Outreach Priority Score for EdTech Startup

Evan Beck ecb9981@nyu.edu
Jovita Gandhi jg8316@nyu.edu
Simran Khullar sk12108@nyu.edu
Dipesh Tharu Mahato dm6259@nyu.edu

Abstract

IterLight, an ed tech start up, must prioritize outreach across many plausible funders despite severe small-data constraints: only a small set of organizations can be confidently labeled as positive EdTech funders, while negative labels are largely unavailable[2]. We develop an Outreach Priority Scoring (OPS) framework that estimates a probability-like engagement score, denoted $p(x)$, and combines it with expected grant capacity to construct an expected-value ranking of organizations. The OPS pipeline begins with cleaning an intentionally high-recall, LLM-generated funder list by identifying true financial funders, standardizing grant-size ranges, engineering geographic-relevance features, and constructing cleaned mission statements suitable for embedding. We use transformer-based embeddings, UMAP dimensionality reduction, and HDBSCAN clustering[4, 1, 6, 5] to uncover latent "success-like" neighborhoods guided by a small proxy-success funder set. We then develop two heuristic scoring functions and evaluate sensitivity to embedding models and similarity metrics. In addition, we generate three LLM-based rankings as standalone qualitative prioritization tools that operate directly on narrative funder descriptions. Finally, we incorporate limited real outreach outcomes to calibrate $p(x)$ via a constrained supervised model, treating supervised learning as a refinement layer rather than a replacement for OPS. Overall, OPS yields actionable prioritization under extreme label scarcity and provides a scalable foundation that improves as real-world outreach data accumulates.

1 Objective

IterLight must prioritize outreach among many potential funders under extreme label scarcity. Only a small set of organizations can be identified as EdTech-aligned proxy successes, and reliable negative labels are unavailable. Rather than aiming for high-accuracy outcome prediction, we develop an interpretable decision-support framework that produces a stable and defensible outreach ranking.

We define Outreach Priority Scoring (OPS) as an expected-value-style ranking:

$$\mathrm{OPS}(x) = p(x) \times g(\mu(x)),$$

where $p(x)$ is a probability-like engagement score for organization x, $\mu(x)$ is expected grant capacity derived from typical minimum and maximum grant sizes, and $g(\cdot)$ is a monotone capacity transform. OPS is designed to guide action (who to contact first), not to guarantee funding. In this report we evaluate two choices of $g(\cdot)$: a raw-capacity form $g(\mu) = \mu$ and a log-scaled form $g(\mu) = \log(1 + \mu)$ to reduce dominance from extreme grant estimates.

2 Data Source and Preprocessing

2.1 Dataset Origin

The initial dataset was produced by a structured LLM prompt requesting a comprehensive, high-recall list of organizations that could plausibly fund IterLight. Because this step intentionally favors recall, it includes many non-funders and noisy metadata. Each entry includes fields such as category, geographic scope, mission/pillars, and typical grant-size descriptions.

2.2 Feature Engineering

Filtering to True Financial Funders: OPS focuses on expected *financial* outcomes. A key preprocessing step is identifying organizations that genuinely disburse monetary grants or sponsorship funding. Each entry was manually reviewed using publicly available sources (e.g., program descriptions, grant histories, CSR reports, and when applicable IRS Form 990 filings). We create a binary indicator `Financial Funder` and retain only organizations with `Financial Funder = 1`. This removes false positives introduced by high-recall generation and yields the funder set used for all subsequent modeling.

Standardizing Grant Size and Defining $\mu(x)$: Raw "typical grant size" fields were inconsistent (vague descriptors, missing values, contradictions). We standardized these into typical minimum and maximum grant amounts when available and define expected grant capacity as:

$$\mu(x) = \frac{\mathrm{Min}(x) + \mathrm{Max}(x)}{2}.$$

This provides a stable numeric quantity for the expected-value component of OPS.

Engineering Geographic Focus: Geographic relevance matters operationally for IterLight. We derive a binary indicator `geo_focus` that equals 1 if an organization operates in or regularly supports programs in IterLight-relevant regions (NY, NJ, CT, PA, MA, VT, Washington D.C., or Ontario), and 0 otherwise. We translate this into a multiplicative weight inside $p(x)$ (Section 3.2).

Cleaning Mission Statements for Embedding: Mission statements are the primary semantic signal used for embeddings and clustering. We cleaned mission text by removing: (i) monetary references (to reduce leakage into similarity), (ii) geographic mentions (to avoid confounding with `geo_focus`), and (iii) extraneous punctuation/formatting. Missing missions were manually filled when available through public sources. The resulting `mission_clean` field is embedded for representation-learning steps.

2.3 Modeling Fields Retained

The quantitative OPS core uses: `Organization Name`, `mission_clean`, `geo_focus`, `Typical Min Grant`, `Typical Max Grant`, and `Financial Funder`. Additional fields (category and strategic notes) are retained for interpretation and the LLM pipeline.

3 OPS Method: Representation Learning and $p(x)$ Construction

3.1 Pipeline Overview

OPS estimates engagement likelihood under small-data constraints by combining mission-text semantic alignment, unsupervised structure discovery, and proxy-success anchoring. Concretely, we embed each organization's `mission_clean` using a transformer-based text embedding model (e.g., Sentence-BERT or modern general-purpose embeddings) [8, 4, 1], then compute semantic alignment between each organization and IterLight via a similarity metric in embedding space. To expose neighborhood structure beyond raw pairwise similarities, we apply UMAP to obtain a lower-dimensional manifold representation [6], and cluster organizations with HDBSCAN to recover latent thematic groups without committing to a fixed number of clusters [5]. We then use a proxy-success set of known EdTech-aligned funders to identify a "success-like" region of the embedding space and translate this signal into cluster-level priors. Finally, we combine alignment, cluster prior, cluster confidence, and geography into a calibrated likelihood score $p(x)$ (see calibration discussion [3, 7]), and multiply by the expected grant capacity $\mu(x)$ to obtain the outreach priority score (OPS).

3.2 Defining $p(x)$

We define $p(x)$ as a probability-like engagement score formed by multiplying four interpretable components: (i) mission alignment $\mathrm{align}(x)$, the semantic similarity between organization x and IterLight in embedding space; (ii) a cluster-level prior $\mathrm{prior}(\mathrm{cluster}(x))$ that upweights clusters closer to proxy-success funders; (iii) a soft-fit confidence term $\mathrm{conf}(x)$ that downweights ambiguous or noisy cluster membership; and (iv) a geographic weight $w_{\mathrm{geo}}(x)$ derived from `geo_focus`.

$$p(x) \;\propto\; \mathrm{align}(x) \,\cdot\, \mathrm{prior}(\mathrm{cluster}(x)) \,\cdot\, \mathrm{conf}(x) \,\cdot\, w_{\mathrm{geo}}(x).$$

We treat $p(x)$ as a calibrated ranking signal rather than a literal probability; it is designed to be refined as outreach outcomes accumulate.

3.3 Experimental Structure

We produce 12 rankings to evaluate sensitivity and provide complementary perspectives: eight heuristic OPS variants (two heuristics $\times$ four embedding/similarity settings), three LLM-based rankings (three prompting methods), and one supervised calibration ranking (logistic regression trained on outreach response labels). The goal is not to claim a single "true" ranking, but to identify stable high-priority organizations and understand how design choices affect prioritization.

