

VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance

Edward Andrew P. Alverde¹, James Robert F. Dangbis², Voke Michael Oghenekaro³, Paul Emmanuelle Quimpo⁴, Johnny F. Verzola, MTS⁵

¹College of Computer Studies and Engineering, Lorma Colleges, edwardandrew.alverde@lorma.edu,

²College of Computer Studies and Engineering, Lorma Colleges, jamesrobert.dangbis@lorma.edu,

³College of Computer Studies and Engineering, Lorma Colleges, voke.oghenekaro@lorma.edu,

⁴College of Computer Studies and Engineering, Lorma Colleges, paulemmanuelle.quimpo@lorma.edu, and

⁵College of Computer Studies and Engineering, Lorma Colleges, johnny.verzola@lorma.edu

Abstract

Pet owners often struggle to interpret symptoms, apply basic first aid, or judge when a veterinary visit is necessary. In provincial settings such as La Union, Philippines, limited clinic access and geographic distance compound these gaps. This study designed, developed, and evaluated VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance, based on a Retrieval-Augmented Generation (RAG) architecture. The system was implemented with a Flutter mobile interface, a Django backend, PostgreSQL for transactional data, and ChromaDB for semantic retrieval, with Gemini 3 Flash as the language model. VetBot performs triage assessment, delivers general health guidance and safety-oriented first aid instructions, and provides limited home-based care advice for low-risk cases. Supporting modules include pet profiling with a digital vet card, a clinic locator covering La Union, and accessibility features comprising speech-to-text, text-to-speech, and bilingual English and Tagalog support.

The RAG pipeline was evaluated with the RAGAS framework on 99 clinical test queries. Faithfulness scored 0.889 and Answer Relevancy 0.906, indicating that responses drew on retrieved veterinary sources and addressed user queries directly. Context Precision (0.753) and Context Recall (0.691) reflected moderately high retrieval performance, with room for improvement in retrieval coverage and configuration. Answer Correctness scored 0.517; this lower score does not mean VetBot gave wrong answers, but rather that the metric compared responses word-for-word against a single pre-written reference answer, so any response that conveyed the same information in different wording was scored lower even if it was factually correct. Nine (9) licensed veterinarians evaluated VetBot's outputs on a 5-point Likert scale across nine criteria, yielding an overall mean of 4.56. Triage Report Utility received the highest rating (4.67); non-urgent scenarios received the highest urgency-level mean (4.65), followed by urgent (4.57) and emergency (4.47), with all three urgency levels remaining within the Strongly Agree range. Thirty-two (32) pet owners completed the System Usability Scale (SUS), producing a score of 79.61 (Grade A-, "Excellent"). The findings indicate that VetBot delivers reliable, clinically consistent pet health guidance and connects owners to professional veterinary services without replacing clinical judgment.

Keywords: *Veterinary Chatbot, Retrieval-Augmented Generation, Triage Assessment, Large Language Model, System Usability Scale, RAGAS, Pet Health Guidance*

1. Introduction

Pet owners regularly encounter situations in which their pets display concerning symptoms, including persistent fatigue, loss of appetite, minor injuries, and signs of distress. Existing research has shown that many pet owners struggle to judge the seriousness of such symptoms, particularly when veterinary care is constrained by geographic distance, financial limitations, or insufficient first aid knowledge. These conditions are most pronounced in rural and low-resource settings, where clinics often operate only during standard daytime hours. In these circumstances, owners commonly turn to online resources whose accuracy is difficult to verify. Because many cases require veterinary attention without constituting emergencies, owners are left uncertain about whether immediate action is warranted. The resulting delays in treatment, and the avoidable distress they cause, point to a need for accessible, reliable, and clinically validated decision-support tools.

Transportation and geographic distance have been consistently identified as primary structural barriers to veterinary access across multiple regions (Ng & Koh, 2023). In urbanized contexts with restrictive transport policies and high travel costs, pet owners in remote or low-income areas face delayed consultations, particularly during emergencies (Ng & Koh, 2023). The literature treats this barrier not as an isolated logistical issue but as a determinant of treatment timing and, consequently, of clinical outcomes.

Comparable transportation barriers exist in the Philippine setting, particularly in rural areas. In La Union, advanced and 24-hour emergency facilities are concentrated in San Fernando City, where clinics such as New Creation Animal Clinic and Diagnostic Center and PrimeCare Animal Recovery Clinic operate (New Creation Animal Clinic and Diagnostic Center, n.d.; PrimeCare Animal Recovery Clinic, n.d.) requiring residents of outlying municipalities to travel substantial distances. In Bacnotan and San Juan, owners depend largely on general-practice clinics operating on limited daytime schedules (Neervet Animal Clinic, n.d.; ELYU Veterinary Clinic, n.d.). These transportation constraints delay access to veterinary care and elevate health risks for companion animals in time-sensitive cases.

Treatment cost has been consistently identified as a primary determinant of veterinary consultation decisions (Bir et al., 2020; Park et al., 2021), with financial capacity exerting influence on lower-income households (Park et al., 2021). Across these studies, the recurring finding is that financial constraint shapes not only whether owners seek care but when. Consequently, owners commonly delay or skip recommended veterinary visits.

Income disparity in the Philippine setting amplifies the effects documented in the international literature. Regional studies attribute incomplete adherence to preventive measures, including vaccination and spaying or neutering, to financial constraints (San José et al., 2019; Dizon et al., 2022). Although awareness of responsible pet ownership is widespread, that awareness does not reliably translate into consistent care expenditure, producing a higher rate of untreated conditions and elevated risk of complications.

Pet-owner preparedness for veterinary emergencies remains low across surveyed populations, with most owners reporting no formal training in handling seizures, fractures, or other urgent conditions (MRCVS Online, 2024; Oberholtzer & Hofmeister, 2020). The gap persists even where willingness to learn is high. Oberholtzer and Hofmeister (2020) found that only a small proportion of owners could perform cardiopulmonary resuscitation

(CPR) on pets despite expressed interest in acquiring the skill, indicating that the constraint is access to training rather than motivation.

Comparable awareness-practice gaps are documented in the Philippine setting, where residence and educational attainment have been identified as significant determinants of owner knowledge and practice (Dizon et al., 2022; San José et al., 2019). Rural owners in particular show systematically lower preparedness, a pattern that increases reliance on non-clinical information sources during emergencies and compounds the access barriers already established in the regional literature.

Pet owners increasingly consult online platforms for veterinary advice, particularly social media groups, despite acknowledging concerns about source reliability (Kogan, Little, & Oxley, 2021). The persistence of this behavior, even where users themselves cite absent professional moderation and reliance on anecdotal experience as known weaknesses, points to a substitution effect: when veterinary access is delayed, inconvenient, or financially burdensome, accessibility and immediacy outweigh information quality in owner decision-making.

Equivalent patterns appear in developing contexts, including the Philippines, where adoption of digital veterinary tools has grown alongside the same regulation-and-validation concerns (Arshad et al., 2025; Diez & Wohlbe, 2025). The collective conclusion in this literature is that owner reliance on untested digital sources raises the risk of misinformation in emergencies and complex health conditions, defining a need that current consumer-facing applications do not yet meet.

These compounding barriers point to the need for an accessible, reliable, and scalable solution. Artificial intelligence (AI) presents one such direction. AI-based veterinary systems have been shown to improve clinical decision-making, support more accurate diagnosis, and reduce the cost of access compared to traditional consultations (Prashanna & Jaitly, 2025; Jin et al., 2025). Within this category, chatbots offer continuous availability and lower-cost consultation pathways (Jokar et al., 2024), and machine-learning-based first aid systems for dogs provide further evidence that AI can deliver immediate, context-sensitive assistance (Minoque et al., 2023).

The current developments in artificial intelligence have significantly increased its use in veterinary and medical practice. Domain-adapted Large Language Models have demonstrated the capacity to process and classify complex veterinary clinical data. VetLLM, a fine-tuned Alpaca-7B model, attained an F1 score of 0.747 on free-form veterinary notes from the Colorado State University Veterinary Teaching Hospital dataset, showing that general-purpose LLMs can be effectively adapted to veterinary-specific tasks (Jiang et al., 2024). Equally, VetBERT, created by Hur et al. (2020), used domain adaptation and instance selection to make a more precise classification of disease syndromes than general veterinary clinical notes. This model was trained using more than 15 million veterinary clinical records in the VetCompass Australia database, which has a broad range of species and clinical conditions. In order to overcome linguistic disparity between general biomedical and veterinary text, VetBERT adapted the BioBERT architecture and was fine-tuned on annotated veterinary syndrome datasets, yielding considerable improvements in syndrome classification and clinical concept extraction, and demonstrating the applicability of BERT-style transformers to domain-specific veterinary language (Hur et al., 2020). These models demonstrate that AI can be meaningfully applied

in veterinary practice, though challenges remain, particularly around data transparency and the high cost of fine-tuning.

While domain-adapted models such as VetLLM and VetBERT demonstrate strong performance within specialized veterinary datasets, they often require extensive domain-specific data, computational resources, and continuous fine-tuning to remain effective. In contrast to this, recent studies have begun to examine whether general-purpose Large Language Models (LLMs), trained on broader and more diverse datasets, can achieve comparable performance in veterinary-related tasks without the same level of domain-specific customization.

