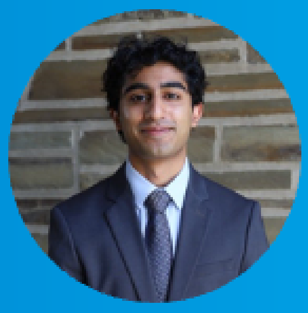
# Paws & Effect: Exploring the Impact of Vaccinations and Sterilizations



Prateek Gautam

Ankur Malik

## INTRODUCTION

India accounts for **36% of global rabies deaths**, with over **17.4 million dog bites annually**. Street dogs are the main transmission vector, and national laws prohibit killing, making **humane sterilization and vaccination** the only sustainable approach.

## PROBLEM

India's rabies control ecosystem is hindered by **siloed data and limited analytical integration**. It remains unclear which areas most urgently require sterilization based on dog density and bite trends, and what budget or staffing levels are needed to scale programs.

## OBJECTIVE

**Develop a data-driven framework** to quantify stray dog populations, assess intervention effectiveness, and guide resource allocation for rabies control. This is **supported by an intuitive tool** that helps the Humane team easily interpret model results and optimization outputs.

## DATA SOURCES

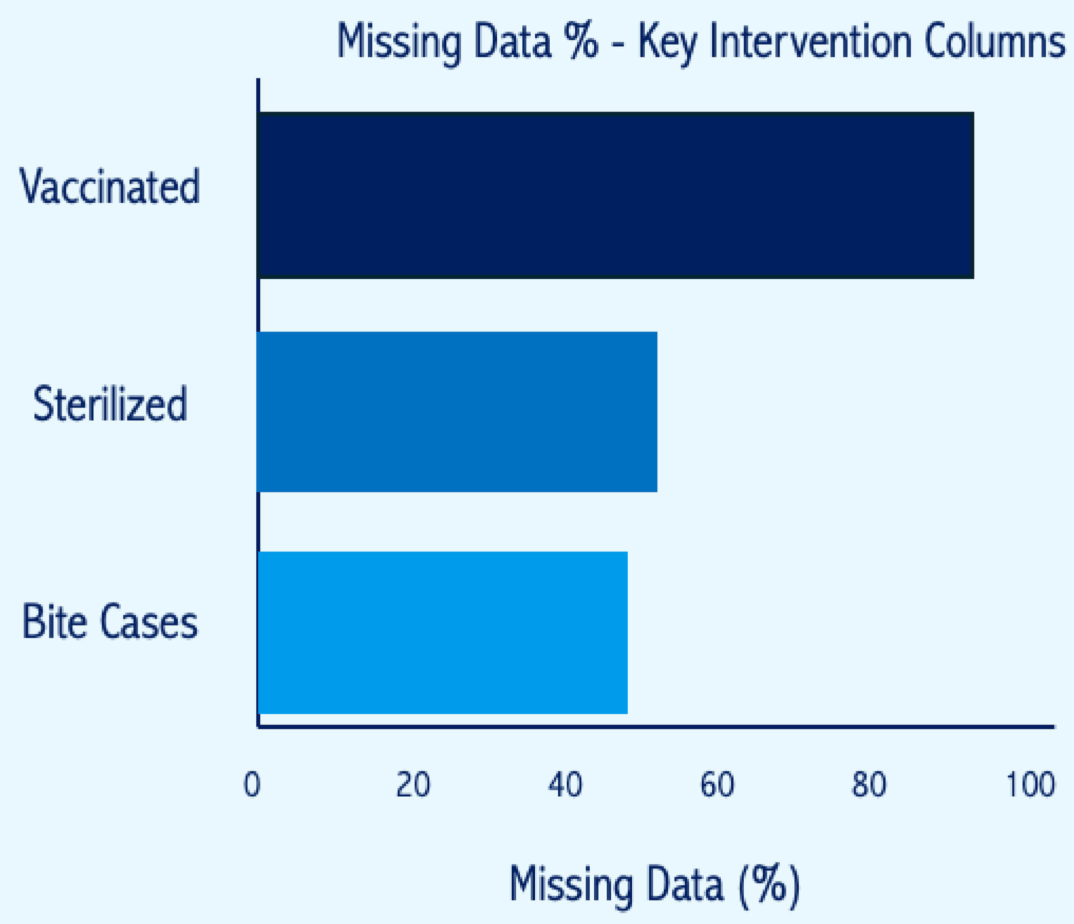
**2000+** PDF Reports

**50+** Cities and Municipalities

- Street dog counts and household surveys
- Government rabies and bite reports
- Peer-Reviewed studies and NGO reports
- Spay-neuter records and follow-up surveys

## DATA ENGINEERING

Much of the data existed as numbers buried in reports, so **we built a data extraction pipeline** to consolidate them into a structured CSV. This used a combination of OCR techniques and OpenAI API calls. Many of the data points were also missing or inconsistent, which we addressed through targeted imputation techniques as well as manual research beyond HWFA resources.




