

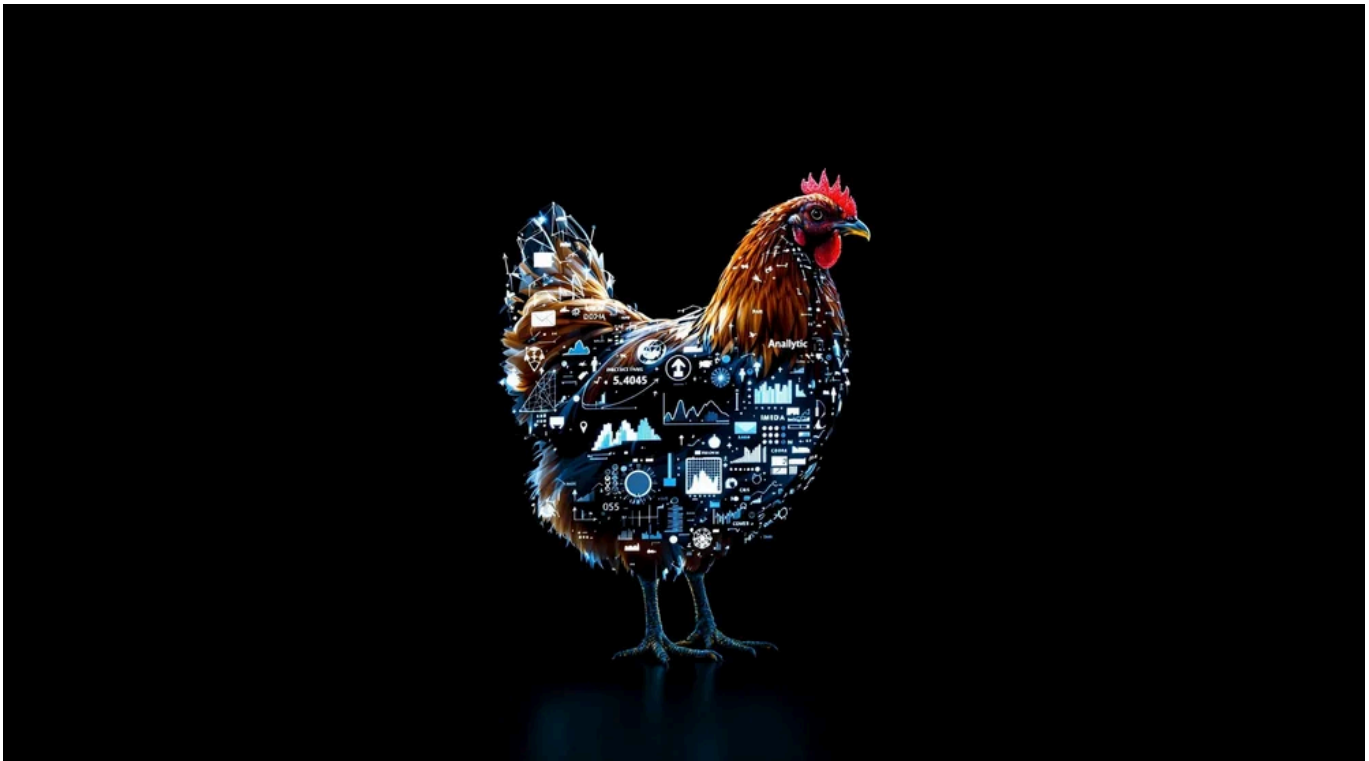


Tools for Activists Tools for Developers
Knowledge Base

Donate

Research & Reports

Building Predictive Models for Effective Animal Advocacy



Animal advocacy organizations collectively possess valuable data on what works and what doesn't, yet this knowledge remains trapped in organizational silos. Many groups (especially smaller ones) lack the technical resources to extract meaningful insights from their own data, let alone benefit from movement-wide patterns. This report details our collaborative project to break down these barriers by aggregating cross-organizational data and developing easy-to-use prediction tools that any advocacy organization can access.

Through partnerships with over 30 organizations, we've created standardized datasets and built a suite of machine learning models that can predict advocacy effectiveness across multiple dimensions. Our models achieve predictive accuracy that outperforms industry benchmarks, with predictions typically falling within 0.12-0.20 points of actual outcomes on a standardized 0-1 scale.

These tools are now freely available to all animal advocacy organizations, regardless of size or technical capacity.

Data Collection and Standardization Pipeline

Building a Collaborative Dataset

The foundation of this project was establishing data sharing agreements with diverse animal advocacy organizations and recruiting individual animal advocates for providing human feedback. This required months of relationship building, addressing privacy concerns, and creating secure data transfer protocols. The resulting dataset encompasses:

- Social media engagement metrics across platforms
- Email campaign performance data
- Website analytics and conversion metrics
- AI chatbot interaction logs
- Qualitative feedback and response data

This unprecedented collaboration allowed us to analyze patterns across the movement that would be impossible to detect within any single organization's data.