Recent benchmarking studies indicate that general-purpose LLMs perform competitively on veterinary tasks. On a set of 250 veterinary undergraduate-level multiple-choice questions, ChatGPT o1Pro and ChatGPT 4.5 achieved the highest accuracy rates at 90.4% and 90.8%, respectively, while other models scored between 64.8% and 87% (Alonso Sousa et al., 2025), reflecting wide variation in reasoning and visual-understanding capabilities across contemporary LLMs. In a related clinical setting, GPT-4 Omni achieved 96.9% sensitivity and 97.6% specificity when extracting information from feline health records (Wulcan et al., 2025).

These benchmarks establish technical competence in controlled or professional settings but leave the owner-facing case unaddressed. Pet owners face three specific constraints that examination-format and clinical-record-extraction evaluations do not capture: limited symptom interpretation, insufficient first aid knowledge, and restricted access to veterinary care. Bridging the benchmark literature to owner-facing deployment therefore requires deliberate design constraints around non-diagnostic tasks such as triage assessment, first aid guidance, and general care recommendations.

Prompt-engineering techniques alone can substantially extend the reasoning capability of broadly trained LLMs without domain-specific fine-tuning. Kojima et al. (2022) demonstrated that the prefix “Let’s think step by step” enabled InstructGPT (text-davinci-002) and PaLM (540B parameters) to function as zero-shot reasoners, producing accuracy gains of approximately 40–79 percentage points across arithmetic (MultiArith), symbolic (Last Letter), and logical (Date Understanding) reasoning tasks.

Comparative work indicates that general-purpose models offer practical advantages over fine-tuned counterparts for conversational and generative tasks (Rzādeczka et al., 2024; Villa et al., 2024). Villa et al. (2024) attribute these advantages to greater versatility, faster deployment, and lower maintenance costs, with fine-tuned systems retaining a clear edge only in narrow domains requiring highly specialized terminology. Rzādeczka et al. (2024) report convergent evidence at the dialogue level, finding that GPT-3.5 and GPT-4 produce more contextually coherent responses and more accurate intent recognition than specialized conversational agents. The collective implication is that a general-purpose LLM provides an adequate foundation for a chatbot supporting routine canine and feline care, provided that domain-specific accuracy is supplied by other architectural means.

The hallucination problem in general-purpose LLMs has produced Retrieval-Augmented Generation (RAG) as the dominant architectural response (Neha, Bhati, & Shukla, 2025). Hallucination refers to fluent but unsupported output, a failure mode that is particularly costly in health applications. RAG frameworks improve factual grounding, reduce hallucinations, and increase interpretability in healthcare contexts (Neha et al., 2025), and domain-specific implementations such as the RAGMed architecture

demonstrate measurable clinical accuracy gains when validated medical knowledge is retrieved before response generation (Patil et al., 2025). The convergent finding across these studies is that RAG complements general-purpose LLMs by anchoring outputs in authoritative source material rather than learned priors.

Veterinary and biomedical AI research extends these capabilities into diagnostic support, predictive analytics, and individualized health monitoring (Akinsulie et al., 2024; Pereira et al., 2023). Akinsulie et al. (2024) document veterinary AI systems supporting early disease detection, medical image interpretation, and real-time clinical decision support, while Pereira et al. (2023) report measurable accuracy and efficiency gains in veterinary radiology from automated pattern recognition. Considered alongside RAG-based generation, this body of work establishes that LLM-based systems can extend accessibility, reliability, and decision-support capacity in contemporary veterinary practice.

Despite an established literature on barriers to veterinary access, gaps in pet-owner knowledge, and the growing role of AI in veterinary medicine, an applied research gap remains. Existing decision-support systems and AI evaluations focus predominantly on clinical decision support, diagnostic accuracy, and professional users, with little attention to pet owners as end users. Much of the available pet health software and online information lacks source attribution, geographic specificity, and veterinarian validation, which limits reliability in first aid and emergency contexts. There is no integrated system that combines symptom-based triage assessment, first aid guidance, and location-specific clinic information for provincial or low-resource settings such as La Union, Philippines.

Based on these developments, this study developed VetBot, an AI-driven veterinary chatbot that provides general health guidance, safety-oriented first aid guidance, and limited home-based care guidance for dogs and cats. The system allows pet owners to enter observed symptoms and other relevant information, such as species, breed, and behavior, after which it performs a triage assessment to determine the urgency of the situation and recommend appropriate next steps. Depending on the triage outcome, the system provides limited home-based care guidance for low-risk cases, alongside safety-oriented first aid guidance and clear escalation instructions when veterinary attention is recommended. VetBot combines a Large Language Model for conversational reasoning with a Retrieval-Augmented Generation (RAG) framework to produce reliable responses backed by verified veterinary data.

VetBot implements four core functions: triage assessment, safety-oriented first aid guidance, limited home-based care guidance for low-risk cases identified during triage, and information retrieval from verified veterinary resources. VetBot's pet profiling subsystem allows owners to register multiple pets with profiles containing name, species, breed, age, sex, behavior patterns, vaccination status, allergies, known conditions, and current medications. These profiles let VetBot tailor responses to the specific pet under discussion and serve as the basis for the digital vet card, which is transmitted with each clinic booking to support pre-consultation review. Accessibility features include speech-to-text input, text-to-speech output, and bilingual operation in English and Tagalog. These components work toward a single goal: giving pet owners in La Union a reliable first point of contact for pet health concerns, in a language they can use, without requiring a clinic visit as the starting point.

This study contributes to the fields of veterinary care, animal welfare, and computer science. For licensed veterinarians, VetBot serves as a digital support tool that complements professional practice. It helps pet owners arrive at clinics better informed, having already taken appropriate non-invasive initial actions when applicable, which contributes to more efficient consultations. For pet owners, the study provides an accessible and practical platform for managing their pets' health and well-being. The study also contributes to computer science and artificial intelligence research by documenting an end-to-end Retrieval-Augmented Generation implementation in a veterinary triage context, an application area underrepresented in existing RAG literature, and by providing comparative data points for subsequent work on RAG performance in low-resource clinical domains, where standardized evaluation has been identified as a literature gap (Amugongo et al., 2025).

The development of VetBot was supported by a modern technology stack selected to meet the system's requirements for efficiency, scalability, and accessibility. Flutter and Dart were selected for cross-platform mobile development, satisfying the requirement that VetBot reach pet owners across both Android and iOS through single-codebase compilation. Django was selected as the backend framework to provide secure user authentication, REST API endpoints, structured request-response handling for the triage pipeline, and direct integration with the RAG retrieval and generation stages. PostgreSQL was selected as the relational database to provide ACID-compliant transactional integrity for pet medical records, since the data is inherently relational in structure: users own multiple pets, each pet has multiple recorded conditions and medications, and each booking links a pet to a clinic and a triage report. At the inference layer, a general-purpose Large Language Model (Gemini 3 Flash) was paired with a Retrieval-Augmented Generation pipeline that retrieves validated passages from a curated veterinary knowledge base before generation, which constrains the model to produce content traceable to authoritative sources.

2. Objectives

This study designed and developed VetBot, an AI-driven veterinary chatbot that integrates a Large Language Model (LLM) with Retrieval-Augmented Generation (RAG) to deliver triage assessment, general health guidance, safety-oriented first aid guidance, and limited home-based care guidance for dogs and cats. The study further evaluated the system's performance and usability in supporting pet owners' decisions when professional veterinary care is delayed, geographically inaccessible, or temporarily unavailable, while ensuring that cases requiring professional attention are escalated to licensed veterinarians.

Specifically, the study: (1) designed and developed a system architecture for VetBot, an AI-driven veterinary chatbot that integrates a Large Language Model with Retrieval-Augmented Generation to support triage assessment and deliver general health, safety-oriented first aid, and limited home-based care guidance for canines and felines; (2) evaluated the performance of VetBot in terms of accuracy, reliability, and responsiveness using Retrieval-Augmented Generation Assessment (RAGAS) and expert validation by licensed veterinarians; and (3) determined the usability level of VetBot using the System Usability Scale (SUS) among pet owners.

3. Materials and Methods

This study used a mixed-methods research design combining quantitative and qualitative approaches to evaluate the system's performance and acceptability. Quantitative data was collected through three instruments: Retrieval-Augmented Generation Assessment (RAGAS) for technical AI performance, an Expert Validation Form completed by nine (9) licensed veterinarians using a 5-point Likert scale across nine criteria including medical accuracy, clinical suitability, and safety, and the System Usability Scale (SUS) for user experience. Qualitative data were gathered through open-ended questions within the same expert validation form and analyzed thematically to surface professional observations not captured by Likert ratings. The combined design supported evaluation of both technical functionality and clinical alignment.

This study was conducted within the province of La Union, which served as the primary locale for the evaluation of VetBot. The selection of the province, rather than limiting the scope to a single city such as San Fernando City, allowed for broader representation of pet owners and licensed veterinary practitioners. This approach reflected real-world conditions where access to veterinary services may vary across different municipalities within La Union. The sample comprised forty-one (41) respondents: nine (9) licensed veterinarians and thirty-two (32) pet owners, alongside five (5) IT expert validators who reviewed the expert evaluation instrument. Two non-probability sampling techniques were used. Purposive sampling was applied to the veterinarians, who were selected based on clinical qualifications, practice experience, and active engagement in veterinary work within La Union. Convenience sampling was applied to the pet owners, who were selected from those accessible and willing to participate during the data-collection window. Pet-owner participants were required to (1) own at least one dog or cat, (2) be at least 18 years of age, (3) have basic experience in pet care or in handling minor health concerns, and (4) be willing to undergo usability testing and complete the System Usability Scale.