While OPS is evaluated across multiple heuristic and representation variants for sensitivity analysis, one specific configuration, 'H1-A' was selected to drive real outreach decisions. This baseline OPS used a bi-encoder transformer (BGE-Large) with cosine similarity for mission alignment and the raw expected grant capacity $\mu(x)$. This configuration was chosen for its interpretability, stability under clustering, and operational simplicity. All outreach outcomes collected during the project therefore correspond to this baseline OPS ordering, and subsequent supervised calibration is anchored to this same configuration. Variant definitions are summarized in Table 5.

4 Rankings Constructed

4.1 Heuristic OPS Rankings

We implement two expected-value formulations. Heuristic 1 uses raw grant capacity:

$$f_1(x) = p(x) \cdot \mu(x).$$

Heuristic 2 reduces dominance from extreme $\mu(x)$ values by log-scaling:

$$f_2(x) = p(x) \cdot \log(1 + \mu(x)).$$

Within each heuristic, we evaluate four representation variants (transformer model $\times$ similarity metric) to test sensitivity to semantic representation.

4.2 LLM-Based Rankings

Quantitative similarity and clustering are intentionally structured and interpretable, but some qualitative signals are difficult to encode (e.g., philanthropic framing, constraints on partnership models, and narrative emphasis). We therefore generate three LLM-based rankings as *standalone qualitative lenses* that operate directly on standardized narrative funder descriptions (Appendix A.1).

Prompt 1 scores **direct narrative alignment** to IterLight's mission statement and functions as a broad high-recall screen. Prompt 2 performs **proxy-success pattern matching** using a small set of known positive examples and is the only LLM ranking that we optionally compare to the baseline OPS top tier, because it uses the same pre-outreach information regime (Appendix A.2). Prompt 3 conditions on observed outreach outcomes and is therefore treated strictly as a **post-hoc sensitivity probe** illustrating how outcome feedback can shift narrative prioritization; it is not used to assess OPS or claim predictive accuracy.

4.3 Supervised Calibration of H1-A

We collected a limited set of real outreach outcomes (Section 5.4) and fit a constrained supervised model to calibrate $p(x)$ while preserving the structure and assumptions of the OPS framework. Supervised calibration is anchored to the baseline OPS configuration that was used to drive the initial outreach campaign, H1-A. Throughout the paper, H1A and H1B use cosine similarity, while H1C and H1D use cross-encoder similarity (Table 5).

This baseline OPS employed a bi-encoder transformer (BGE-Large) with cosine similarity for mission alignment and the untransformed expected grant capacity $\mu(x)$. This configuration was selected for three practical reasons. First, cosine similarity in a bi-encoder setting yields stable, smooth similarity scores that behave well under clustering and ranking, whereas cross-encoder similarity, while potentially more expressive, introduces pair-specific interactions that are harder to interpret globally and less practical at scale. Second, the raw expected grant capacity avoids introducing additional nonlinear transformations during outreach, ensuring that ranking decisions reflect a transparent tradeoff between engagement likelihood and funding magnitude. Third, among the embedding models evaluated, BGE-Large produced the most interpretable semantic neighborhoods: proxy-success funders formed coherent clusters, and nearest-neighbor relationships aligned closely with domain intuition.

The supervised model is treated as a refinement layer, not a replacement for OPS. We fit a regularized logistic regression using the same engineered OPS features (mission alignment, cluster prior/proximity, soft-fit confidence, geographic focus, and grant capacity) to estimate a calibrated probability of response $\hat{p}(x)$[3, 7]. Here, $\hat{p}(x)$ is a calibrated estimate of short-window response probability under our outreach protocol, whereas the heuristic $p(x)$ is a probability-like ranking signal constructed from interpretable components (Section 3.2). The resulting supervised expected-value ranking is:

$$f_{\text{sup}}(x) = \hat{p}(x) \cdot \mu(x).$$

Importantly, this model estimates the probability of receiving a short-window outreach response under our specific contact protocol, not the probability of securing funding. By preserving the OPS feature structure and learning only how these components map to observed engagement, supervised calibration incorporates real-world feedback without overfitting or undermining interpretability.

5 Evaluation & Results

We report results along five practical axes. Rather than assessing predictive accuracy, this section characterizes ranking behavior, design sensitivity, and decision updates under small-data constraints, focusing on stability, sensitivity, and directional consistency with limited behavioral feedback. Full tables and plots are provided in the Appendix.

5.1 Ranking Stability Across Heuristic Variants

We assess robustness by comparing overlap in top-k organizations across the eight heuristic OPS variants. This analysis examines how changes in embedding model, similarity function, and scoring formulation affect ranking consistency, with particular attention to stability among the highest-ranked organizations. Across variants, OPS exhibits strong stability at the top of the ranking. Averaged over all pairwise comparisons, the top-10 lists share 8.6 organizations on average (Jaccard similarity 0.76), with even the least similar variant pair sharing at least 7 of the top 10 organizations. Similar patterns hold at $k = 20$, indicating that OPS consistently identifies a robust high-priority tier despite substantial changes in representation and scoring design. Full overlap and Jaccard statistics for multiple values of k are reported in Appendix A.3.1. As k increases, agreement declines through the mid-ranked region, reflecting increased sensitivity where candidate quality is less clearly separated and small modeling differences induce reordering. From an applied perspective, these results indicate that OPS is reliable for identifying a core set of high-priority outreach candidates, while fine-grained ordering beyond the top tier should be interpreted cautiously. To connect stability outcomes to modeling choices, we additionally analyze a stratified outreach shortlist (15 top, 10 mid, 5 bottom) under all variants (Appendix A.3.2). Two patterns are most relevant for applied use. First, cross-encoder variants induce greater positional reordering among marginal candidates (higher rank dispersion), but are less sensitive to the underlying embedding model: within each heuristic, the cross-encoder shortlists are identical across BGE and GTE (Jaccard = 1.0), cross-encoder scoring dominated shortlist composition more than embedding backbone changes. Second, changing the scoring formulation (Heuristic 1 vs Heuristic 2) alters candidate inclusion more than switching embedding backbones within the same formulation, suggesting that modeling effort should prioritize the choice of capacity transformation and scoring structure over embedding selection when operating under small-data constraints.

5.2 Effect of Log-Scaling Grant Capacity

We isolate the effect of grant-capacity transformation by comparing heuristic variants that differ only in whether expected grant size is used in raw or log-scaled form. This analysis examines how scaling choices influence relative prioritization while holding all other modeling components fixed. Log-scaling does not materially affect the stability of the highest-ranked organizations. As shown by the top-k overlap analysis in Section 5.1, the core top tier remains largely invariant to modeling choices, including whether grant capacity is transformed. Differences instead arise in the mid-ranked region, where grant size plays a larger role in resolving tradeoffs among organizations with similar alignment scores. Shortlist comparisons indicate that changing the capacity transformation (Heuristic 1 vs. Heuristic 2) alters candidate inclusion more than switching embedding backbones within a fixed heuristic (Appendix A.3.3). This suggests that grant-capacity scaling is a meaningful design choice for marginal prioritization, particularly under noisy or uncertain grant-size estimates. From an applied perspective, log-scaling may provide a more conservative and robust prioritization strategy for early-stage outreach under small-data constraints.

5.3 LLM Ranking Behavior

The three LLM prompts are qualitative narrative heuristics with increasing access to information. Prompt 1 reflects broad thematic alignment and exhibits high recall with limited discrimination; **we use it primarily to illustrate that narrative fit is widespread in the candidate pool and therefore insufficient on its own for precise prioritization**. Prompt 2 incorporates proxy-success pattern matching and produces sharper separation by recognizing structural similarities to known EdTech-oriented funders; **we use it as a complementary qualitative signal and (optionally) as a consistency check against the baseline OPS top tier, since both operate without outreach labels**. Prompt 3 conditions on observed outreach outcomes and is therefore treated strictly as a post-hoc sensitivity probe illustrating how outcome feedback can shift narrative prioritization; **we use it to motivate the value of feedback-driven refinement, not as an evaluator**. We do not treat any LLM output as a benchmark or evaluator of OPS; instead, LLM rankings are used to aid qualitative interpretation and to surface sources of uncertainty not captured by embedding-based similarity. Full prompt text, scoring outputs, and an optional diagnostic comparison between Prompt 2 and the baseline OPS top-30 are provided in Appendix A.2.

