibility, scoring, draft review, and submission readiness) that the workflow deliberately retains as human decisions.

Grant opportunities were not handled by a single drafting routine. Each opportunity was classified by type, and the dispatcher routed it to the skill best suited to that type. Table 4 lists the grant-type classifications used in the pipeline and the corresponding skill module.

Table 4
Grant-type classifications and the corresponding skill module.

Grant type	Typical route or submission form	Skill module
Innovation prizes and company challenges	Company-level application or challenge portal	Commercial and innovation-prize drafting skill
Research and scientific grants	Investigator-led scientific call, often partner-mediated	Research-grant drafting skill
Partner-route opportunities and letters of intent	Outreach email or letter of intent to a prospective partner	Outreach and first-contact skill
Accelerators and pitch-based programs	Slide-deck submission and pitch event	Pitch-deck generation skill
All types (cross-cutting)	Submission tracking and weekly pipeline reporting	Tracking and pipeline-reporting skills

Discussion

This paper describes a governed grant development workflow for a health-AI company operating in a funding environment where opportunities are fragmented, eligibility rules are heterogeneous, and evidence-generation needs are urgent. The main contribution is methodological. The workflow converts grant discovery from an ad hoc search process into a traceable pipeline with structured triage, decision-aware deduplication, relative scoring, and repository-based artifact preservation.

The same large language model capabilities that accelerate grant discovery and proposal drafting have prompted a governance response from publishers, funders, and institutional review bodies. Major publishers have established policies on AI use in research outputs, explicitly addressing accountability, transparency, and the risk that generated content may not be traceable to verifiable sources [7, 16, 17]. These policies shift accountability to the researcher: AI assists in preparation, but the researcher owns every scientific claim in the final product. The same principle applies to grant applications. A proposal that uses large language models for discovery or drafting but cannot trace its factual claims to verified sources, or that was submitted without explicit researcher review, creates both integrity risk and reputational risk for the applying organization. A governed workflow addresses this not by restricting the use of large language models, but by ensuring that every output passes through an auditable researcher review checkpoint before use.

Main findings

Throughput improved only after the search strategy became more contextual. Generic grant searches are insufficient for a company like Arkangel AI. Relevant opportunities may appear under AI, clinical decision support, digital health, health equity, antimicrobial resistance, oncology, payer innovation, women-led entrepreneurship, corporate innovation, regional development, compute access, or partner-route academic funding. The pipeline therefore needed to search by institutional context, not only by

topic. This is likely relevant to other low- and middle-income country health-AI teams whose funding fit depends on geography, entity type, partner relationships, and disease-specific evidence [1, 5].

Beyond search strategy, deduplication emerged as a governance function rather than a data-cleaning step. A naive search system may repeatedly return the same ineligible grants. The workflow addressed this by storing reasons for do-not-resurface, monitor, partner-needed, and business-development routing. This preserved researcher judgment and made future searches more efficient. In practical terms, the pipeline learned which opportunities should not consume reviewer time again.

Scoring, in turn, must be honest about uncertainty. The Win Readiness Score is useful precisely because it avoids being called a probability. Early pipelines rarely have enough outcomes to estimate calibrated win rates [18]. A readiness score can still improve prioritization if it uses explicit inputs and is updated when outcomes accumulate.

These operational components require explicit boundaries when large language models support the drafting process. Published guidance warns against hallucinated references, privacy leaks, overreliance, and funder-policy violations [8, 19]. The governance rules described in the Methods section operationalize these boundaries as workflow constraints rather than discretionary checks. Because these constraints are embedded in the skill modules rather than applied manually, they are enforced automatically on every invocation: each drafting skill verifies citations, blocks disclosure of confidential funder communications and unpublished strategy, and flags unresolved eligibility or partner dependencies for the researcher rather than fabricating content. Encapsulating the guardrails in versioned skills also makes them auditable, since any change to a constraint is reviewed and tracked in the same repository as the proposals themselves [10].