To address the first objective on system architecture, the researchers conducted a literature analysis to inform architectural decisions and component selection. Sources were gathered from academic databases such as Google Scholar, IEEE Xplore, PubMed, ACM Digital Library, and Semantic Scholar, using search terms including “veterinary chatbot,” “LLM architecture,” “retrieval-augmented generation,” “RAG framework,” “AI pet health,” “chatbot system design,” and “conversational AI in healthcare.” Authoritative veterinary references were consulted to construct the system's knowledge base, including the American Veterinary Medical Association (AVMA), the Merck Veterinary Manual, the American Animal Hospital Association (AAHA), the Cornell Feline Health Center, and VCA Animal Hospitals. The analysis examined existing AI chatbot system structures, technology stack comparisons (Flutter versus React Native; Django versus FastAPI; PostgreSQL versus MongoDB), RAG implementation strategies in healthcare, LLM benchmarking studies, and triage classification approaches in veterinary and medical AI systems.

To address the second objective, the researchers adopted Retrieval-Augmented Generation Assessment (RAGAS) as the primary performance evaluation framework. Es et al. (2024) introduced RAGAS for the automated evaluation of RAG pipelines, addressing the challenge that RAG architectures must be assessed across distinct but interdependent components: the retrieval system's ability to surface relevant context

passages, the language model's ability to use those passages faithfully, and the overall quality of the generated response. The original framework defined three core metrics: Faithfulness, Answer Relevance, and Context Relevance that target these dimensions independently and support objective, scalable evaluation without dependence on human-annotated ground truth (Es et al., 2024). The RAGAS Python library extends this set with Context Precision, Context Recall, and Answer Correctness, which provide more granular assessment of retrieval quality and semantic alignment with reference answers. RAGAS was selected for VetBot because its evaluation dimensions correspond directly to the properties VetBot was designed to uphold: factual grounding, retrieval relevance, and response quality.

A dataset of 99 clinical test queries was developed by the researchers through a structured development process, covering canine and feline scenarios across three urgency levels (non-urgent, urgent, and emergency), with curated reference answers aligned to authoritative veterinary sources to serve as the evaluation benchmark. Query items were formulated based on common canine and feline clinical presentations documented in the same authoritative veterinary sources used to build VetBot's knowledge base. Each query was written to reflect the kind of symptom description or care concern a pet owner would realistically submit through the chatbot interface, using plain, non-clinical language consistent with the target user population. Query items within each urgency level were further varied by species, symptom type, and clinical scenario to reduce evaluative bias toward any single case category. Reference answers were curated directly from the same authoritative veterinary sources used to construct the knowledge base. These reference answers served as the ground truth against which RAGAS computed semantic alignment metrics, particularly Answer Correctness.

The responses of VetBot were programmatically assessed using the RAGAS Python library to automatically calculate evaluation metrics. RAGAS generated numerical ratings on a scale of 0.0–1.0 across five dimensions: faithfulness, which measured the extent to which responses were grounded in retrieved veterinary sources and free from hallucinations; answer relevancy, which measured how directly responses addressed user queries; context precision, which measured whether retrieved documents were focused and relevant to the query; context recall, which determined whether all necessary information had been retrieved from the knowledge base; and answer correctness, which measured the semantic alignment between the generated response and a curated reference answer to assess whether the system conveyed factually accurate information.

Prior to the expert evaluation, the Expert Evaluation Form was reviewed and validated by five (5) IT expert validators to verify its appropriateness for assessing VetBot. Each validator assessed the instrument using a structured validation sheet, evaluating its structure, clarity, and relevance in capturing key aspects of the system's outputs. Their feedback was used to refine and finalize the form before its administration to licensed veterinarians.

Beyond automated evaluation, the factual accuracy and clinical suitability of VetBot's responses were validated by nine (9) licensed veterinarians using an Expert Evaluation Form employing a 5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). The instrument presented six predefined canine and feline scenarios spanning three urgency levels, with two scenarios per level (non-urgent, urgent, and emergency). For each scenario, evaluators rated VetBot's output on nine criteria: medical

accuracy, clinical suitability, urgency classification, escalation and referral, safety, appropriate scope, completeness, triage report utility, and vet portal chatbot appropriateness. The first seven criteria covered the pet-owner chatbot output, the eighth assessed the AI-generated triage report intended for veterinarians, and the ninth evaluated the veterinarian portal chatbot's responses in a professional context. Weighted means were computed for each criterion and interpreted using a five-point agreement scale to gauge evaluator consensus.

To address the third objective, the researchers conducted structured usability testing sessions with thirty-two (32) pet owners, who served as VetBot's primary end users. Each session began with a brief demonstration of the chatbot interface, pet profile management, and clinic locator, after which participants interacted independently with VetBot by exploring pet-care scenarios of their own choosing and assessing the clarity and usefulness of its guidance. Following the session, participants completed the System Usability Scale (SUS), a standardized 10-item instrument rated on a 5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). SUS yields a single usability score from 0 to 100, interpreted using established score bands and grading criteria. The instrument was administered only to pet-owner participants; licensed veterinarians completed the expert validation of clinical accuracy and appropriateness rather than the SUS.

This study adhered to established ethical standards to protect all participants. All participants, including licensed veterinarians and pet owners, were provided with a clear explanation of the purpose and objectives of the study, including the procedures involved and their role in the research process. Participation was voluntary, and participants were informed that they could withdraw at any time without penalty. Written informed consent was obtained from all participants prior to data collection. All personal data collected from participants were treated as confidential. Data were anonymized during analysis and reporting to ensure that individual respondents could not be identified. The collection, storage, processing, and disposal of personal data complied with the Data Privacy Act of 2012 (Republic Act No. 10173). VetBot was designed solely as an informational and supportive tool and does not perform diagnosis, prescribe treatment, or replace professional veterinary care.

This study adopted Agile as its development methodology. The development process was organized into iterative sprints, each focusing on enhancing specific system components and functionalities. The development of VetBot followed a structured Agile cycle consisting of planning, design, implementation, testing, and review and refinement phases. These phases were continuously repeated across multiple sprints to support progressive system improvement. The implementation stage developed the core components using Flutter for the mobile interface, Django for backend processing, and PostgreSQL for data management, while the AI layer integrated Gemini 3 Flash as the language model for conversational reasoning and response generation, supported by ChromaDB as the vector database for semantic retrieval of verified veterinary information. The testing stage validated system features and conducted iterative assessments of AI-generated responses using RAGAS across multiple sprint cycles, with each run informing subsequent refinements to the retrieval configuration, prompt design, and knowledge base.

Quantitative analysis was applied to RAGAS metrics, expert evaluation data, and SUS scores. RAGAS scores were interpreted using a five-level scale: High (0.80–1.00), Moderately High (0.60–0.79), Moderate (0.40–0.59), Low (0.20–0.39), and Very Low

(0.00–0.19). Expert evaluation data were analyzed by computing the weighted mean for each of the nine criteria across all six scenarios, with mean scores interpreted using a standard five-point agreement scale: Strongly Agree (4.21–5.00), Agree (3.41–4.20), Neutral (2.61–3.40), Disagree (1.81–2.60), and Strongly Disagree (1.00–1.80). Results were further examined by urgency level (non-urgent, urgent, and emergency) to identify whether system performance varied across different case severities. Qualitative responses from the open-ended questions in the expert validation form were analyzed through thematic analysis to identify recurring themes, patterns, and observations related to the system's strengths, weaknesses, and areas for improvement. Usability was assessed using data from the System Usability Scale (SUS). Individual and mean SUS scores were calculated following the standard scoring procedure developed by Brooke (1996), with the overall SUS score mapped to its corresponding grade and adjective rating using the curved grading scale established by Sauro and Lewis (2016) and adjective descriptors adapted from Bangor, Kortum, and Miller (2008, 2009).

4. Results

This section presents the findings of the design, development, and evaluation of VetBot, organized according to the three research objectives. The first part describes the implemented architecture and system components of VetBot. The second part presents the results of the performance evaluation of the Retrieval-Augmented Generation pipeline, including automated RAGAS evaluation and expert validation by licensed veterinarians. The third part presents the usability evaluation conducted among pet owners.

To address the first research objective, the researchers built VetBot as an integrated platform composed of multiple interconnected modules. The system architecture is organized into four interconnected layers: the Frontend layer, implemented as a Flutter mobile interface; the Backend layer, built on Django API and business logic; the Database layer, managed through PostgreSQL; and the AI Layer, which integrates an LLM with a RAG framework and a triage-based decision support component. Communication across these layers follows three distinct flows. Runtime requests and responses are transmitted between the frontend and backend through API calls. Data storage and retrieval operations connect the backend to PostgreSQL, providing persistent management of user, pet, and session information. AI and RAG processing flows link the backend to the AI Layer, supporting embedding generation, vector retrieval from ChromaDB, and LLM-driven response generation.