5.4 Outreach Outcomes as Post-Hoc Diagnostic

One heuristic configuration (H1-A) was selected *a priori* as the operational baseline for real-world outreach. Using this fixed OPS ordering, we contacted 30 organizations, stratified across the top (15), middle (10), and bottom (5) regions of the ranking. For each contacted organization, we recorded whether any response occurred within a two-week window. These engagement outcomes are treated as short-horizon behavioral signals and are used solely as a post-hoc diagnostic of ranking alignment rather than as a measure of predictive accuracy. Among organizations for which a response label was observed, engagement exhibits a coarse monotonic relationship with ranking tier. Responses occur in both the top and middle tiers of the baseline OPS ordering, while no responses are observed among bottom-tier organizations. Although the absolute number of responses is small and the outreach sample was not randomly drawn, this pattern is directionally consistent with OPS assigning higher priority to organizations that are more likely to engage. Importantly, this analysis is not intended to evaluate model performance or estimate response probabilities. Because outreach targets were selected using a stratified design and engagement outcomes are sparse, the results should be interpreted as a qualitative sanity check rather than a statistical test. Nevertheless, the absence of responses in the lowest-ranked tier suggests that OPS does not systematically mis-rank low-engagement organizations above higher-priority candidates. Full tier-level counts and labeled organization lists are reported in Appendix A.3.5.

5.5 Calibrated Next-Outreach List (Top-30 Uncontacted)

The heuristic OPS score is intentionally constructed for *small-data* settings, where outcome labels are scarce and unreliable negatives are largely unavailable. After completing the first outreach round (30 organizations), we use the resulting response indicator (Label: responder vs. non-responder) to perform a lightweight supervised *calibration* step. This calibration step can be viewed as the quantitative analogue of Prompt 3: both incorporate outcome feedback, but the supervised model does so transparently using OPS features rather than narrative reasoning. The goal is not to replace OPS with a fully supervised model, but to estimate a calibrated engagement probability $\hat{p}(x)$ that preserves the structure of Heuristic 1. To keep the calibrated stage consistent with the baseline outreach configuration, the supervised model uses the same representation and similarity pipeline as Heuristic 1: BGE-Large bi-encoder embeddings and cosine similarity, cluster proximity weights, and soft cluster fit. Importantly, the supervised model does *not* apply any log-scaling or nonlinear transformation to grant capacity. Grant magnitude is treated as a separate, downstream quantity used only when converting calibrated engagement probability into an expected-value prioritization. We fit a constrained logistic regression with class-balancing on the contacted subset, using the following engineered inputs: (i) similarity to IterLight (`sim_to_iter`), (ii) cluster weight (`cluster_weight`), (iii) soft-fit confidence (`soft_fit`), and (iv) geographic focus (`Geo_Focus`). This produces a calibrated probability $\hat{p}(x)$ for each organization in the candidate set.

Calibrated expected value ranking. To generate a practical "who to contact next" list, we combine the calibrated probability with the *raw* expected grant capacity $\mu(x)$ (the row-wise mean of typical min/max grant size):

$$\mathrm{EV}_{\text{cal}}(x) \;=\; \hat{p}(x) \cdot \mu(x).$$

We remove already-contacted organizations and rank the remaining candidates by $\mathrm{EV}_{\text{cal}}(x)$. The resulting top-30 uncontacted organizations are listed in Appendix A.4. table 6, along with baseline OPS components for context (similarity, cluster weight, soft-fit, geography, and $\mu(x)$).

6 Limitations

This work is limited by noisy inputs and scarce labels. The funder list was LLM-generated and manually verified, so residual errors may remain, and OPS encodes explicit heuristic assumptions (cluster priors, geography) that are interpretable but not guaranteed optimal. Representation learning and clustering are sensitive to model and hyperparameters, and supervised calibration uses a small, operational label (two-week response) where non-response may reflect process friction rather than poor fit.

7 Conclusion

IterLight faces an outreach prioritization problem under extreme label scarcity, motivating the development of Outreach Priority Scoring (OPS) which combines a probability-like engagement signal $p(x)$ with expected grant capacity. OPS is not designed to predict funding outcomes, but to produce a stable and defensible ordering for action. Across eight heuristic variants OPS consistently identifies a small, robust top-priority tier. This invariance under substantial modeling changes provides the primary basis for trust: while mid-ranked ordering is sensitive and should be treated as an exploration region, the highest-priority candidates remain largely unchanged. Analyzing grant-capacity scaling shows that $\log(1 + \mu(x))$ preserves the top tier while reducing domination by extreme or noisy estimates, primarily affecting marginal prioritization. This sensitivity analysis informs robustness rather than redefining the operational pipeline, which remains anchored to the raw-capacity formulation for interpretability and consistency with outreach decisions. LLM-based rankings serve as complementary qualitative lenses rather than evaluators of OPS. Prompt 1 highlights the high recall and low discrimination of narrative alignment alone; Prompt 2 provides proxy-success pattern matching that can surface uncertainty when it diverges from OPS; and Prompt 3 is strictly post-hoc, illustrating how feedback can shift narrative priorities without serving as validation. Limited outreach outcomes provide a behavioral sanity check consistent with OPS tiering and enable a lightweight supervised calibration that refines $p(x)$ into $\hat{p}(x)$. Overall, OPS is trustworthy not because it predicts outcomes, but because it yields a stable, interpretable prioritization under uncertainty and improves transparently as real-world feedback accumulates.

Code Availability. All code used for data preprocessing, OPS construction, sensitivity analysis, LLM-based ranking, and supervised calibration is available in a public GitHub repository: `https://github.com/BecksterScience/IterLight`.

References

[1] Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*, 2024.

[2] Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2008.

[3] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

[4] Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*, 2023.

[5] Leland McInnes, John Healy, and Steve Astels. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205, 2017.

[6] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[7] Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 2005.

[8] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*, 2019.

A Appendix

A.1 LLM Prompts (Full)

A.1.1 Prompt to combine funder information

```
Prompt = """
You are given structured information about a funding organisation.
Combine all fields into one coherent, well-written text block that captures the
organisation's mission, geographic scope, funding behaviour, grant size information,
priorities, and any other relevant details.
Do NOT add any new information. Do NOT omit any field unless it is empty.
Write in clear prose, not bullet points. This combined text will be used for
downstream semantic analysis and scoring.

Name: {NAME}
Mission Statement: {MISSION_STATEMENT}
Category: {CATEGORY}
Geographic Scope: {GEOGRAPHIC_SCOPE}
Geo Focus: {GEO_FOCUS}
Typical Minimum Grant Size: {MIN_GRANT}
Typical Maximum Grant Size: {MAX_GRANT}

Task:
Create a single cohesive paragraph merging all information.
Output ONLY the final paragraph.
"""
```