Implications for health-AI organizations in low- and middle-income settings

For health-AI teams in low- and middle-income countries, the workflow has broader relevance. Digital health sustainability literature emphasizes financing barriers, fragmented markets, and the need for strategic partnerships [1, 5, 20, 21]. Health-AI organizations frequently operate without dedicated research development offices or grant operations infrastructure. Evidence generation must coexist with product development cycles, limited staff, and institutional pressure to demonstrate clinical or commercial relevance before evidence is available [22].

A grant pipeline is one way to manage those barriers. It helps identify which funders support evaluation, which require local or international partners, which opportunities are too low-fit, and which narratives need more evidence before submission. In this sense, grant development is part of implementation capacity, not merely fundraising administration. A governed workflow that uses large language models for search expansion and drafting while preserving researcher review and traceability can partially substitute for the research development operations resources that many health-AI teams in low- and middle-income countries cannot maintain.

A workflow without issue-based state tracking, deduplication memory, or citation audit checkpoints risks producing output that lacks a recoverable decision history, cannot be verified against eligibility criteria, and may include fabricated references that would only be detected by the funder [19]. Governance components address these risks not by restricting model use but by adding structured accountability at each step.

To support adoption by other teams, Table 5 distinguishes workflow components that are broadly transferable from those specific to the organization’s context.

Table 5
Transferable versus organization-specific workflow components.

Component	Transferable	Notes
Context-based discovery approach	Yes	Any team can anchor searches in its product areas, geography, and disease focus
Triage state logic (proceed, monitor, partner-needed, do-not-resurface)	Yes	Labels and states are adaptable to any issue-based tracking system
Win Readiness Score concept	Yes	The eligibility-fit-base rate-bonus structure is generalizable
Decision-aware deduplication	Yes	The principle of storing discard reasons applies regardless of tooling
Governance rules	Yes	Privacy, source preservation, researcher review requirements apply universally
Artifact preservation via version control	Yes	Any team can adapt the commit-and-link pattern
Specific scoring weights	No	Calibrated to Arkangel’s product portfolio, geography, and evidence stage
US partner requirement	No	Specific to US federal funding eligibility constraints
European partner requirement	No	Specific to EU funding mechanisms
Arkangel repository structure	No	Implementation detail; other teams will use their own tooling
Colombian regulatory framing	No	Other organizations follow their jurisdiction’s requirements

Teams adopting a similar workflow should expect to recalibrate scoring weights, partner routes, and triage criteria to their own organizational context.

The workflow did not remove researcher decision-making. Researcher review remained necessary for eligibility interpretation, partner strategy, scientific claims, privacy constraints, final submission readiness, and decisions to discard or monitor an opportunity. The large language model-supported

components increased drafting and search capacity, but governance depended on explicit labels, issue comments, pull requests, and researcher review.

This is consistent with published grant-writing guidance: large language models may support clarity, feedback, and iteration, but no model can replace funder-specific judgment, scientific ownership, or expert review [8].

Existing guidance on AI in grant writing focuses mainly on how individual researchers should use large language models responsibly [8]. That guidance is necessary but incomplete for organizations. An organization also needs operating infrastructure: opportunity state, source tracking, scoring, versioned drafts, partner-route memory, and submission records. This manuscript extends the conversation from individual prompt use to team-level grant operations.

Teams seeking to adopt a similar workflow should start with governance rather than automation. The minimum viable system is not a model interface. It is a structured opportunity database with clear labels, official sources, eligibility rules, discard logic, and human review points. Large language models can then support search expansion, summarization, first-pass drafting after human framing, and consistency checks.

Partner requirements also deserve treatment as first-class data. For Arkangel AI, United States, European, and Colombian institutional partners were not incidental: they were recurring unlockers for entire classes of funding. Recording these structural blockers helps the business-development and evidence teams coordinate and avoids repeated discovery of the same structurally closed opportunities [4, 21].