## OPTIMIZATION PREVIEW

$$\max \quad \alpha_1 \sum_{r,t} w(r)\rho(r) \cdot \frac{P_0(r) - P(r,t)}{P_0(r)} + \alpha_2 \sum_{r,t} w(r)\rho(r) \cdot \frac{V(r,t)}{P_0(r)} + \alpha_3 \sum_{r,v} x(r,v)\rho(r)w(r)$$

The **objective function maximizes a weighted combination of three priorities**: reducing stray dog populations, increasing vaccination coverage, and efficiently allocating resources. **Each component is weighted by region size and rabies risk.**

$$\sum_{r,t} s(r,t) \cdot c_s \le B_s, \quad \sum_{r,t} v(r,t) \cdot c_v \le B_v, \quad B_s + B_v \le B$$

**These constraints ensure that spending on sterilizations and vaccinations stays within their respective budgets**, and that their combined allocation does not exceed the total available budget.

$$\sum_{r} s(r,t) \le K_s, \quad s(r,t) \le P(r,t) \cdot \sum_{s} y(s) \cdot \gamma_s(s)$$

$$\sum_{s} y(s) = 1$$

**These constraints manage sterilization planning** by (1) limiting total monthly procedures to system capacity, (2) capping regional sterilizations based on scenario-specific rates, and (3) ensuring exactly one scenario is selected for optimization.

$$s(r,t) \le \sum_{v \in V} x(r,v) \cdot \alpha_v \quad \forall r \in R,\ t \in T$$

**This constraint ensures that the number of sterilizations in any region and time period does not exceed the total operational capac**ity provided by allocated resources (such as veterinarians, equipment, or facilities).

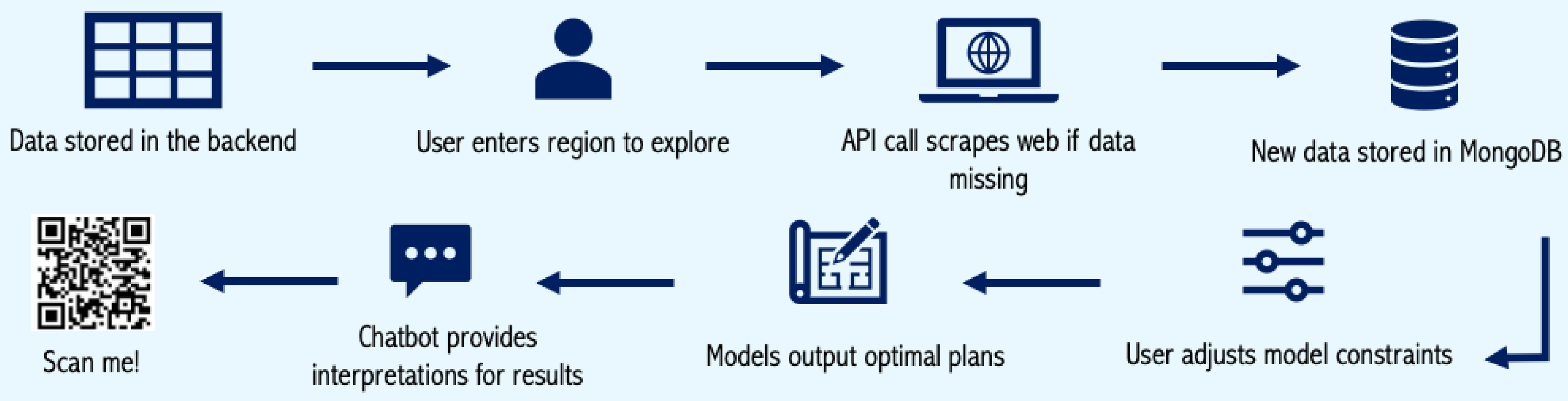
## MODELING

1. Hybrid LSTM Model (~40% weight): We use a bidirectional LSTM architecture augmented with additional covariates to capture sequential dependencies and temporal patterns. LSTMs are well-documented for effectively handling time-series classification tasks even with sparsity (Karim et al., 2017)

2. XGBoost Model (~30% weight): Gradient boosting machines, such as XGBoost, are robust for structured data with heterogeneous features, enabling us to capture complex non-linear relationships that are not purely temporal (Chen and Guestrin, 2016).

3. Prophet Model (~30% weight): Facebook Prophet is explicitly designed for time-series forecasting with missing data and irregular intervals, making it particularly well-suited to our sparse datasets (Taylor and Letham, 2018).

## INTERACTIVE TOOL




## CONCLUSION

The combination of modeling, optimization, and stakeholder engagement provides **a strong foundation for data-driven decision-making in street dog population management.** The models serve as a **flexible framework that can be continuously refined as more data becomes available.** Our interactive web tool empowers the Humane team to explore model outputs, assess trade-offs, and incorporate their on-the-ground expertise into planning.