The core functionality of VetBot is supported by a Retrieval-Augmented Generation pipeline that integrates a curated veterinary knowledge base with a Large Language Model to produce responses drawn from verified veterinary information. The knowledge base draws from authoritative veterinary sources, including the American Veterinary Medical Association (AVMA), the Merck Veterinary Manual, the American Animal Hospital Association (AAHA), the Cornell Feline Health Center, and VCA Animal Hospitals. The pipeline operates through two distinct flows. The first is the indexing and preparation flow, in which veterinary knowledge sources, organized by species and category, are preprocessed and divided into chunks of 900 characters with a 120-character overlap using LangChain's RecursiveCharacterTextSplitter. These chunks are then passed through BAAI/bge-base-en-v1.5, a HuggingFace embedding model that produces 768-

dimensional vector representations, and stored in ChromaDB, a SQLite-backed vector database that serves as the retrieval layer for the pipeline.

The second flow is the runtime query flow, triggered when a user submits a query through the mobile application. For multi-turn conversations, VetBot first rewrites the user's follow-up message into a self-contained search query using Gemini 3 Flash, preserving conversational context during retrieval. The rewritten query is encoded by BAAI/bge-base-en-v1.5 into a query embedding, which is then matched against ChromaDB through similarity search to retrieve the top 20 candidate chunks. A cross-encoder, ms-marco-MiniLM-L-6-v2, re-ranks these candidates by scoring each query-chunk pair for relevance and retains only the top 6. The re-ranked context is assembled by the Prompt Builder alongside the pet profile and conversation history, and the assembled prompt is passed to Gemini 3 Flash for response generation. The output includes guidance text, a triage level classification, escalation recommendations where applicable, and cited sources.

VetBot includes a triage assessment module that classifies pet conditions into three triage levels: non-urgent, urgent, and emergency. This classification is driven by a combination of prompt-based reasoning, in which Gemini 3 Flash evaluates the user's symptom descriptions against the retrieved veterinary knowledge base, and a dialogue state detection layer that monitors VetBot's own responses for urgency-related language in real time. The assessment begins with an emergency scan, where VetBot checks the user's input against known red-flag symptoms such as collapse, seizures, difficulty breathing, and severe bleeding. If no immediate emergency is detected, VetBot asks one to two focused follow-up questions to gather sufficient context before arriving at a triage classification. For cases where veterinary consultation is recommended, VetBot generates a structured vet-side report from the chat transcript and pet profile data, containing a conversation summary, triage assessment, and the assigned triage level, formatted for efficient review by veterinary professionals.

Supporting modules include a pet profile and digital vet card module that enables pet owners to register and manage detailed health records (name, species, breed, age, sex, vaccination history, deworming records, weight logs, and medical history); a veterinary clinic locator that allows users to access information about nearby veterinary clinics registered in the system, covering selected clinics within La Union; a veterinary portal and appointment management module that allows licensed veterinarians to register their clinic details and manage appointment requests submitted by pet owners; a telemedicine and messaging module that enables direct communication between pet owners and veterinarians through an in-app messaging thread; and accessibility and localization features comprising speech-to-text input, text-to-speech output, and bilingual operation in English and Tagalog. PostgreSQL was used as the primary database management system, integrated with the Django backend through Django's Object Relational Mapper. PostgreSQL follows ACID specifications, which maintain data consistency and integrity regardless of the number of concurrent users.

To evaluate the performance of VetBot in terms of accuracy, reliability, and responsiveness, the researchers used the Retrieval-Augmented Generation Assessment (RAGAS) framework. The evaluation was conducted using a dataset of 99 dog- and cat-related clinical queries with curated reference answers aligned to authoritative veterinary sources. The final evaluation results are presented in Table 1.

Table 1. RAGAS Evaluation Results for VetBot (Final Run, N = 99)

Metric	Mean	Interpretation
Faithfulness	0.889	High
Answer Relevancy	0.906	High
Context Precision	0.753	Moderately High
Context Recall	0.691	Moderately High
Answer Correctness	0.517	Moderate

As shown in Table 1, VetBot achieved High scores in both Faithfulness (0.889) and Answer Relevancy (0.906), indicating that responses were strongly grounded in the curated veterinary knowledge base and consistently addressed user queries. Context Precision (0.753) and Context Recall (0.691) both fell within the Moderately High range, indicating that most retrieved passages were relevant to the query and generally contained sufficient information to support the expected answer. Answer Correctness (0.517) was the lowest among the five metrics and fell within the Moderate range; this result, however, does not necessarily indicate that VetBot produced incorrect responses, since RAGAS measures correctness by comparing the generated response against a single fixed reference answer, and any response that conveys the same information differently will receive a lower score even if it is factually accurate.

To complement the automated RAGAS evaluation, the factual accuracy and clinical suitability of VetBot's outputs were validated by nine (9) licensed veterinarians using an Expert Evaluation Form. The evaluation covered six predefined canine and feline scenarios representing three levels of urgency: non-urgent (2 scenarios), urgent (2 scenarios), and emergency (2 scenarios). For each scenario, evaluators rated VetBot's output across nine criteria using a 5-point Likert scale. Table 2 (see page 13) presents the overall weighted mean scores for each of the nine evaluation criteria across all six scenarios.

Table 2. Expert Evaluation Results: Overall Weighted Mean per Criterion (N = 9)

Criterion	Mean Score	Interpretation
Medical Accuracy	4.46	Strongly Agree
Clinical Suitability	4.43	Strongly Agree
Urgency Classification	4.46	Strongly Agree
Escalation and Referral	4.59	Strongly Agree
Safety	4.61	Strongly Agree
Appropriate Scope	4.61	Strongly Agree
Completeness	4.61	Strongly Agree
Triage Report Utility	4.67	Strongly Agree
Vet Portal Chatbot Appropriateness	4.63	Strongly Agree
Overall Mean	4.56	Strongly Agree

As shown in Table 2, the overall weighted mean across all nine criteria was 4.56, which falls within the Strongly Agree range. Among the nine criteria, Triage Report Utility received the highest weighted mean (4.67), suggesting that the AI-generated triage reports provided to veterinarians were considered well-structured and useful for supporting case review and informed decision-making prior to consultation. Meanwhile, Clinical Suitability received the lowest weighted mean (4.43); despite being the lowest-rated criterion, the score still fell within the Strongly Agree range. To examine whether VetBot's performance varied across different levels of urgency, the weighted mean for each criterion was further analyzed by scenario type. Table 3 (see page 14) presents the results grouped by urgency level.

Table 3. Expert Evaluation Results: Weighted Mean per Criterion by Urgency Level (N = 9)

Criterion	Non-urgent	Urgent	Emergency
Medical Accuracy	4.50	4.50	4.39
Clinical Suitability	4.44	4.44	4.39
Urgency Classification	4.39	4.56	4.44
Escalation and Referral	4.72	4.56	4.50
Safety	4.72	4.61	4.50
Appropriate Scope	4.72	4.61	4.50
Completeness	4.72	4.61	4.50
Triage Report Utility	4.83	4.68	4.50
Vet Portal Chatbot Appropriateness	4.78	4.61	4.50
Overall Mean	4.65	4.57	4.47

As shown in Table 3, non-urgent scenarios received the highest overall mean (4.65), followed by urgent scenarios (4.57) and emergency scenarios (4.47). All three urgency levels fell within the Strongly Agree range, indicating that licensed veterinarians consistently rated VetBot's outputs positively regardless of case severity. The difference between the highest and lowest urgency-level means was only 0.18, suggesting that VetBot maintained a stable level of performance across all three urgency levels. The first seven criteria, which assessed the pet owner chatbot output, produced a combined weighted mean of 4.54. The eighth criterion (triage report utility) received a mean of 4.67, and the ninth criterion (veterinarian portal chatbot appropriateness) received a mean of 4.63.

Qualitative feedback from open-ended questions was analyzed through thematic analysis. In terms of key strengths, evaluators consistently highlighted VetBot's usability and practical value, describing the system as user-friendly for both veterinarians and pet owners, with the ability to gather relevant information and provide appropriate suggestions rather than overwhelming the user with excessive information. VetBot's accuracy, timely scheduling functionality, and ability to assist pet owners in locating nearby veterinary clinics were also identified as notable strengths. Regarding weaknesses, most evaluators reported no significant concerns, with several responding with "N/A" or "None"; one area for improvement identified was the need to improve response time when processing user queries. Recommendations included prioritizing direction of pet owners to the nearest clinic capable of handling emergency cases, with the option to reduce or bypass additional questions during emergency situations; addition of a vet-to-vet communication feature to facilitate coordination during patient transfers between clinics; and expansion of the knowledge base to cover rare cases in small animal practice.

To address the third research objective, the System Usability Scale (SUS) was administered to thirty-two (32) pet owners following structured usability testing sessions. Individual SUS scores were computed following the standard scoring procedure developed by Brooke (1996), and the overall mean SUS score was calculated to produce a single usability measure on a scale of 0–100. VetBot received an overall mean SUS score of 79.61, corresponding to a grade of A– and an adjective rating of Excellent on the curved grading scale. The score exceeded the benchmark of 68 reported by Sauro and Lewis (2016) across 241 industrial usability studies, placing VetBot above the empirical average. To further examine the usability perceptions of pet owners, individual item mean scores were computed for each of the ten SUS statements. Table 4 presents the mean scores for each item.