Equally important, artifact preservation should happen by default. Every submitted package, outreach narrative, funder brief, and scoring rationale should be recoverable. This reduces rework and supports future submissions when similar calls reopen [3]. Version-controlled repositories provide the infrastructure; the governance rules provide the discipline.

Strengths and limitations

This study has several strengths. The workflow was developed and evaluated using real operational artifacts from a functioning grant pipeline, not from a hypothetical design. All process metrics were derived from version-controlled repository records, providing a traceable evidence base. The structured scouting log offers granular, run-level data on search yield, filtering, and deduplication. The workflow explicitly preserves four researcher review checkpoints, documenting where human judgment remains required rather than claiming full automation.

The study also has limitations that constrain interpretation. It describes a single-company workflow whose categories, scoring weights, and partner routes reflect one organization's products, geography, and evidence portfolio. Other teams would need to adapt the rubric, and results should not be generalized to teams with different institutional contexts or funding environments. The current

evaluation is operational, not outcome-based: throughput, issue creation, and duplicate detection do not prove higher proposal quality or funding success. Future versions should measure researcher time, reviewer scores, submission rates, advancement rates, awards, and amount secured.

Some source material was missing from the repository at the time of drafting, and this analysis relied on the structured metrics available rather than expanding unsupported transcript details. The Win Readiness Score includes heuristic inputs, since public acceptance rates are not always available and adjustments for for-profit or foreign entities can only approximate funder behavior; the score should remain a prioritization tool until sufficient outcome data allow calibration. Privacy and confidentiality constraints also limit how much actual proposal text and partner strategy can be shared in a journal manuscript.

Repository metrics reflect operational records maintained for workflow management rather than prospectively collected research data, and denominators and time windows vary across metrics as specified in the Methods. The throughput improvement from approximately 5 to 6 grants per day to 18 per cycle reflects a change in search strategy rather than a controlled comparison and should not be interpreted as a causal estimate of workflow impact. The formative evaluation was conducted by members of the development and writing team, which limits independence; however, internal formative evaluation is appropriate at this stage because the objective was to describe system design and identify governance gaps. The analyzed cases represent completed writing programs and may not capture framework behavior during projects that were paused, canceled, or substantially redirected by human decision-making. Finally, the large language model citation audit checkpoint applies to research outputs produced by the system, but grant proposals submitted to external funders require additional manual verification for every factual claim, since hallucinated citations in a grant application are a credibility risk that automated checks may not fully prevent [19].

Future evaluation

The next version of this work should evaluate outcomes across repeated cycles. Useful metrics include time from opportunity discovery to go-or-no-go decision, time from start-draft to submission-ready package, proportion of opportunities discarded for structural ineligibility, proportion requiring partner route, researcher review burden, number of submitted grants, proportion advancing to next stage, funding secured, reviewer feedback quality, and reuse rate of prior proposal assets. These measures should be specified prospectively using implementation-outcome constructs such as feasibility, adoption, fidelity, acceptability, and sustainability [18].

Several of these measures can already be reported for the current operating period (March to June 2026) and are presented here as baseline values rather than as validated outcomes. Time from opportunity discovery to a documented go/no-go decision had a median under one hour and a mean of approximately eleven hours, reflecting a long tail of opportunities held for clarification. Time from the start-draft decision to a submission-ready package ranged from approximately one to twenty-nine days (mean roughly ten to twelve days). Approximately 88% of discarded opportunities were rejected for

structural ineligibility rather than weak fit, and roughly half of the eligible opportunities required a partner or letter-of-intent route. The researcher review burden corresponded to four explicit human checkpoints per opportunity (eligibility, scoring, draft review, and submission readiness). As of the analytic cutoff date, 52 applications had been submitted during the operating period (March to June 2026), and 5 of the active opportunities had advanced to an interview or finalist stage; subsequent repository activity was not included in this analysis. Reuse of prior proposal assets was substantial, with reusable content carried forward across related opportunities in roughly two-thirds to nine-tenths of cases. The remaining measures (funding secured and independent reviewer-feedback quality) require longer follow-up and are deferred to the prospective evaluation described above. These baseline values are retrospective operational estimates derived from internal workflow records and were not independently validated.