A.1.2 Prompt for Method 1

```
prompt_1 = """
You are evaluating whether a funding organisation is likely to support a company
called IterLight.

IterLight is an education technology company advancing how K-12 students learn
by connecting academic skills with the passions that drive them--starting with
    sports.
Its mission is to make learning more engaging, equitable, and evidence-based through
innovative, data-driven approaches that empower every learner to reach their
    potential.

You will be given one organisation's combined description. Using only the
    information
about IterLight and this organisation, assess how likely it is that the organisation
would fund or sponsor IterLight's work.

Consider factors such as:
- alignment with education, youth, equity, innovation, technology, or sports
- history of funding education or youth-focused programs
- interest in community impact or student outcomes
- whether their mission and priorities overlap with IterLight's goals
- whether they typically give grants or sponsorships of any kind
- their region of funding

Task:
1. Provide a funding-likelihood score between 0 and 1
(0 = no alignment, 1 = extremely strong alignment)
2. Provide a short explanation (2-4 sentences) describing the main reasons
for the score.
"""
```

A.1.3 Prompt for Method 2

```
prompt_2 = """
You are evaluating how likely a funding organisation is to support a company
called IterLight.

IterLight is an education technology company advancing how K-12 students learn
by connecting academic skills with the passions that drive them--starting with
    sports.
Its mission is to make learning more engaging, equitable, and evidence-based through
innovative, data-driven approaches that empower every learner to reach their
    potential.

Below is a set of "good funders"--organisations that have historically funded work
aligned with IterLight's mission. These represent positive examples of funders who
support youth development, education innovation, equity, or community-based learning
initiatives. They provide a pattern for what successful funders look like.

Here are the known good funders:
{GOOD_FUNDERS_TEXT}

Your task is to evaluate how likely a *new* organisation is to fund or sponsor
IterLight, using BOTH:
1. IterLight's mission and work, and
2. The characteristics shared by the good funders listed above.

When scoring, consider:
- alignment with K-12 education, youth programs, equity, sports, learning engagement
- interest in innovation, technology, data-driven education
- grant-making patterns reflected in the good funders
- signs of philanthropic or community-focused priorities
```

```
- similarity in mission, geographic focus, or funding behavior

Task:
1. Provide a funding likelihood score between 0 and 1.
2. Provide a brief explanation (2-6 sentences) describing the main reasons
for the score, explicitly referencing overlap with the good funders when relevant.

Output your answer strictly in the following JSON format:

{
"score": <numeric_score_between_0_and_1>,
"reasoning": "<short explanation>"
}

Here is the organisation to evaluate:
{ORG_TEXT}
"""
```

A.1.4 Prompt for Method 3

```
prompt_3 = """
You are evaluating how likely a funding organisation is to respond positively or
at least engage with outreach from a company called IterLight.

IterLight is an education technology company advancing how K-12 students learn
by connecting academic skills with the passions that drive them--starting with
    sports.
Its mission is to make learning more engaging, equitable, and evidence-based through
innovative, data-driven approaches that empower every learner to reach their
    potential.

Below is real outreach outcome data:
- Organisations labeled "1" responded to IterLight's outreach within two weeks.
- Organisations labeled "0" did not respond.

This dataset reflects actual behavioural patterns that indicate which types of
organisations are more likely to engage with IterLight.

Here are the organisations that responded (label = 1):
{RESPONDER_TEXT}

Here are the organisations that did not respond (label = 0):
{NONRESPONDER_TEXT}

Your task is to evaluate how likely a *new* organisation will respond or engage
with IterLight, based on:
1. IterLight's mission,
2. The characteristics of past responders, and
3. The contrasting characteristics of past non-responders.

When scoring, consider factors such as:
- thematic alignment with respondents (education, youth, learning, equity, sports,
    innovation)
- funding behaviour, grant-making patterns, mission alignment
- organisational scale, scope, and philanthropic focus
- similarities to organisations that previously engaged with IterLight
- dissimilarities to those that consistently did not respond

Task:
1. Provide a likelihood score between 0 and 1 indicating how likely this new
organisation is to respond to outreach.
2. Provide a brief explanation (2-6 sentences) describing the main reasons,
explicitly referencing alignment with responders or contrast with non-responders.
```

```
Output your answer strictly in the following JSON format:
{
"score": <numeric_score_between_0_and_1>,
"reasoning": "<short explanation>"
}
"""
```

A.2 LLM Diagnostics (Optional): Prompt–OPS Comparison and Interpretation

This appendix subsection documents how LLM-based rankings relate to OPS without treating LLM output as a validation benchmark. Because Prompt 3 conditions directly on observed outreach outcomes, it is strictly post-hoc and is not compared against OPS ordering to avoid label leakage and hindsight bias. Accordingly, the only LLM–OPS comparison reported here uses Prompt 2 (proxy-success pattern matching), which relies on the same type of information available prior to outreach (organization narrative descriptions and a proxy-success example set).

Comparison setup. We restrict attention to the original baseline OPS configuration (H1-A) that was used to select outreach targets. We take the baseline top-30 organizations under H1-A (prior to any outreach outcomes), and examine how Prompt 2 scores distribute across these organizations. This comparison is intended as a qualitative consistency check: it asks whether narrative proxy-success reasoning broadly agrees with the structured OPS prioritization in the subset that OPS itself deemed highest priority.

How to interpret agreement and disagreement. Partial convergence between Prompt 2 and baseline OPS is expected because both encode proxy-success information, but their inductive biases differ. OPS is driven by embedding-space similarity, cluster priors, and grant-capacity weighting, while Prompt 2 emphasizes narrative cues such as philanthropic framing, stated mechanisms of support, and perceived similarity to the proxy-success examples. Therefore, disagreements are treated as informative uncertainty signals rather than errors: cases where OPS ranks an organization highly but Prompt 2 scores it moderately may indicate that the organization looks structurally promising in embedding space but has narrative or operational constraints not captured in OPS features.

Where the underlying data live. Prompt 2 scores and short rationales for the organizations used in this diagnostic are reported in Appendix A.5. Baseline OPS ranks and the outreach shortlist construction are reported in Appendix A.3.2.

A.3 Results Tables and Plots

This appendix aggregates supporting tables and diagnostics referenced in Section 5. It is intended to provide full transparency and reproducibility without overloading the main text.

A.3.1 Top-k Stability Across Heuristic Variants

For each pair of the eight heuristic OPS variants (28 pairwise comparisons), we compute (i) the overlap in the sets of top-k organizations and (ii) the Jaccard similarity between those sets, for multiple values of k. These statistics summarize how sensitive ranking outcomes are to changes in embedding model, similarity function, and scoring formulation.

Table 1 reports the mean, standard deviation, minimum, and maximum overlap and Jaccard similarity across all pairwise comparisons. Stability is highest at small values of k, indicating strong agreement among the highest-ranked organizations. Even at $k = 10$, the minimum observed overlap is 7 organizations, demonstrating that the core priority tier is largely invariant to modeling choices. Agreement declines as k increases into the mid-ranked region, where variants exhibit greater sensitivity and reordering.

At very large values of k, overlap increases again as the cutoff approaches the size of the shared candidate universe. This effect reflects coverage rather than meaningful ranking agreement and should not be interpreted as increased robustness.

Table 1: Top-k overlap and Jaccard similarity across OPS heuristic variants. Statistics are aggregated over all 28 pairwise variant comparisons.