Table 4. Individual Item Mean Scores of the System Usability Scale (N = 32)

SUS Item	Mean Score
1. I think that I would like to use VetBot frequently.	4.41
2. I found VetBot unnecessarily complex.	2.00
3. I thought VetBot was easy to use.	4.50
4. I think that I would need the support of a technical person to be able to use VetBot.	2.19
5. I found the various functions in VetBot were well integrated.	4.47
6. I thought there was too much inconsistency in VetBot.	1.94
7. I would imagine that most people would learn to use VetBot very quickly.	4.53
8. I found VetBot very cumbersome to use.	2.00
9. I felt very confident using VetBot.	4.31
10. I needed to learn a lot of things before I could get going with VetBot.	2.25
Composite SUS Score (0–100)	79.61

As shown in Table 4, positively worded items (1, 3, 5, 7, and 9) consistently received high mean scores, while negatively worded items (2, 4, 6, 8, and 10) received low mean scores, indicating favorable usability perceptions across both dimensions. Among the positively worded items, Item 7 (“I would imagine that most people would learn to use VetBot very quickly”) received the highest mean score of 4.53, followed closely by Item 3 (4.50) and Item 5 (4.47). Among the negatively worded items, Item 6 (“I thought there was too much inconsistency in VetBot”) received the lowest mean score of 1.94, indicating that users perceived VetBot as consistent in its behavior and interface. The highest-scoring negatively worded items were Item 10 (2.25) and Item 4 (2.19), suggesting that some

participants needed brief familiarization before interacting with VetBot comfortably, consistent with the general nature of AI-based applications.

5. Discussion

The findings across the three research objectives converge on a single position: VetBot is technically feasible to build, clinically acceptable to licensed veterinarians, and usable by pet owners in the population it was designed to serve. Each of these claims rests on a distinct evidence base, and reading them together is what makes the system's contribution defensible as a decision-support tool rather than a clinical replacement.

The implementation results from the first objective imply that integrating an LLM with a Retrieval-Augmented Generation pipeline across a Flutter, Django, PostgreSQL, and ChromaDB stack is technically feasible for delivering veterinary guidance to pet owners. Separating transactional data from semantic retrieval, and routing all owner-facing responses through retrieved veterinary sources, gives the system the conditions to support safety and reliability. This separation matters because veterinary correctness cannot be audited inside a model's pre-trained weights; it has to be sourced from material that can be inspected, updated, and held to a standard. The architectural decision to anchor responses in authoritative veterinary sources before generation, rather than to ask the model to reason from prior knowledge, is therefore not a stylistic preference but a precondition for safety in this application area.

The performance evaluation results from the second objective imply that VetBot's outputs are clinically acceptable for owner-facing triage support. The Faithfulness score of 0.889 and the expert-validation mean of 4.56 indicate that responses stayed grounded in verified veterinary sources and were judged safe by licensed veterinarians, even in emergency scenarios where the bar for accuracy is highest. The Answer Correctness score of 0.517 deserves careful interpretation, since the metric compared each VetBot response to a single fixed reference answer; valid paraphrases of correct information were penalized under this scoring method. The high Faithfulness score and the Strongly Agree expert ratings imply that VetBot's responses were clinically sound even when they were not phrased identically to the reference, which is the expected outcome for an open-ended generative system. The Context Recall score of 0.691 implies a moderate gap in retrieval coverage that is best addressed at the retrieval layer rather than at the generation layer.

The qualitative feedback from veterinarians implies that the system's coverage has room to grow. Their suggestions regarding emergency-clinic prioritization, vet-to-vet coordination during patient transfers, and inclusion of rare small-animal cases point to areas where future iterations of VetBot could be expanded. The Triage Report Utility score of 4.67, the highest among the nine criteria, implies that the clinic-facing component of VetBot held strong practical value in the eyes of licensed veterinarians, suggesting that this aspect of the system resonated most directly with professional practice. The slightly lower scores in emergency scenarios (4.47 overall mean) may reflect the inherently higher expectations evaluators applied when assessing guidance for critical situations, where precision in urgency classification, escalation instructions, and safety-oriented language became more consequential. The difference between the highest and lowest urgency-level means was 0.18, and all three urgency levels remained within the Strongly Agree range, which we read as a sign of stable rather than uniform performance across case severities.

The usability results from the third objective imply that VetBot is usable enough to function as an owner-facing tool for the population it was designed for. Scoring above the 68-point benchmark, with no single usability area flagged as a consistent difficulty, means that pet owners with varying technical backgrounds can reach veterinary guidance without specialized training. The slightly higher scores on Items 4 and 10 point to brief familiarization as a useful onboarding addition rather than a structural usability barrier. The alignment between the SUS result and the expert evaluation is informative: VetBot was regarded as both professionally reliable by veterinarians and accessible to everyday users, which matches its design purpose as a veterinary guidance tool for pet owners rather than a clinical instrument for veterinarians.

Taken together, the three sets of findings establish that a RAG-based veterinary chatbot can serve as a genuine decision-support tool for pet owners in a provincial Philippine setting, on the condition that retrieval is grounded in authoritative veterinary sources and that the system limits itself to non-diagnostic outputs. The empirical contribution of this study to the broader RAG literature is the demonstration that the same architectural patterns developed for human medical and educational applications transfer effectively to a veterinary owner-facing context, where the literature has identified standardized evaluation as a gap (Amugongo et al., 2025). The remaining limitations, particularly Answer Correctness scoring artifacts, retrieval coverage gaps, and inherently stricter expectations under emergency conditions, define the agenda for the next iteration rather than weaknesses in the system as deployed for the population studied.

6. Conclusion

This study designed, developed, and evaluated VetBot, an AI-driven veterinary chatbot for canine and feline care guidance. VetBot's development and evaluation produced three findings that together establish its viability as a decision-support tool for pet owners in La Union.

The first objective, to design and develop VetBot's system architecture, was realized through an integrated platform built on a Flutter mobile interface, a Django backend, a PostgreSQL database, and a ChromaDB-backed Retrieval-Augmented Generation pipeline. The architecture was structured to support triage assessment, general health guidance, safety-oriented first aid guidance, and limited home-based care guidance for dogs and cats. Separating transactional data management in PostgreSQL from semantic retrieval in ChromaDB maintained data integrity while supporting efficient access to the curated veterinary knowledge base. Additional modules, including pet profiling with a digital vet card, a veterinary clinic locator, a veterinary portal with appointment management, and accessibility features such as speech-to-text, text-to-speech, and bilingual support, were implemented and integrated into a unified platform.

The second objective, to evaluate VetBot's performance in terms of accuracy, reliability, and responsiveness, was addressed through RAGAS evaluation and expert validation by licensed veterinarians. Faithfulness scored 0.889 and Answer Relevancy scored 0.906, both interpreted as High, confirming that responses were grounded in retrieved veterinary sources and consistently addressed user queries. Context Precision (0.753) and Context Recall (0.691) reflected moderately high retrieval performance. Answer Correctness scored 0.517, which reflected the metric's strict semantic alignment against fixed reference answers rather than a clinical deficiency. Expert validation by nine

(9) licensed veterinarians yielded an overall weighted mean of 4.56, interpreted as Strongly Agree across all nine evaluation criteria. Triage Report Utility received the highest rating at 4.67. Non-urgent scenarios received the highest overall mean (4.65), followed by urgent scenarios (4.57) and emergency scenarios (4.47), yet all three urgency levels remained within the Strongly Agree range, indicating that the system maintained consistent performance across all case severities.

The third objective, to determine VetBot's usability level among pet owners, was addressed through the System Usability Scale, yielding an overall score of 79.61, corresponding to a grade of A- and an adjective rating of Excellent. This score exceeded the empirical benchmark of 68 reported by Sauro and Lewis (2016) across 241 industrial usability studies. Item-level analysis showed that participants rated the system quick to learn, well-integrated, and consistent in its behavior, with no single usability area emerging as a persistent difficulty. The slightly higher scores on items related to technical support need and prior learning suggested that a brief onboarding walkthrough could further reduce the familiarization period for first-time users.

Across all three objectives, VetBot demonstrates that a RAG-based veterinary chatbot can serve as a genuine decision-support tool, helping pet owners in La Union take informed first steps and connect with professional veterinary services, without replacing clinical judgment.

Based on the results and findings, the researchers recommend the following for future development. On retrieval, the Context Recall score of 0.691 indicates that a portion of relevant passages were not consistently retrieved; a hybrid pipeline combining vector similarity with lexical matching such as BM25 would address cases where symptom qualifiers are split across chunks. On latency and triage logic, veterinarians identified response time as the only area requiring improvement, and Clinical Suitability scores were the lowest of the nine criteria; streaming outputs and asynchronous API calls should be implemented to reduce perceived latency, and the clarifying-question logic should be expanded to gather more context from users before triage classification. On onboarding, item-level SUS responses on technical support need and prior learning slightly pulled down the overall score of 79.61, suggesting that a guided walkthrough for first-time users is appropriate. On scope, veterinarians recommended expanding the knowledge base to cover rare small-animal cases and adding an emergency-clinic prioritization feature and a vet-to-vet coordination function for patient transfers; near-term expansions should prioritize emergency-clinic routing through live availability APIs and structured multimodal input with veterinarian-validated annotation. On sustainability, the Triage Report Utility score of 4.67 indicated that the clinic-facing component was more immediately useful in practice than the owner-facing one, so a B2B2C pilot engaging La Union clinics, supported by a digital triage intake dashboard, is recommended; later stages may include anonymized aggregated trend reporting for public-health authorities, subject to ethics review and the Data Privacy Act of 2012.