A stronger evaluation could compare pre-implementation and post-implementation periods. It could also sample proposal packages for independent review using criteria such as funder fit, specificity, evidence grounding, clarity, budget coherence, and risk of generic language. This would test whether increased throughput compromises or improves quality.

Conclusions

A governed, large language model-supported grant development workflow can make opportunity discovery and proposal preparation more systematic, auditable, and scalable for health-AI teams operating in resource-constrained settings. The workflow converts broad grant searches into a managed pipeline with structured triage, decision-aware deduplication, relative scoring, artifact preservation, and explicit researcher review checkpoints at every decision stage. Concretely, the workflow retains four human checkpoints, one in each phase: an eligibility check during discovery, a scoring and prioritization check during triage, a draft-review check before a proposal is considered ready, and a submission-readiness check before anything is sent to a funder. Early operational data support the feasibility of this approach for a Colombian health-AI company, but do not establish that automation improves grant success rates. The next evidence step is to evaluate proposal quality, researcher effort, reviewer outcomes, and funding results across multiple cycles, and to assess whether the workflow generalizes to teams with different funding environments and organizational resources.

Declarations

Ethics approval and consent to participate

This study did not require ethics committee approval or participant consent. No patient or participant data were collected or analyzed. The workflow description draws exclusively on organizational process artifacts, internal repository records, version histories, and issue-level workflow logs, with no identifiable personal or clinical information. This type of methodological process description is classified as research without risk under Colombian Resolution 8430/1993.

Consent for publication

Not applicable.

Competing Interests

Katherine Monsalve, Natalia Castaño-Villegas, Jose Zea, and Laura Velásquez are employees or co-founders of Arkangel AI, the organization whose internal workflow is described in this manuscript. The paper describes internal organizational methodology and does not evaluate competing products or services.

Funding

This research received no external funding. The workflow described was developed as part of the internal research and evidence development operations of Arkangel AI.

Author Contribution

KMB conceptualized the study, drafted all sections of the manuscript, and coordinated the submission process. NCV implemented the grants workflow, contributed operational data, and reviewed the final manuscript. JZ contributed to system architecture design and review of technical sections. LV provided strategic oversight and final review of the manuscript. AI-assisted drafting tools were used to prepare initial text under human review; the corresponding author takes full responsibility for the integrity of the content. All authors read and approved the final manuscript.

Acknowledgement

Not applicable.

Data Availability

The operational scouting logs and pipeline state data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to their proprietary nature and the inclusion of unpublished grant application materials.

References

1. World Health Organization. Global strategy on digital health 2020–2025. Geneva: World Health Organization. (2021). Available from: <https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf>