	Overlap				Jaccard			
k	Mean	Std	Min	Max	Mean	Std	Min	Max
10	8.61	0.79	7	10	0.76	0.13	0.54	1.00
20	16.25	1.86	14	20	0.70	0.14	0.54	1.00
50	34.14	9.17	23	50	0.55	0.23	0.30	1.00
100	85.57	7.66	72	100	0.76	0.12	0.56	1.00

A.3.2 Shortlist Sensitivity Diagnostic (Rank Dispersion and Set Overlap)

Because the outreach shortlist intentionally sampled organizations from the top, middle, and bottom of the baseline ranking, we additionally examine how shortlisted organizations behave under alternative heuristic variants. For each heuristic variant's 30-organization shortlist, we evaluate the rank position of each shortlisted organization under the full ordering produced by every other variant, and summarize variability via mean rank, standard deviation of rank, and range (min/max). We also report Jaccard overlap between the sets of 30 shortlisted organizations across variants to assess whether design changes alter candidate inclusion.

Interpretation and implications for OPS design. This diagnostic is intended to translate ranking sensitivity into actionable guidance for IterLight under small-data constraints. Because only a limited number of organizations can be confidently labeled as positive funders, the goal of OPS evaluation is not predictive accuracy but robustness of prioritization: whether OPS identifies a stable high-priority tier and how modeling choices affect ordering elsewhere in the ranking.

The top-k stability results reported in Appendix A.3.1 establish that the highest-ranked organizations are largely invariant to modeling choices. In contrast, the shortlist analysis deliberately probes sensitivity beyond the top tier by examining a stratified sample of organizations drawn from the top, middle, and bottom of the baseline ranking. Consequently, average shortlist ranks near the middle of the full ordering and large maximum ranks are expected and should not be interpreted as instability of the core priority tier.

Rank dispersion statistics (Table 3) reveal systematic differences across similarity methods. Shortlists generated using cross-encoder similarity exhibit higher rank variability across heuristic variants than cosine-based approaches, indicating that cross-encoders induce sharper distinctions and greater reordering among marginal candidates. At the same time, cross-encoder shortlists are identical across embedding backbones (BGE vs. GTE), as reflected by unit Jaccard similarity in Table 4. This suggests that once a cross-encoder is used, shortlist composition is dominated by the similarity model itself rather than by the underlying embedding representation.

In contrast, changing the scoring formulation (Heuristic 1 versus Heuristic 2) produces larger changes in shortlist composition than switching embedding models within a fixed heuristic. Across many heuristic pairs, Jaccard overlap between shortlists is moderate, indicating genuine uncertainty about candidate inclusion outside the stable top tier. This pattern implies that modeling effort should prioritize the choice of capacity transformation and scoring structure over embedding selection when operating in low-label regimes.

From an applied perspective, these results motivate an exploit–explore outreach strategy. IterLight can confidently prioritize organizations that consistently appear in the top tier across variants, while treating mid-ranked and marginal shortlist candidates as an exploration pool. Rather than over-interpreting fine-grained rank differences in this region, outreach can be diversified across variants and refined iteratively using engagement feedback, as discussed in Section 5.4.

A.3.3 Effect of Grant-Capacity Scaling on Marginal Prioritization

This subsection provides additional interpretation of the impact of grant-capacity transformation on OPS behavior. While the top-k stability analysis demonstrates that high-priority organizations are robust to modeling choices, shortlist-level diagnostics reveal how capacity scaling affects prioritization among marginal candidates.

Table 2: Shortlist sizes for each heuristic variant (all shortlists contain 30 organizations).

Variant	Shortlist size
H1_bge_cos	30
H1_bge_cross	30
H1_gte_cos	30
H1_gte_cross	30
H2_bge_cos	30
H2_bge_cross	30
H2_gte_cos	30
H2_gte_cross	30

Table 3: Shortlist-level rank dispersion across variants. Each row summarizes, for a given variant's 30-organization shortlist, the distribution of each shortlisted organization's rank statistics measured across the eight full rankings.

	Mean rank			Std rank			Missing ranks	
Shortlist	Mean	Median	Max	Mean	Median	Max	Mean	Max
H1_bge_cos	50.85	42.62	144.43	11.73	7.72	46.94	0.03	1
H1_bge_cross	48.26	37.44	144.12	15.13	7.33	57.01	0.03	1
H1_gte_cos	53.85	44.50	144.12	12.96	8.72	46.94	0.00	0
H1_gte_cross	48.26	37.44	144.12	15.13	7.33	57.01	0.03	1
H2_bge_cos	50.64	33.87	144.12	10.79	5.29	49.70	0.00	0
H2_bge_cross	47.39	34.87	134.50	15.43	13.60	57.01	0.03	1
H2_gte_cos	55.80	37.25	144.43	9.11	5.13	46.12	0.03	1
H2_gte_cross	52.01	47.75	139.88	16.88	12.98	51.38	0.00	0

Using raw expected grant size amplifies differences among organizations with large estimated funding capacity, increasing their influence on the overall ranking. In regions of the ranking where alignment scores are similar, this can cause a small number of organizations with extreme or uncertain grant estimates to dominate prioritization. This effect is reflected in lower shortlist overlap between Heuristic 1 (raw capacity) and Heuristic 2 (log-scaled capacity), relative to variants that differ only in embedding model.

Log-scaling compresses extreme grant-size values, reducing the leverage of large but uncertain estimates and allowing alignment-driven factors to play a larger role in marginal ranking decisions. As a result, organizations with strong mission fit but moderate funding capacity are more likely to surface in the mid-ranked region under log-scaled variants. This behavior is consistent with the observed shortlist Jaccard patterns and rank dispersion statistics reported in Appendix A.3.2.

From an operational standpoint, this distinction reflects a tradeoff between upside-seeking and robustness. Raw expected value prioritization emphasizes potential maximum returns but is more sensitive to noise in grant-size estimates, whereas log-scaled capacity yields more stable and risk-aware prioritization when grant information is sparse or heterogeneous. Given IterLight's small-data setting, log-scaling serves as a principled default for early outreach, with raw capacity becoming more appropriate as grant estimates are refined through engagement feedback.

A.3.4 Additional Tables and Figures

A.3.5 Outreach Outcome Diagnostic Details

Tables in analysis reports labeled outreach outcomes by ranking tier under the baseline OPS configuration. Organizations with missing response labels are excluded. Because outreach targets were intentionally sampled across ranking regions, observed response rates should not be interpreted as estimates of underlying engagement probability. Instead, the diagnostic assesses whether observed engagement contradicts or aligns with the OPS prioritization.

Table 4: Jaccard similarity between the sets of 30 shortlisted organizations across heuristic variants.

	H1_bge_cos	H1_bge_cross	H1_gte_cos	H1_gte_cross	H2_bge_cos	H2_bge_cross	H2_gte_cos	H2_gte_cross
H1_bge_cos	1.000	0.500	0.538	0.500	0.304	0.277	0.333	0.224
H1_bge_cross	0.500	1.000	0.364	1.000	0.333	0.304	0.304	0.304
H1_gte_cos	0.538	0.364	1.000	0.364	0.364	0.250	0.364	0.250
H1_gte_cross	0.500	1.000	0.364	1.000	0.333	0.304	0.304	0.304
H2_bge_cos	0.304	0.333	0.364	0.333	1.000	0.395	0.333	0.277
H2_bge_cross	0.277	0.304	0.250	0.304	0.395	1.000	0.333	0.333
H2_gte_cos	0.333	0.304	0.364	0.304	0.333	0.333	1.000	0.333
H2_gte_cross	0.224	0.304	0.250	0.304	0.277	0.333	0.333	1.000

Table 5: Heuristic variants evaluated (transformer/similarity).

Variant	Configuration
H1-A / H2-A	BGE-Large + cosine similarity
H1-B / H2-B	GTE-Large + cosine similarity
H1-C / H2-C	BGE-Large + cross-encoder similarity
H1-D / H2-D	GTE-Large + cross-encoder similarity

Responses are observed exclusively in the top and middle tiers, with no responses among bottom-tier organizations. Given the small sample size and heterogeneous outreach conditions, this result supports directional consistency between OPS ranking and observed engagement while highlighting the need for additional feedback to refine prioritization in future outreach rounds.