Creating a Unified Measurement Framework

To transform this diverse data into actionable insights, we developed a comprehensive standardization pipeline that addressed both objective performance metrics and subjective quality assessments.

Standardizing Objective Performance Metrics

Our first challenge was standardizing performance data from different channels and organizations into comparable metrics. We developed a three-stage process:

1) Initial Scoring and Channel-Specific Weighting: We created weighted scoring systems for each channel that prioritized meaningful engagement over surface-level metrics:

- **Social Media:** We weighted substantive interactions (shares, comments) higher than passive engagement (likes, views).
- **Email Outreach:** For corporate campaigns, commitments made in responses received the highest weight, followed by response sentiment, with open rates weighted lowest.
- **Email Marketing:** Click-through rates were weighted most heavily, while negative metrics (spam complaints, unsubscribes) were subtracted from the final score.
- **Website Analytics:** Concrete actions (donations, petition signatures) received the highest weights, followed by engagement metrics (time spent, return visits).
- **Chatbot Interactions:** Explicit behavior change commitments received the highest weight, followed by self-reported belief changes and positive sentiment.

2) Distribution Normalization: To address the statistical challenges of comparing different metrics:

- We applied logarithmic transformations to metrics with extreme outliers, such as viral social media posts. This prevented exceptional but rare successes from dominating our analysis while still recognizing their significance.
- We used z-score standardization for more normally distributed metrics, allowing us to identify content that performed exceptionally well relative to typical results within a specific context.

3) Group Size Balancing: To ensure fair comparisons between organizations of different sizes:

- We created a dual scoring system that balanced absolute impact (channel-wide standardization) with relative effectiveness (group-level standardization).
- This approach ensured that while a large organization's greater reach was reflected in scores, smaller organizations could still achieve competitive ratings by consistently outperforming their own benchmarks.

Standardizing Data for Subjective Metrics

While objective performance data provided a foundation for our Text Performance model, we needed additional approaches to capture subjective dimensions like trustworthiness, emotional impact, and rationality. We developed an innovative multi-perspective data generation strategy:

1. Human Volunteer Assessments: We recruited volunteers to rank and score advocacy content according to specific criteria, providing a human baseline for subjective quality judgments.
2. Synthetic Perspective Generation: To capture a broader range of viewpoints than would be possible with human volunteers alone, we used LLMs to simulate diverse perspectives, for example:
 - Animal Perspectives: "You are a chicken in a factory farm. Rank these messages based on how effectively they would help improve your welfare."
 - Target Audience Perspectives: "You are a middle-aged man in Ohio who eats meat regularly. Rank these messages based on how persuasive you find them."
 - Decision-Maker Perspectives: "You are a corporate sustainability officer at a major food company. Rank these advocacy emails based on how likely they would be to influence your policies."
5. Balanced Integration Approach: We combined human and synthetic feedback, using synthetic perspectives to fill gaps and provide diverse viewpoints that would be difficult to capture through human volunteers alone. This hybrid approach ensured our models could understand both typical human reactions

and the perspectives of underrepresented stakeholders, including the animals themselves.

By integrating both objective performance data and multi-perspective subjective assessments, our standardization pipeline created a unified framework for understanding advocacy effectiveness across its many dimensions.

Please note that the prompts provided above are simplified examples intended to illustrate the concept, the actual script we used for generating synthetic feedback data can be found [here](#).

Model Development Journey

Initial Approach and Limitations

Our first attempt used a taxonomy-based approach with over 1,000 factors hypothesized to influence persuasiveness, drawn from behavioral science, social psychology, and marketing frameworks. However, this method proved inadequate:

- Predictions showed unacceptably high inaccuracy ($\pm 150\%$)
- The model failed to capture nuanced relationships between content characteristics and outcomes
- Results were essentially unusable for practical planning purposes

Breakthrough with Text Embeddings

We pivoted to an embedding-based approach that transformed text into numerical representations capturing semantic relationships. This immediately improved results:

- Prediction accuracy improved to within $\pm 29.60\%$ of actual outcomes
- The model began showing practical utility for advocacy planning

Final Model Architecture and Performance

Building on these insights, we developed a suite of nine specialized DistilBERT-based text regression models, each focused on a different dimension of advocacy effectiveness. Our models fall into two complementary categories:

1. Performance Prediction Model: Trained on actual performance data from our standardized metrics pipeline, this model predicts how content will perform in real-world advocacy contexts.
2. Subjective Quality Models: Trained on human and synthetic feedback, these models predict various subjective dimensions of content quality:
 - Perceived Trustworthiness
 - Level of Rationality
 - Emotional Impact
 - Perceived Insightfulness
 - Cultural Sensitivity
 - Potential Influence
 - Effect on Animals
9. Preference Prediction Model: This model provides an aggregate prediction across all subjective quality dimensions, offering a holistic assessment of how content will resonate with audiences based on our multi-perspective subjective assessments.