7. Acknowledgements

The researchers extend their earnest appreciation to those who provided the necessary resources and guidance throughout the duration of this study. First and foremost, the researchers offer their deepest praise to the Almighty God for His presence and for

imbuing them with the guidance, strength, and sustenance of faith needed to successfully finish this study.

To their thesis adviser, Mr. Johnny F. Verzola, MTS, for sharing his knowledge and expertise with them and providing his full support, as well as for believing in their abilities to move forward with this study. To their teacher in charge, Dr. Janelli M. Mendez, DIT, for her valuable insights and for consistently steering the researchers toward academic excellence while providing clarity and support during every stage of the requirement process.

To the Dean of the College of Computer Studies and Engineering and chairperson, Mr. Jeffrey B. Layco, MIS, in recognition of his transformative feedback, distinguished authority, and the patience he extended throughout the development of this study. To their panelists, Dr. Janelli M. Mendez, DIT, and Engr. Shekiro R. Raposas, MIS, MTS, for dedicating their time and for providing constructive feedback that significantly improved the quality of this study. To the faculty of the College of Computer Studies and Engineering, for the mentorship and help they provided, which were essential to the researchers' progress and the development of this project.

The researchers also dedicate this study to their parents, whose constant encouragement provided the foundation for their efforts, and to their friends for their steady support and shared perspectives. Lastly, the researchers thank the veterinary professionals who participated in the expert evaluation; their willingness to share their expertise was fundamental as the researchers worked to refine this study.

8. References

- Akinsulie, O. C., Idris, I., Aliyu, V. A., Shahzad, S., Banwo, O. G., Ogunleye, S. C., Olorunshola, M., Okedoyin, D. O., Ugwu, C., Oladapo, I. P., Gbadegoye, J. O., Akande, Q. A., Babawale, P., Rostami, S., & Soetan, K. O. (2024). The potential application of artificial intelligence in veterinary clinical practice and biomedical research. *Frontiers in Veterinary Science*, 11, 1347550. <https://doi.org/10.3389/fvets.2024.1347550>
- Alonso Sousa, S., Bukhari, S. S. U. H., Steagall, P. V., Bęczkowski, P. M., Giuliano, A., & Flay, K. J. (2025). Performance of large language models on veterinary undergraduate multiple-choice examinations: A comparative evaluation. *Frontiers in Veterinary Science*, 12, 1616566. <https://doi.org/10.3389/fvets.2025.1616566>
- Amugongo, L. M., Mascheroni, P., Brooks, S., Doering, S., & Seidel, J. (2025). Retrieval augmented generation for large language models in healthcare: A systematic review. *PLOS Digital Health*, 4(6), e0000877. <https://doi.org/10.1371/journal.pdig.0000877>
- Arshad, M. F., Ahmed, F., Nonnis, F., Tamponi, C., Scala, A., & Varcasia, A. (2025). Artificial intelligence and companion animals: Perspectives on digital healthcare for dogs, cats, and pet ownership. *Research in Veterinary Science*, 193, 105776. <https://doi.org/10.1016/j.rvsc.2025.105776>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>

- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114–123.
- Bir, C., Ortez, M., Widmar, N. J. O., Wolf, C. A., Hansen, C., & Ouedraogo, F. B. (2020). Familiarity and use of veterinary services by US resident dog and cat owners. *Animals*, 10(3), 483. <https://doi.org/10.3390/ani10030483>
- Bleiweiss, A. (2024, October 1). Evaluating medical RAG with NVIDIA AI endpoints and Ragas. NVIDIA Technical Blog. <https://developer.nvidia.com/blog/evaluating-medical-rag-with-nvidia-ai-endpoints-and-ragas/>
- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Diez, E., & Wohllebe, A. (2025). Mobile apps in veterinary medicine: Literature review on opportunities and benefits for veterinarians, animal owners and animal welfare. *International Journal of Management and Applied Technology*. <https://www.wrpublishing.org/index.php/ijmat/article/view/247>
- Dizon, T. J. R., Saito, N., Inobaya, M., Tan, A., Reñosa, M. D. C., Bravo, T. A., Endoma, V., Silvestre, C., Salunga, M. A. O., Lacanilao, P. M. T., Guevarra, J. R., Kamiya, Y., Lagayan, M. G. O., Kimitsuki, K., Nishizono, A., & Quiambao, B. P. (2022). Household survey on owned dog population and rabies knowledge in selected municipalities in Bulacan, Philippines: A cross-sectional study. *PLoS Neglected Tropical Diseases*, 16(1), e0009948. <https://doi.org/10.1371/journal.pntd.0009948>
- ELYU Veterinary Clinic. (n.d.). Professional information. *The Filipino Vet*. <https://thefilipinovet.com/veterinarian/ahmed-ali>
- Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2024). RAGAS: Automated evaluation of retrieval augmented generation. *arXiv*. <https://arxiv.org/abs/2309.15217>
- Hur, B., Baldwin, T., Verspoor, K., Hardefeldt, L., & Gilkerson, J. (2020). Domain adaptation and instance selection for disease syndrome classification over veterinary clinical notes. In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing (BioNLP 2020)* (pp. 156–166). <https://doi.org/10.18653/v1/2020.bionlp-1.17>
- Jiang, Y., Irvin, J. A., Ng, A. Y., & Zou, J. (2024). VetLLM: Large language model for predicting diagnosis from veterinary notes. *Pacific Symposium on Biocomputing*, 29, 120–133. <https://pubmed.ncbi.nlm.nih.gov/38160274/>
- Jin, Y., Wu, G., Seo, J., Park, S., Hur, S., Aliyeva, D., Park, J., & Kim, K. (2025). AI veterinary assistance: Enhancing clinical decision-making in animal healthcare. *IEEE Access*, 13, 119292–119304. <https://ieeexplore.ieee.org/document/11077153>
- Jokar, M., Abdous, A., & Rahmanian, V. (2024). AI chatbots in pet health care: Opportunities and challenges for owners. *Semantic Scholar*.

- <https://www.semanticscholar.org/paper/AI-chatbots-in-pet-health-care%3A-Opportunities-and-Jokar-Abdous/0e6220089b2fe51183780ed9a1954cd1a2747501>
- Khasanova Zafarkizi, M., & Suh, Y. (2025). Design and performance evaluation of LLM-based RAG pipelines for chatbot services in international student admissions. *Electronics*, 14(15), 3095. <https://doi.org/10.3390/electronics14153095>
- Kogan, L. R., Little, S., & Oxley, J. (2021). Dog and cat owners' use of online Facebook groups for pet health information. *Health Information & Libraries Journal*, 38(3), 203–223. <https://pubmed.ncbi.nlm.nih.gov/34051032/>
- Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *arXiv*. <https://doi.org/10.48550/arXiv.2205.11916>
- Meng, W., Li, Y., Chen, L., & Dong, Z. (2025). Using the retrieval-augmented generation to improve the question-answering system in human health risk assessment: The development and application. *Electronics*, 14(2), 386. <https://doi.org/10.3390/electronics14020386>
- Minoque, N., Graciella, F., & Tanamal, R. (2023). Utilization and management of technology in making decisions to assist owners in first aid for dogs. *Journal of Technology and Engineering*, 15(4), 45–55. <https://doi.org/10.54554/jtec.2023.15.04.003>
- MRCVS Online. (2024, May 10). Owners lack vital pet first-aid knowledge, survey finds. <https://mrcvs.co.uk/en/news/22435/Owners-lack-vital-pet-first-aid-knowledge%2C-survey-finds>
- Neervet Animal Clinic. (n.d.). Professional information. *The Filipino Vet*. <https://thefilipinonet.com/veterinarian/ken-oliver-badua>
- Neha, F., Bhati, D., & Shukla, D. K. (2025). Retrieval-augmented generation (RAG) in healthcare: A comprehensive review. *AI*, 6(9), 226. <https://doi.org/10.3390/ai6090226>
- New Creation Animal Clinic and Diagnostic Center. (n.d.). Our services. <https://newcreationvet.my.canva.site/>
- Ng, K. Y., & Koh, K. (2023). Toward equitable public transportation with pets: Accessing veterinary care under mobility constraints in Hong Kong through taxi fare analysis. *International Journal of Sustainable Transportation*, 18, 264–274. <https://www.semanticscholar.org/paper/Toward-equitable-public-transportation-with-pets%3A-Ng-Koh/2ceb0f41b28e9ba8a0c0845e27beead1063dc602>
- Oberholtzer, K., & Hofmeister, E. (2020). Knowledge, attitudes, and perceptions of pet owners regarding CPR in dogs and cats. *Journal of Veterinary Emergency and Critical Care*, 30(4), 456–462. <https://doi.org/10.1111/vec.12997>
- Park, R. M., Gruen, M. E., & Royal, K. (2021). Association between dog owner demographics and decision to seek veterinary care. *Veterinary Sciences*, 8(1), 7. <https://doi.org/10.3390/vetsci8010007>