A.4 Calibrated Top-30 Uncontacted Organizations

Table 6 lists the top-30 *uncontacted* organizations under the calibrated expected value score $\mathrm{EV}_{\text{cal}}(x) = \hat{p}(x)\mu(x)$ (Section 5.5). We include $\hat{p}(x)$ (logistic calibration output) alongside the baseline OPS components used to construct it (similarity to IterLight, cluster weight, soft-fit confidence, and geographic focus) and the raw expected grant capacity $\mu(x)$.

The purpose of this appendix is operational: it provides a concrete, reproducible shortlist that can be used to drive the next outreach wave while keeping the ranking logic transparent and auditable.

A.5 LLM Ranking Results & Reasoning

These tables show the scores given by the LLM and the rationale used per organisation. The order of results includes of the top 15, middle 10 and lowest 5 scores. Finally, the model used to generate these results was Claude sonnet 4.

Table 7: LLM Prompt 1 Rankings: Direct Semantic Alignment with IterLight

Organization	Score	Prompt 1 Rationale
IES SBIR (ED/IES)	0.95	There is extremely strong alignment between IES SBIR and IterLight. The organization supports evidence-based educational research, directly matching IterLight's data-driven K–12 mission. As a federal program with grants in the \$250K–\$1M range, it is well-suited to fund innovative education technology.
Owl Ventures	0.90	Owl Ventures is a specialized EdTech venture capital firm with global reach and substantial funding capacity. Its portfolio includes K–12 platforms such as Kiddom, demonstrating direct interest in IterLight's target market and AI-driven learning approach.
Golden State Warriors Community Foundation	0.90	The foundation's focus on education, youth development, and STEM closely aligns with IterLight's mission. IterLight's sports-based learning model naturally complements the foundation's professional sports affiliation and equity goals.

Organization	Score	Prompt 1 Rationale
Houston Rockets – Clutch City Foundation	0.85	The foundation prioritizes Houston-area youth education and has an established grant program. Its NBA affiliation aligns strongly with IterLight's sports-integrated learning approach and K–12 focus.
San Antonio Spurs – Spurs Give	0.85	Spurs Give focuses on youth development and community impact, aligning well with IterLight's educational mission. The sports connection strengthens the match, and its grant range supports meaningful EdTech initiatives.
Philadelphia Union Foundation	0.85	This foundation supports youth education and community development through professional sports, directly matching IterLight's sports-based K–12 learning model, with minor geographic constraints.
Dallas Cowboys Foundation	0.85	The foundation's emphasis on youth education, school partnerships, and equity aligns strongly with IterLight's mission. Its geographic focus on North Texas could support regionally targeted expansion.
Portland Timbers Community Fund	0.85	The fund supports youth education initiatives and, as an MLS organization, aligns naturally with IterLight's sports-integrated learning platform.
Carolina Panthers Charities	0.85	The foundation's dual emphasis on education and athletics strongly matches IterLight's mission. Geographic limitations require local impact demonstration.
Real Salt Lake Foundation	0.85	The foundation supports youth development and education through a professional sports lens, making it a strong thematic match for IterLight's model.
Chicago Bulls Charities	0.85	Chicago Bulls Charities has a strong record of funding youth education and community programs. Its sports affiliation creates direct synergy with IterLight's learning-through-sports approach.
Toronto FC – MLSE Foundation	0.85	The foundation uses sports to support youth development, closely aligning with IterLight's mission. Funding capacity is substantial, with geographic considerations for Canada-based programs.
TGR Foundation (Tiger Woods)	0.85	TGR Foundation focuses on education technology, STEM, and equity, directly matching IterLight's data-driven learning objectives and sports connection.
Kansas City Chiefs Foundation	0.85	The foundation emphasizes youth education and athletics, aligning strongly with IterLight's mission, with some geographic constraints.
Vancouver Whitecaps FC Community	0.85	As a professional soccer organization supporting youth education and wellness, the foundation closely aligns with IterLight's sports-education integration model.
New York Yankees Foundation	0.85	The foundation focuses on youth education, equity, and social responsibility, aligning well with IterLight's mission, with regional geographic considerations.
Orlando Magic Youth Foundation (OMYF)	0.85	OMYF supports at-risk youth through education initiatives and sports engagement, closely matching IterLight's K–12 equity-driven learning approach.
DonorsChoose (Technology Classroom Projects)	0.85	DonorsChoose strongly aligns with IterLight's EdTech mission, though its smaller grant sizes may limit organizational-scale funding.
Chan Zuckerberg Initiative (CZI)	0.85	CZI's focus on educational technology and whole-child development aligns strongly with IterLight's mission, with substantial national funding capacity.

Organization	Score	Prompt 1 Rationale
Under Armour Foundation	0.85	The foundation's focus on youth engagement and sports aligns closely with IterLight's sports-based educational model and national reach.
Nike Community Impact Fund	0.85	Nike's focus on youth development and equity aligns well with IterLight's mission of connecting learning with sports passions.
Sacramento Republic FC	0.85	This organization integrates sports with youth education and community impact, matching IterLight's model, with regional scope considerations.
MIT Solve (Global Learning & Education Tracks)	0.85	MIT Solve's emphasis on global education innovation and standardized grants aligns strongly with IterLight's K–12 EdTech mission.
New Orleans Pelicans	0.80	The organization supports youth development and education through sports, aligning with IterLight's mission, with regional funding constraints.
Dallas Stars Foundation	0.80	The foundation supports youth education in a major metro area and aligns with IterLight's sports-based learning approach.
Minnesota Wild Foundation	0.30	While youth-focused, the foundation prioritizes medical and hockey development over educational technology, limiting alignment.
University of Michigan – Edward Ginsberg Center	0.30	The center funds community education projects but primarily supports internal university initiatives rather than external EdTech organizations.
University of Georgia – Office of Service-Learning	0.30	The office emphasizes service-learning rather than education technology innovation, limiting alignment with IterLight's platform.
Kickstarter	0.30	Kickstarter's project-based crowdfunding model and limited funding scale misalign with IterLight's long-term EdTech objectives.
Philadelphia Eagles Foundation	0.20	The foundation's specialized focus on autism research limits alignment with IterLight's broader education technology mission.

Table 8: LLM Prompt 2 Rankings: Pattern Matching Using Proxy-Success Funders

Organization	Score	Prompt 2 Rationale
MIT Solve (Global Learning & Education Tracks)	0.95	MIT Solve shows exceptional alignment with IterLight's mission and is explicitly listed as a proxy-success funder in the provided examples. Its focus on global learning and education innovation directly matches IterLight's K–12 EdTech objectives, while its social-impact innovation model aligns with IterLight's equity goals. The standardized $10,000 grant size fits well within education technology funding patterns.
The Ohio State University – Office of Outreach	0.85	The Office of Outreach aligns strongly with IterLight's mission through its emphasis on K–12 student success and educational outcomes. As an academic-affiliated funder, it mirrors proxy-success patterns seen in organizations such as MIT Solve and Berkeley SkyDeck, with grant sizes consistent with education innovation support.