2. Greenall-Ota, J., Yapa, H. M., Fox, G. J. & Negin, J. Qualitative Evaluation of mHealth Implementation for Infectious Disease Care in Low- and Middle-Income Countries: Narrative Review. *JMIR Mhealth Uhealth*. **12**, e55189. 10.2196/55189 (2024).
3. Alami, H. et al. Artificial Intelligence Governance in Health Systems: Systematic Review of Frameworks and Integrative Model Proposal. *J. Med. Internet Res.* **28** (1), e87448. 10.2196/87448 (2026).
4. Bill and Melinda Gates Foundation. Jun 10. New Global Partnership Launches Call for Proposals to Evaluate AI Tools in Primary Health Care in Low- and Middle-Income Countries. Published 2026 Feb 20. Accessed (2026). Available from: <https://www.gatesfoundation.org/ideas/media-center/press-releases/2026/02/ai-impact-health>
5. Irihamye, E. et al. Sustainable by Design: Digital Health Business Models for Equitable Global Health Impact in Low-Income and Low-Middle-Income Countries. *Mayo Clin. Proc. Digit. Health*. **3** (4), 100261. 10.1016/j.mcpdig.2025.100261 (2025).
6. Liang, W. et al. Quantifying large language model usage in scientific papers. *Nat. Hum. Behav.* **9** (5), 928–943. 10.1038/s41562-025-02273-8 (2025).
7. Stokel-Walker, C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* **18** 10.1038/d41586-023-00107-z (2023 Jan).
8. Seckel, E., Stephens, B. Y. & Rodriguez, F. Ten simple rules to leverage large language models for getting grants. *PLOS Comput. Biol.* **20** (3), e1011863. 10.1371/journal.pcbi.1011863 (2024).
9. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization. (2021). Available from: <https://www.who.int/publications/i/item/9789240029200>
10. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST, 2023). 10.6028/NIST.AI.100-1
11. Shaw, J. et al. Research ethics and artificial intelligence for global health: perspectives from the global forum on bioethics in research. *BMC Med. Ethics*. **25** (1), 46. 10.1186/s12910-024-01044-w (2024).
12. Jimenez, D. et al. PANDORA: An AI model for the automatic extraction of clinical unstructured data and clinical risk score implementation. In: IEEE Conference on Artificial Intelligence (CAI); 2025. p. 1682. (2025). 10.1109/CAI64502.2025.00280
13. Villa, M. C. et al. Arkangel AI: A conversational agent for real-time, evidence-based medical question-answering. *Intell. Based Med.* **12**, 100274. 10.1016/j.ibmed.2025.100274 (2025).
14. Hryciw, D. H. & Tangney, M. Writing with AI in academia: ethical considerations and practical guidance. *Adv. Physiol. Educ.* **48** (4), 791–800. 10.1152/advan.00140.2024 (2024).
15. Colombia Ministerio de Salud. *Resolution 8430 of 1993, establishing scientific, technical and administrative standards for health research* (Ministerio de Salud, 1993).
16. Biswas, S. ChatGPT and the Future of Medical Writing. *Radiology* **307** (2), e223312. 10.1148/radiol.223312 (2023).

17. Ganjavi, C. et al. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *BMJ* **384**, e077192. 10.1136/bmj-2023-077192 (2024).
18. Proctor, E. K. et al. Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Adm. Policy Ment Health*. **38** (2), 65–76. 10.1007/s10488-010-0319-7 (2011).
19. Resnik, D. B. & Hosseini, M. Hallucinated citations produced by generative artificial intelligence may constitute research misconduct when citations function as data in scholarly papers. *Acc. Res.* **15** 10.1080/08989621.2026.2645390 (2026 Mar).
20. Wahl, B., Nassiri-Ansari, T., Redpath, D. D., Allotey, P. & Schwalbe, N. Artificial intelligence for research capacity strengthening: Two reviews and a pathway to shift power in global health. *PLOS Digit. Health*. **5** (4), e0001302. 10.1371/journal.pdig.0001302 (2026).
21. Kiosia, A. et al. Current data science capacity building initiatives for health researchers in LMICs: global and regional efforts. *Front. Public. Health*. **12**, 1418382. 10.3389/fpubh.2024.1418382 (2024).
22. Sutton, R. T. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit. Med.* **3** (1), 17. 10.1038/s41746-020-0221-y (2020).