- Patil, R., Abbidi, M., & Fannon, S. (2025). RAGMed: A RAG-based medical AI assistant for improving healthcare delivery. *AI*, 6(10), 240. <https://doi.org/10.3390/ai6100240>
- Pereira, A. I., Franco-Gonçalo, P., Leite, P., Ribeiro, A., Alves-Pimenta, M. S., Colaço, B., Loureiro, C., Gonçalves, L., Filipe, V., & Ginja, M. (2023). Artificial intelligence in veterinary imaging: An overview. *Veterinary Sciences*, 10(5), 320. <https://doi.org/10.3390/vetsci10050320>
- Prashanna, J., & Jaitly, V. (2025). AI-powered veterinary chatbot for automated diagnosis. Semantic Scholar. <https://www.semanticscholar.org/paper/AI-Powered-Veterinary-Chatbot-for-Automated-Prashanna-Jaitly/d1862286b2eb35307b2657ddfacc884bf85bda10d>
- PrimeCare Animal Recovery Clinic. (n.d.). PrimeCare Animal Recovery Clinic – La Union. Facebook. <https://www.facebook.com/yourcompletevetcareforless/>
- Rządeczka, M., Sterna, A., Stolińska, J., Kaczyńska, P., & Moskalewicz, M. (2024). The efficacy of conversational artificial intelligence in rectifying the theory of mind and autonomy biases: Comparative analysis. arXiv. <https://doi.org/10.48550/arXiv.2406.13813>
- San José, R., Magsino, P. J., & Bundalian, R. (2019). Pet owners' awareness on RA 9482 (Anti-Rabies Act of 2007) in Magalang, Pampanga, Philippines. *Heliyon*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6529737/>
- Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research* (2nd ed.). Morgan Kaufmann.
- Villa, L., Carneros-Prado, D., Dobrescu, C. C., Sánchez-Miguel, A., Cubero, G., & Hervás, R. (2024). Comparative analysis of generic and fine-tuned large language models for conversational agent systems. *Robotics*, 13(5), 68. <https://doi.org/10.3390/robotics13050068>
- Wulcan, J. M., Jacques, K. L., Lee, M. A., Kovacs, S. L., Dausend, N., Prince, L. E., et al. (2025). Classification performance and reproducibility of GPT-4 omni for information extraction from veterinary electronic health records. *Frontiers in Veterinary Science*. <https://www.frontiersin.org/journals/veterinary-science/articles/10.3389/fvets.2024.1490030/>

9. Appendices

APPENDIX A Letter of Request for IT Expert Validation



Campus of Learning and Innovation
CUI Bldg. Urbiztondo, San Juan, La Union 2514 Philippines

April 9, 2026

NAME OF VALIDATOR
Address of Validator

Good day!

We are fourth-year Bachelor of Science in Computer Science students from LORMA Colleges, currently conducting our thesis project entitled "**VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance**" as part of the requirements for **CS THESIS 2**. This study aims to design and develop **VetBot**, an AI-driven veterinary chatbot that provides triage assessment, general health guidance, safety-oriented first aid instructions, and limited home-based care recommendations for dogs and cats. The system integrates a Large Language Model (LLM) with a Retrieval-Augmented Generation (RAG) framework to ensure that responses are grounded in reliable and validated veterinary information.

In line with this, we respectfully request your expertise to evaluate our system, particularly in terms of its accuracy, reliability, functionality, and overall suitability as a support tool for pet owners. Your professional insights, comments, and recommendations will be invaluable in improving the quality, validity, and credibility of our research.

Rest assured that all information gathered will be treated with strict confidentiality and will be used solely for academic purposes. Your participation will greatly contribute to the successful completion of this study.

Thank you very much for your time and kind consideration.

Respectfully yours,


Edward Andrew Alverde
BSCS-IV
Lorma Colleges


James Robert Dangbis
BSCS-IV
Lorma Colleges


Paul Emmanuelle Quimpo
BSCS-IV
Lorma Colleges


Voke Michael Oghenekaro
BSCS-IV
Lorma Colleges


Johnny P. Verzola
Thesis Adviser
Faculty, College of Computer Studies and Engineering
Lorma Colleges

APPENDIX C

Letter of Request for Veterinarian Participation



Campus of Learning and Innovation
CU Bldg, Urbiztondo, San Juan, La Union 2514 Philippines

Date: April 11, 2026

Animal Land Veterinary Diagnostic Center
City of San Fernando, La Union

Dear Veterinary Team of Animal Land Veterinary Diagnostic Center:

Good day!

We are fourth-year students of Bachelor of Science in Computer Science at Lorma Colleges. We are currently conducting a thesis study entitled: "**VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance**" as partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science.

In this regard, we respectfully request the expertise of your licensed veterinarians to participate in an **expert validation session**. This evaluation will focus on **VetBot's clinical accuracy, safety, and usability** from a professional veterinary perspective. Recognizing the busy nature of your clinic's operations, the session is designed to be conducted at your convenience. We are more than happy to facilitate this evaluation through an online platform, a personal visit to your clinic, or an asynchronous review that fits within your team's preferred pace and schedule.

Your participation will contribute to the credibility of our study and provide recognition for your institution in our thesis acknowledgments, methodology, and possibly in research forums.

This study strictly adheres to the provisions of the **Data Privacy Act of 2012 (Republic Act No. 10173)**. Any personal information collected will be processed lawfully, fairly, and solely for academic research purposes.

Data Privacy Safeguards:

- Your participation is **voluntary**.
- You may decline to answer any question or withdraw from the study at any time without penalty.
- All information gathered will be treated with **strict confidentiality**.
- Your identity will be anonymized in the research report unless you provide explicit permission to disclose your name.
- Interview recordings (if permitted) will be used only for transcription purposes and will be securely stored and deleted after the completion of the study.

By agreeing to participate, you acknowledge that you have been informed of the purpose of the study and your rights as a research participant.

Should you have any concerns regarding this research, you may contact one of our researchers via email at jamesrobert.dangbis@lorma.edu or by phone at **09270697332**.


We sincerely hope for your favorable consideration. Your participation will greatly contribute to the success and credibility of this study.

Thank you very much for your time and support.

Respectfully yours,


Edward Andrew Alverde
BSCS-IV
Lorma Colleges


James Robert Dangbis
BSCS-IV
Lorma Colleges


Paul Emmanuelle Quimpo
BSCS-IV
Lorma Colleges


Volke Michael Oghenekaro
BSCS-IV
Lorma Colleges


Johnny F. Verzola
Thesis Adviser
Faculty, College of Computer Studies and Engineering, Lorma Colleges

APPENDIX D

Expert Evaluation Form for Licensed Veterinarians

VetBot | Expert Evaluation Form

EXPERT EVALUATION FORM

VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance

Evaluator Name / Anonymous	
Years of Practice	
Specialization / Focus	
Date of Evaluation	

Instructions

This form assesses the accuracy, clinical suitability, and safety of VetBot's generated chat responses in predefined canine and feline scenarios at different urgency levels (Mild / routine-style cases, Urgent, and Emergency — aligned with VetBot's triage vocabulary above).

For each scenario you will see pet details and reported symptoms. There is no "model answer" or checklist of what VetBot should say—please use your professional judgment together with the rating criteria below.

Rate each response using a 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree).

Important — scope of VetBot

- VetBot serves as an informational and supportive resource. It offers guidance and possible differentials, but it does not provide definitive diagnoses, prescriptions, or drug dosages.
- It is not a substitute for examination by a licensed veterinarian.
- Please evaluate responses within this intended scope.

Scenario Overview

#	Scenario	Species	Intended Urgency Level
1	Minor skin irritation	Canine	Mild / monitor or Non-urgent / routine
2	Mild diarrhea	Feline	Mild / monitor or Non-urgent / routine
3	Persistent vomiting	Canine	Urgent
4	Limping and swelling	Feline	Urgent
5	Suspected poisoning	Canine	Emergency
6	Difficulty breathing	Feline	Emergency

Scenario 1: Minor Skin Irritation in a Dog

Species	Dog
Breed	Labrador Retriever
Age	3 years old
Sex	Male
Reported Symptoms	Dog has mild redness and occasional scratching on the belly for 2 days. The dog has no open wounds or discharge. He is eating, drinking, and acting normally.

Evaluation:

Criterion	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Medical Accuracy <i>The response contains factually correct veterinary information.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clinical Suitability <i>The first aid/care recommendations are appropriate for the presented case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urgency Classification <i>The urgency level assigned matches the clinical severity of the case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escalation & Referral <i>The system appropriately directs the user to seek professional care when needed.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safety <i>The response avoids harmful, misleading, or dangerous recommendations.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Appropriate Scope <i>Avoids definitive diagnosis and prescription for pet owners; uses appropriate uncertainty language</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness <i>The owner-facing chat response covers main risks, monitoring, and when to seek vet/ER without critical omissions.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Triage Report Utility <i>The AI-generated triage summary for veterinarians is suitable for review: chief concern, timeline, key signs clear; urgency plausible; useful for pre-consult / handoff (not misleading or empty)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vet Portal Chatbot (professional mode) <i>Replies appropriate for a professional audience (depth and tone); consistent with patient context and triage report;</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Scenario 2: Mild Diarrhea in a Cat

Species	Cat
Breed	Domestic Shorthair
Age	2 years old
Sex	Female
Reported Symptoms	The cat had a soft stool earlier today. Cat is alert, playful, eating/drinking. Cat is not vomiting, lethargic, and does not have blood in stool.