Organization	Score	Prompt 2 Rationale
Brooklyn Nets Foundation	0.85	The foundation's focus on youth education, athletics, and community development closely matches IterLight's sports-based learning model. Its grant flexibility and emphasis on educational opportunity align well with proxy-success funder characteristics, with minor geographic constraints.
Los Angeles Lakers Youth Foundation	0.85	This foundation strongly aligns with IterLight's equity-driven education mission through its focus on underserved youth and integration of sports with learning. Its grant range mirrors those of several proxy-success education funders, with geographic scope as the primary limitation.
LA Clippers Foundation	0.85	The LA Clippers Foundation emphasizes educational advancement for children and community impact, closely reflecting proxy-success funder priorities. Its professional sports affiliation strengthens alignment with IterLight's sports-integrated learning approach.
Indiana Pacers Foundation	0.85	The foundation aligns strongly with IterLight's focus on equity, youth empowerment, and education. Its grant scale fits within the range observed among proxy-success funders, though geographic scope may limit broader deployment.
Austin FC – 4ATX Foundation	0.85	Austin FC's foundation shows strong alignment with proxy-success funders through its emphasis on youth development, education, and closing opportunity gaps. Its substantial funding capacity and sports affiliation create an excellent thematic match with IterLight.
Detroit Pistons Foundation	0.85	The foundation's focus on youth education, leadership, and equity mirrors the characteristics of proxy-success funders. Its NBA affiliation further strengthens alignment with IterLight's sports-based EdTech model.
Denver Nuggets Community Relations	0.85	Denver Nuggets Community Relations supports youth education initiatives and demonstrates strong alignment with proxy-success patterns seen in education-focused sports foundations. Grant sizes are consistent with education technology initiatives.
Cleveland Cavaliers Community Foundation	0.85	This foundation's focus on education, entrepreneurship, and youth development closely aligns with IterLight's mission and mirrors funding behavior observed among proxy-success funders.
Winnipeg Jets True North Youth Foundation	0.75	The foundation demonstrates strong thematic alignment through sports-integrated education and equity initiatives, similar to proxy-success funders, though its geographic focus and hands-on programming model may limit direct funding.
Atlanta Falcons Youth Foundation	0.75	The foundation aligns with IterLight's sports-based engagement model and youth focus, though its primary emphasis on physical fitness and local scope slightly reduces alignment compared to proxy-success funders.
Washington Wizards – Monumental Sports & Entertainment Foundation	0.75	This foundation reflects proxy-success characteristics through substantial community grants, professional sports affiliation, and support for local education initiatives.
Sacramento Kings Foundation	0.75	The foundation supports youth literacy and education initiatives, aligning with IterLight's K–12 mission, though limited geographic scope and innovation emphasis reduce alignment relative to proxy-success funders.

Organization	Score	Prompt 2 Rationale
Oklahoma City Thunder Foundation	0.75	The foundation emphasizes educational opportunities for underserved youth, aligning with IterLight's equity goals and proxy-success funder patterns, with regional scope constraints.
New York Knicks Garden of Dreams Foundation	0.75	This foundation aligns through youth education, sports affiliation, and meaningful grant sizes, closely resembling proxy-success funder behavior with some geographic limitations.
Milwaukee Bucks Foundation	0.75	The foundation supports youth education and community development, reflecting proxy-success characteristics, though its Wisconsin-only focus constrains broader applicability.
Miami HEAT Charitable Fund	0.75	The fund's education pillar and sports affiliation align with IterLight's mission, though smaller grant sizes and regional focus reduce alignment compared to proxy-success funders.
Houston Rockets – Clutch City Foundation	0.75	The foundation aligns with IterLight's sports-based K–12 education approach and matches proxy-success funding patterns, with geographic scope as the main constraint.
Charlotte Hornets Foundation	0.75	The foundation demonstrates strong alignment through education initiatives, youth development, and sports affiliation, consistent with proxy-success funder profiles.
Chicago Cubs Charities	0.75	This organization aligns with IterLight's focus on at-risk youth, education, and sports, closely matching proxy-success characteristics despite geographic limitations.
University of Southern California (USC) Good Neighbors	0.75	USC Good Neighbors mirrors proxy-success funders by supporting education initiatives through an academic institution, with grant sizes and mission alignment consistent with IterLight's goals.
Berkeley SkyDeck Fund (UC Berkeley)	0.75	As an identified proxy-success funder, Berkeley SkyDeck strongly aligns with IterLight's mission through its academic accelerator model and standardized $200,000 funding.
New York Yankees Foundation	0.75	The foundation's emphasis on education, youth development, and equity aligns well with proxy-success funder patterns, with regional scope as a limiting factor.
Wilson Sporting Goods	0.75	Wilson aligns with IterLight through its sports-centered youth empowerment mission, reflecting proxy-success characteristics at the intersection of athletics and education.
Atlanta Hawks Foundation	0.85	The Atlanta Hawks Foundation shows exceptional alignment through youth education, equity, and sports integration, closely matching proxy-success funder behavior.
Y Combinator	0.85	Y Combinator is explicitly identified as a proxy-success funder and aligns strongly with IterLight's technology-driven education innovation and scalable impact model.
Google for Education	0.85	Google for Education reflects proxy-success patterns through large-scale educational technology support and strong alignment with IterLight's data-driven learning platform.
Chicago Bulls Charities	0.85	Chicago Bulls Charities closely mirrors proxy-success funders through its focus on youth education, sports integration, and flexible grant capacity.

Table 6: Top-30 uncontacted organizations under calibrated expected value $\mathrm{EV}_{\mathrm{cal}}(x) = \hat{p}(x)\mu(x)$. Columns include calibrated probability $\hat{p}(x)$ and the baseline OPS components used for calibration, plus raw expected grant capacity $\mu(x)$.

Organization	$\hat{p}(x)$	sim_to_iter	cluster_w	soft_fit	Geo	$\mu(x)$	$\mathrm{EV}_{\mathrm{cal}}(x)$
U.S. Dept. of Education (EIR Program)	0.70	0.61	1.00	0.81	1	6,000,000	4,222,888.12
Eat. Learn. Play. Foundation (Stephen & Ayesha Curry)	0.47	0.63	0.25	0.71	0	300,000	140,392.97
Austin FC - 4ATX Foundation	0.11	0.56	0.25	0.85	0	1,050,000	114,009.38
University of Wisconsin-Madison - Community Relations	0.31	0.61	0.50	0.80	0	260,000	81,348.99
Charlotte Hornets Foundation	0.27	0.56	0.15	0.73	0	262,500	69,727.95
Kansas City Chiefs Foundation	0.18	0.68	0.15	0.84	0	375,100	68,240.54
Houston Texans Foundation	0.17	0.56	0.25	0.80	0	400,000	67,054.44
Washington Wizards - Monumental Sports & Entertainment Foundation	0.50	0.49	0.50	0.73	1	115,032.50	56,892.54
New England Patriots Foundation	0.37	0.60	0.25	0.80	1	130,000	47,737.00
New York Knicks Garden of Dreams Foundation	0.43	0.53	0.50	0.77	1	110,000	47,000.04
TGR Foundation (Tiger Woods)	0.87	0.61	1.00	0.71	1	52,500	45,466.26
New Orleans Pelicans	0.12	0.57	0.10	0.81	0	330,000	40,653.13
Washington Commanders Foundation	0.73	0.49	1.00	0.72	1	54,500	40,016.88
Penn State University - Outreach & Engagement	0.29	0.56	0.25	0.81	1	100,000	28,585.91
Brooklyn Nets Foundation	0.27	0.67	0.25	0.87	1	101,800	27,935.76
FC Dallas Foundation	0.12	0.63	0.25	0.88	0	221,415	25,994.63
Chicago Bears Charities	0.21	0.56	0.25	0.77	0	105,000	22,035.00
Boston Celtics Shamrock Foundation	0.41	0.54	0.50	0.79	1	53,000	21,464.62
Chicago Fire FC Foundation	0.08	0.63	0.10	0.88	0	255,000	21,364.38
Utah Jazz Foundation	0.15	0.56	0.25	0.81	0	136,000	20,473.03
San Jose Sharks Foundation	0.13	0.65	0.25	0.88	0	155,000	20,351.26
Calgary Flames Foundation	0.34	0.63	0.50	0.80	0	55,000	18,656.86
Philadelphia 76ers - Sixers Youth Foundation	0.28	0.63	0.25	0.79	0	65,000	17,911.41
Chicago Blackhawks Foundation	0.09	0.58	0.25	0.87	0	177,500	16,334.63
Florida State University Research Foundation	0.54	0.56	1.00	0.78	0	27,500	14,757.77
Philadelphia Eagles Foundation	0.52	0.54	0.10	0.68	1	27,500	14,172.65
LA Kings Care Foundation	0.09	0.58	0.50	0.92	0	149,000	13,871.23
Buffalo Sabres Foundation	0.25	0.62	0.50	0.90	1	55,000	13,471.84
Microsoft Philanthropies	0.51	0.63	0.25	0.75	1	26,000	13,364.36
Colorado Avalanche Community Fund	0.49	0.58	0.50	0.72	0	27,500	13,364.06