Figures

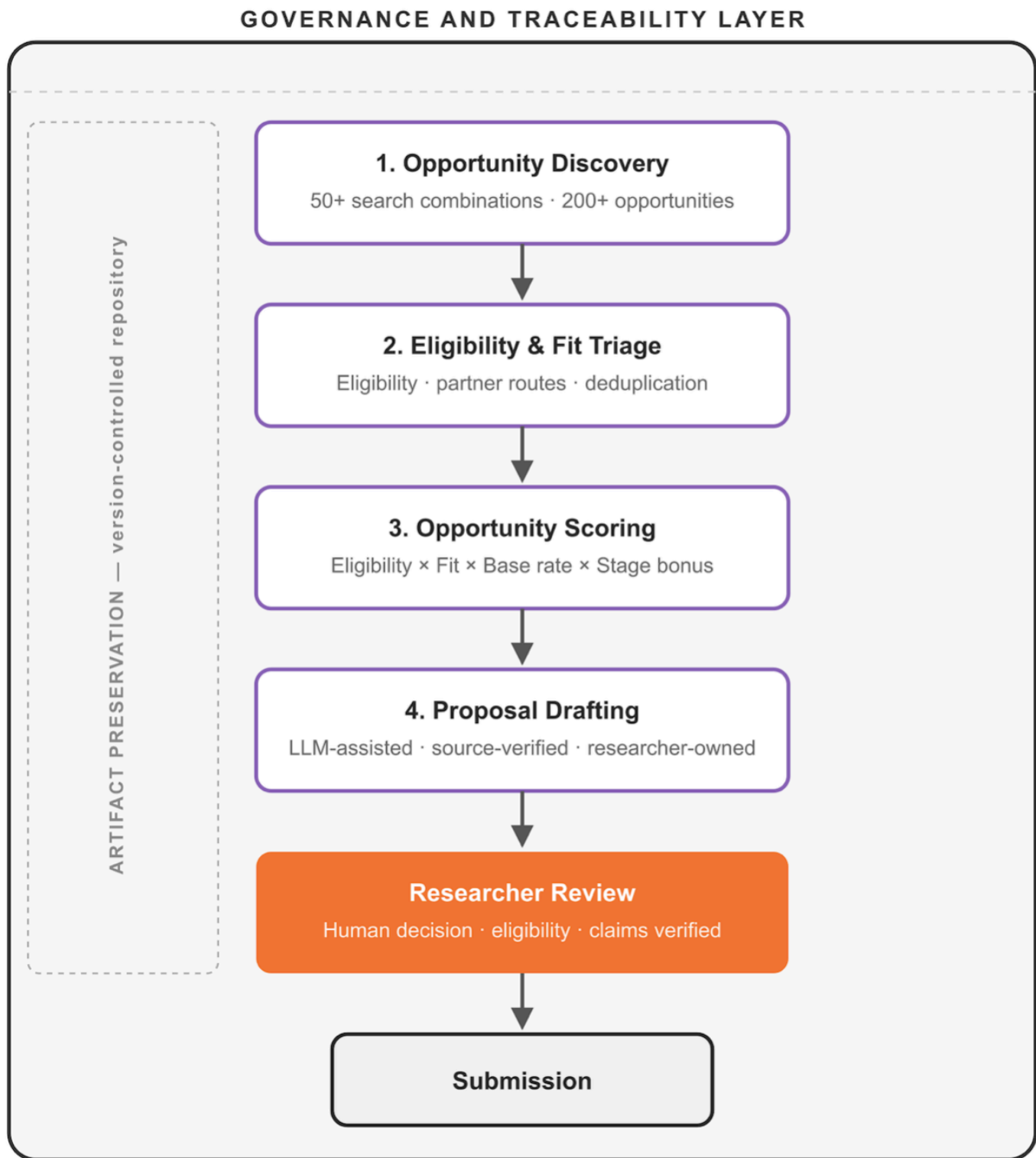


Figure 1

Governed grant development workflow. The four phases (discovery, triage, scoring, and proposal drafting) operate within a governance and traceability layer. All artifacts are preserved in a version-controlled repository, and a researcher review step verifies eligibility, sources, and claims before submission.