This dual approach allows advocates to optimize both for predicted real-world performance and for specific subjective qualities that may be particularly important for their advocacy goals.

Understanding Our Accuracy Metric

The key metric we use to evaluate our models is:

RMSE (Root Mean Square Error)

. This tells us how far off our predictions typically are on our 0-1 scale. For example:

- An RMSE of 0.15 means our predictions are typically within 0.15 points of the actual outcome
- If we predict a piece of content will score 0.7 on trustworthiness, the actual score will typically be between 0.55 and 0.85
- Lower RMSE values indicate more precise predictions

Model Suite Performance

Our final models demonstrated strong predictive capabilities across multiple dimensions:

Model	What It Predicts	Typical Prediction Error (RMSE)
Cultural Sensitivity	How appropriate content is across different cultural contexts	0.12
Preference Prediction	How much audiences will prefer this content	0.13
Perceived Insightfulness	How insightful audiences will find the content	0.13
Emotional Impact	How emotionally resonant the content will be	0.14
Potential Influence	How likely the content is to influence <u>behavior</u> change	0.14
Perceived Trustworthiness	How trustworthy the content will appear	0.15
Level of Rationality	How logically sound arguments will seem	0.18
Text Performance	How well the content will perform across channels	0.18
Effect on Animals	The ultimate impact on animal welfare	0.20

Additional evaluation metrics can be found on the respective [Hugging Face model cards](#)

What These Results Mean in Practice

These accuracy levels translate to practical benefits for advocacy organizations:

Example 1: Email Campaign Testing

Imagine you've drafted five different email subject lines for a campaign promoting plant-based eating. Using our Preference Prediction model (RMSE = 0.13):

1. You run all five options through the model
2. The model predicts scores of 0.45, 0.58, 0.72, 0.61, and 0.49
3. Given the RMSE of 0.13, you can be confident that option #3 (0.72) will outperform option #1 (0.45), since even accounting for prediction error, they don't overlap
4. You select option #3, saving time and resources that would have been spent on A/B testing

Example 2: Corporate Outreach Optimization

When preparing materials for corporate cage-free campaigns, you can use our Trustworthiness model (RMSE = 0.15) to:

1. Draft multiple versions of your outreach letter
2. Identify which version is predicted to appear most trustworthy
3. Be confident that if the model shows a difference of more than 0.30 between versions (twice the RMSE), the higher-scoring version will genuinely be perceived as more trustworthy
4. Select the most trustworthy version to maximize your chances of success

Example 3: Predicting Animal Impact

Our Effect on Animals model (RMSE = 0.20) helps you understand which advocacy approaches will have the greatest ultimate impact:

1. When comparing different campaign strategies (e.g., corporate outreach vs. individual outreach)
2. The model can predict which approach will likely have greater impact on animal welfare
3. While this prediction has the highest error margin (0.20), it still provides valuable guidance for strategic planning
4. Organizations can focus resources on approaches predicted to have the greatest impact

Empowering the Movement with Easy-to-Use Tools

By open-sourcing these models, we're democratizing access to advanced analytics for animal advocacy organizations of all sizes. These tools enable advocates to:

- Test Before Investing: Simulate campaign effectiveness before committing resources
- Optimize Content: Generate multiple variations and identify the most promising options
- Train Specialized AI: Use our models to create advocacy-optimized AI assistants
- Learn Across Organizations: Benefit from movement-wide patterns and insights

How to Use These Tools Today

We've made these prediction tools accessible through:

1. API Access: Integrate predictions directly into your existing workflows
2. Downloadable Models: Run predictions locally for complete privacy

[Full documentation for using our predictive models can be found here.](https://www.openpaws.ai/knowledge-base/building-predictive-models-for-animal-advocacy)

Future Directions

This project represents just the beginning of data-driven animal advocacy. Future work will focus on:

- Developing models that can evaluate images and videos
- Creating specialized models for specific advocacy contexts (corporate outreach, individual outreach, etc.)
- Building automated systems that can generate and optimize advocacy content in real-time

These tools represent a fundamental shift in how animal advocacy can be conducted, from intuition-based to evidence-driven, from siloed to collaborative, and from resource-intensive to accessible. By democratizing access to advanced predictive analytics, we're not just improving individual campaigns but accelerating the entire movement toward more effective advocacy for animals.

 [How AI Works](#)

[Report on the Use of AI
in Animal Advocacy](#) 