Table 9: LLM Prompt 3 Rankings: Responder–Non-Responder Pattern Inference

Organization	**Score**	**Prompt 3 Rationale**
U.S. Dept. of Education (EIR Program)	0.85	The EIR Program shows very strong alignment with organizations that previously responded to IterLight, particularly federal grant programs such as NSF and IES SBIR that prioritize evidence-based educational innovation. Its emphasis on rigorous evaluation, large grant sizes (\$2–10M), and high-need student populations closely matches IterLight's data-driven and equity-focused mission.

Organization	Score	Prompt 3 Rationale
Portland Trail Blazers Foundation	0.85	This foundation aligns strongly with responder patterns observed in similar Portland-based sports organizations, such as the Portland Timbers Community Fund. Its explicit prioritization of education through the *Learn* grants pillar differentiates it from non-responding NBA foundations and increases likelihood of engagement.
Penn State University — Outreach & Engagement	0.75	Penn State Outreach mirrors successful responders through its focus on K–12 STEM education and youth programs. Its collegiate funding structure and grant range align well with institutions such as MIT Solve and Berkeley SkyDeck that previously engaged with IterLight.
University of Tennessee	0.75	The University of Tennessee demonstrates strong alignment with responder institutions through its K–12 educational focus and collaborative approach between higher education and school systems. This contrasts with non-responders that lacked explicit education technology emphasis.
University of Georgia — Office of Service-Learning	0.75	This organization aligns with responders through its educational mission and community engagement focus. However, its service-learning orientation and limited geographic scope slightly reduce alignment relative to broader education-technology funders.
Texas A&M University — Public Partnership & Outreach	0.75	Texas A&M shows strong similarity to responding collegiate funders through its K–12 partnership focus and community engagement mission. The absence of explicit innovation or technology language limits its score relative to top-tier responders.
Nike Community Impact Fund	0.75	Nike's foundation aligns with responder characteristics through its youth-focused, sports-connected mission and substantial corporate funding capacity. Unlike narrowly focused sports foundations that did not respond, Nike's structure suggests greater flexibility and openness to innovation.
Wilson Sporting Goods	0.75	Wilson mirrors responder organizations through its national scope, corporate foundation structure, and emphasis on youth empowerment via sports. These attributes distinguish it from traditional pro-team foundations that showed low response rates.
Auburn University — Office of Public Service	0.75	Auburn's Office of Public Service aligns with responding collegiate institutions through its education-centered mission and youth program focus. While geographically limited, its structural similarity to other engaged academic funders suggests strong likelihood of response.
The Ohio State University — Office of Outreach	0.75	This organization aligns with responder patterns through its focus on K–12 student success and educational outcomes. Smaller grant sizes and localized scope reduce alignment compared to higher-funding responders but remain within successful engagement ranges.
The Miami FC	0.25	Despite thematic overlap with responding MLS organizations, The Miami FC more closely resembles non-responding sports organizations with broader social missions and weaker alignment to education technology innovation.

Organization	Score	Prompt 3 Rationale
Carolina Hurricanes Foundation	0.25	This foundation aligns with non-responding NHL organizations that emphasized traditional youth programming over innovative education technology. Its smaller grant sizes and service-oriented model reduce likelihood of engagement.
Vancouver Whitecaps FC Community	0.25	Although thematically similar to some responding MLS foundations, this organization follows a non-responder pattern characterized by localized scope and traditional community foundation structures.
Carolina Panthers Charities	0.25	This NFL-affiliated foundation mirrors a strong non-responder pattern observed across professional football organizations, despite thematic alignment with youth education and sports.
Minnesota Vikings Foundation	0.25	While the foundation supports youth education and STEM initiatives, its NFL affiliation strongly correlates with historical non-response patterns seen across similar organizations.
Cleveland Browns Foundation	0.25	The Browns Foundation closely matches non-responding professional sports foundations, where organizational type outweighed mission alignment in predicting engagement.
Detroit Lions Charities	0.25	Despite youth and sports alignment, this foundation follows the non-responder profile of NFL-affiliated organizations with traditional grant structures and limited innovation focus.
Seattle Seahawks — Spirit of 12	0.25	This organization demonstrates thematic alignment but fits a broader NFL non-responder pattern, contrasting sharply with the MLS foundations that engaged with IterLight.
Oklahoma City Thunder Foundation	0.25	Although focused on education and underserved youth, the foundation aligns with NBA-affiliated organizations that consistently failed to respond, indicating organizational type as a dominant predictor.
Chicago Bears Charities	0.25	While grant size and mission alignment are positive indicators, the organization's NFL affiliation mirrors strong non-response patterns across similar foundations.
Los Angeles Chargers Impact Fund	0.25	This foundation exhibits characteristics of non-responders, including traditional education funding models and limited emphasis on innovation or technology integration.
Los Angeles Rams Foundation	0.25	Despite youth education focus, the Rams Foundation aligns with NFL non-responder trends characterized by small grant sizes and conventional philanthropic structures.
Baltimore Ravens Foundation	0.25	This foundation matches the non-responder profile seen across NFL organizations, where sports affiliation outweighed thematic alignment in predicting engagement.
Arizona Cardinals Charities	0.25	Arizona Cardinals Charities follows the dominant non-responder pattern among NFL foundations, contrasting with the MLS organizations that showed engagement.
Buffalo Bills Foundation	0.20	The foundation shows weak alignment with responders due to its broad community mission, small grant sizes, and highly localized scope relative to engaged organizations.
Utah Hockey Club Community Engagement	0.15	This organization aligns strongly with NHL non-responders that prioritized traditional sports and health programming over education technology innovation.

Organization	Score	Prompt 3 Rationale
Vegas Golden Knights Foundation	0.15	Despite some thematic overlap, the foundation closely matches the NHL non-responder profile, characterized by local scope and conservative grant structures.
Los Angeles Lakers Youth Foundation	0.15	Although mission-aligned, the foundation mirrors non-responding NBA organizations where outreach engagement was historically low despite similar grant ranges.
Tampa Bay Lightning Community Heroes	0.15	This foundation fits the non-responder pattern among NHL organizations, with rigid grant mechanisms and limited flexibility for innovative EdTech partnerships.
Chicago Blackhawks Foundation	0.15	Despite meaningful grant capacity, the foundation aligns with NHL non-responders where organizational type and traditional funding models dominated engagement outcomes.