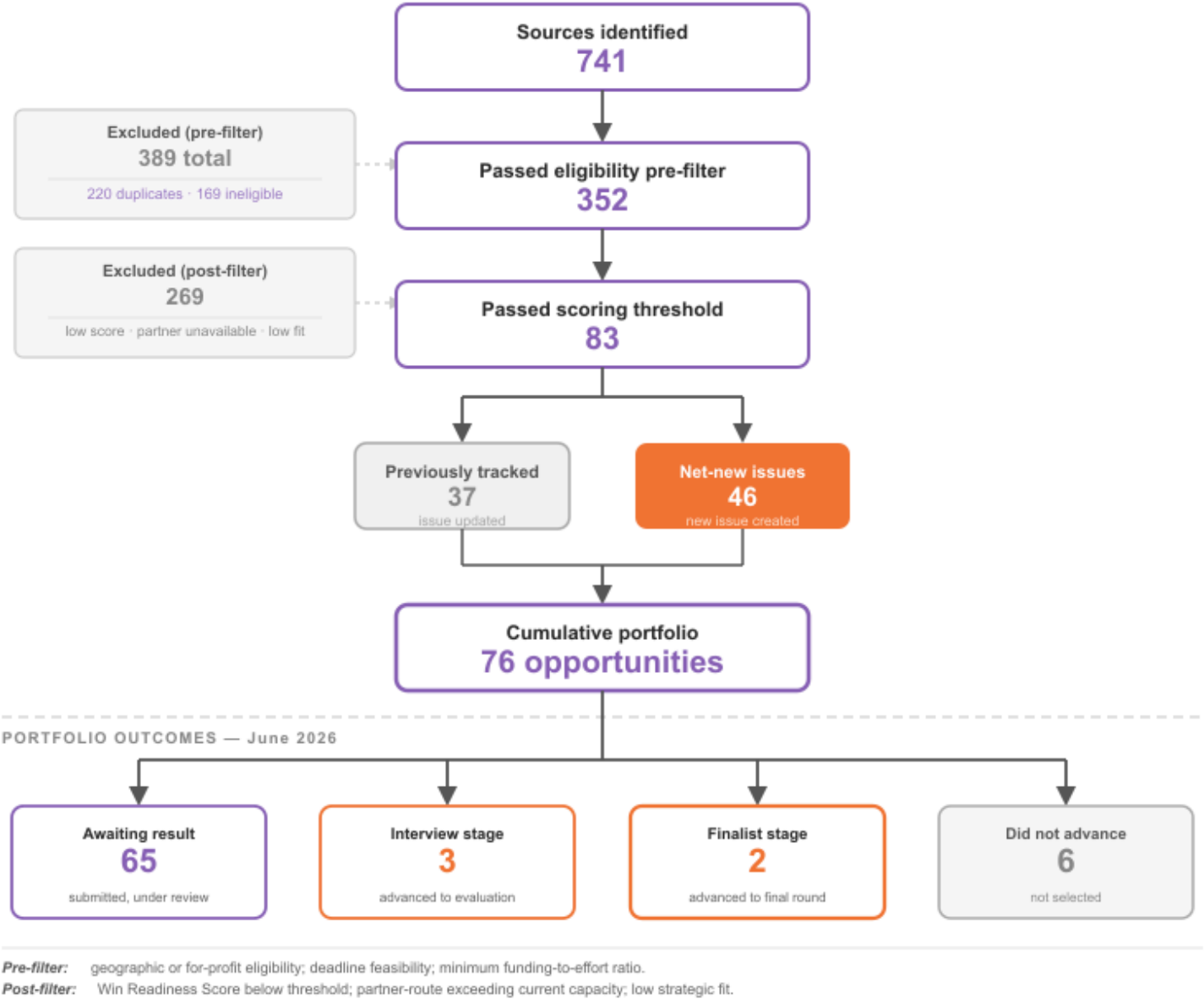


Figure 2

Opportunity scouting funnel and portfolio outcomes.

The left side of the figure summarizes the scouting process across 10 structured cycles (2026-05-29 to 2026-06-05). The right side shows portfolio outcomes as of June 2026 across the full operating period (March to June 2026). Not all portfolio outcomes are attributable to these 10 cycles; earlier discovery and triage activity also contributed to the current pipeline state.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.docx](#)