Evaluation:

Criterion	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Medical Accuracy <i>The response contains factually correct veterinary information.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clinical Suitability <i>The first aid/care recommendations are appropriate for the presented case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urgency Classification <i>The urgency level assigned matches the clinical severity of the case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escalation & Referral <i>The system appropriately directs the user to seek professional care when needed.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safety <i>The response avoids harmful, misleading, or dangerous recommendations.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Appropriate Scope <i>Avoids definitive diagnosis and prescription for pet owners; uses appropriate uncertainty language</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness <i>The owner-facing chat response covers main risks, monitoring, and when to seek vet/ER without critical omissions.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Triage Report Utility <i>The AI-generated triage summary for veterinarians is suitable for review: chief concern, timeline, key signs clear; urgency plausible; useful for pre-consult / handoff (not misleading or empty)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vet Portal Chatbot (professional mode) <i>Replies appropriate for a professional audience (depth and tone); consistent with patient context and triage report;</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Scenario 3: Persistent Vomiting in a Dog

Species	Dog
Breed	Shih Tzu
Age	5 years old
Sex	Female
Reported Symptoms	The dog has vomited 4 times in the past 6 hours. Refusing food, drinking very little water, appearing lethargic. No known ingestion of foreign objects or toxins. Last vaccination was 8 months ago.

Evaluation:

Criterion	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Medical Accuracy <i>The response contains factually correct veterinary information.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clinical Suitability <i>The first aid/care recommendations are appropriate for the presented case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urgency Classification <i>The urgency level assigned matches the clinical severity of the case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escalation & Referral <i>The system appropriately directs the user to seek professional care when needed.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safety <i>The response avoids harmful, misleading, or dangerous recommendations.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Appropriate Scope <i>Avoids definitive diagnosis and prescription for pet owners; uses appropriate uncertainty language</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness <i>The owner-facing chat response covers main risks, monitoring, and when to seek vet/ER without critical omissions.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Triage Report Utility <i>The AI-generated triage summary for veterinarians is suitable for review: chief concern, timeline, key signs clear; urgency plausible; useful for pre-consult / handoff (not misleading or empty)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vet Portal Chatbot (professional mode) <i>Replies appropriate for a professional audience (depth and tone); consistent with patient context and triage report;</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Scenario 4: Urgent Case: Limping and Swelling in a Cat

Species	Cat
Breed	Persian
Age	4 years old
Sex	Male
Reported Symptoms	The cat has been limping on its front left leg since this morning. There is a visible swelling around the paw area. The cat is eating less than usual and hissing when the paw is touched. There is no visible open wound.

Evaluation:

Criterion	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Medical Accuracy <i>The response contains factually correct veterinary information.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clinical Suitability <i>The first aid/care recommendations are appropriate for the presented case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urgency Classification <i>The urgency level assigned matches the clinical severity of the case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escalation & Referral <i>The system appropriately directs the user to seek professional care when needed.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safety <i>The response avoids harmful, misleading, or dangerous recommendations.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Appropriate Scope <i>Avoids definitive diagnosis and prescription for pet owners; uses appropriate uncertainty language</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness <i>The owner-facing chat response covers main risks, monitoring, and when to seek vet/ER without critical omissions.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Triage Report Utility <i>The AI-generated triage summary for veterinarians is suitable for review: chief concern, timeline, key signs clear; urgency plausible; useful for pre-consult / handoff (not misleading or empty)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vet Portal Chatbot (professional mode) <i>Replies appropriate for a professional audience (depth and tone); consistent with patient context and triage report;</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Scenario 5: Emergency Case: Suspected Poisoning in a Dog

Species	Dog
Breed	Beagle
Age	1 year old
Sex	Male
Reported Symptoms	The dog ingested an unknown substance from a neighbor's yard approximately 30 minutes ago. The dog is currently drooling excessively, trembling, and appears disoriented. The dog vomited once that looked foamy.

Evaluation:

Criterion	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Medical Accuracy <i>The response contains factually correct veterinary information.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clinical Suitability <i>The first aid/care recommendations are appropriate for the presented case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urgency Classification <i>The urgency level assigned matches the clinical severity of the case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escalation & Referral <i>The system appropriately directs the user to seek professional care when needed.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safety <i>The response avoids harmful, misleading, or dangerous recommendations.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Appropriate Scope <i>Avoids definitive diagnosis and prescription for pet owners; uses appropriate uncertainty language</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness <i>The owner-facing chat response covers main risks, monitoring, and when to seek vet/ER without critical omissions.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Triage Report Utility <i>The AI-generated triage summary for veterinarians is suitable for review: chief concern, timeline, key signs clear; urgency plausible; useful for pre-consult / handoff (not misleading or empty)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vet Portal Chatbot (professional mode) <i>Replies appropriate for a professional audience (depth and tone); consistent with patient context and triage report;</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Scenario 6: Emergency Case: Difficulty Breathing in a Cat

Species	Cat
Breed	Siamese
Age	7 years old
Sex	Female
Reported Symptoms	The cat is breathing with its mouth open, making wheezing sounds, and sitting in an unusual hunched position. The cat's gums appear pale. It started about 1 hour ago and is getting worse. Cat has no known pre-existing conditions.

Evaluation:

Criterion	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Medical Accuracy <i>The response contains factually correct veterinary information.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clinical Suitability <i>The first aid/care recommendations are appropriate for the presented case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urgency Classification <i>The urgency level assigned matches the clinical severity of the case.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Escalation & Referral <i>The system appropriately directs the user to seek professional care when needed.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Safety <i>The response avoids harmful, misleading, or dangerous recommendations.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Appropriate Scope <i>Avoids definitive diagnosis and prescription for pet owners; uses appropriate uncertainty language</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completeness <i>The owner-facing chat response covers main risks, monitoring, and when to seek vet/ER without critical omissions.</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Triage Report Utility <i>The AI-generated triage summary for veterinarians is suitable for review: chief concern, timeline, key signs clear; urgency plausible; useful for pre-consult / handoff (not misleading or empty)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vet Portal Chatbot (professional mode) <i>Replies appropriate for a professional audience (depth and tone); consistent with patient context and triage report;</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

APPENDIX E

Open-Ended Section of the Expert Evaluation Form

VetBot | Expert Evaluation Form

OPEN QUESTIONS

1. What are the key strengths of VetBot based on the scenarios reviewed?

2. What are the key weaknesses or areas for improvement?

3. Do you have any recommendations for the project?

Thank you for your time and professional input. Your evaluation will contribute to improving VetBot's accuracy, safety, and usefulness as a supportive tool for pet owners & veterinarians.

APPENDIX F

Informed Consent Form for Pet Owners



Campus of Learning and Innovation
CLI Bldg. Urbiztondo, San Juan, La Union 2514 Philippines

INFORMED CONSENT FORM

For Pet Owner Participants

Study Title: VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance
Researchers: Edward Andrew P. Alverde, James Robert F. Dangbis, Voke Michael Oghenekaro, and Paul Emmanuelle Quimpo
Institution: College of Computer Studies and Engineering, Lorma Colleges
Thesis Adviser: Johnny F. Verzola, MTS

We, Edward Andrew Alverde, James Robert Dangbis, Paul Emmanuelle Quimpo, and Voke Michael Oghenekaro, are 4th year students of Bachelor of Science in Computer Science at Lorma Colleges. We are currently conducting a thesis study entitled: **"VetBot: An AI-Driven Veterinary Chatbot for Canine and Feline Care Guidance"** as partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science. This study evaluates VetBot, an AI-driven veterinary chatbot that provides general health guidance, triage assessment, and safety-oriented first aid guidance for dogs and cats. You are being invited to participate in the usability evaluation of VetBot as a pet owner.

In this regard, we respectfully request your consent to participate a brief demonstration of VetBot, interact with the system by performing simulated pet-care scenarios, and complete the **System Usability Scale (SUS)**, a standardized 10-item questionnaire rated on a scale of **1 (Strongly Disagree)** to **5 (Strongly Agree)**. The session is expected to last approximately 30 to 45 minutes.

Your participation is entirely voluntary. You may withdraw at any time without penalty and request that your data be excluded. This study adheres to the Data Privacy Act of 2012 (RA 10173), and all information will be treated with strict confidentiality and anonymized in the report.

For concerns or questions, contact **Paul Emmanuelle Quimpo** at paulemmanuelle.quimpo@lorma.edu or **09391331146**.


By signing below, I confirm that I have read and understood the information above, that I meet the eligibility criteria, and that I agree to participate voluntarily in the usability evaluation of VetBot.

Printed Name and Signature of Participant

Date


Respectfully yours,

Edward Andrew Alverde
BSCS-IV
Lorma Colleges


James Robert Dangbis
BSCS-IV
Lorma Colleges


Paul Emmanuelle Quimpo
BSCS-IV
Lorma Colleges


Voke Michael Oghenekaro
BSCS-IV
Lorma Colleges


Johnny F. Verzola
Thesis Adviser
Faculty, College of Computer Studies and Engineering
Lorma Colleges

APPENDIX G
System Usability Scale (SUS) Questionnaire

VetBot Usability Evaluation – System Usability Scale (SUS)

You have just completed a usability testing session with VetBot, an AI-driven veterinary chatbot designed to provide triage assessment, first aid guidance, and general care recommendations for dogs and cats.

The following questionnaire consists of 10 statements about your experience using VetBot. Please rate each statement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree). There are no right or wrong answers – we are interested in your honest perception of the system.

If you feel you cannot respond to a particular statement, please select the middle option (3).

Your responses will be kept confidential and will only be used for research purposes.

 [Switch accounts](#)



 Not shared

[Next](#)

[Clear form](#)

SECTION 1: INFORMED CONSENT

Please read the following information carefully before proceeding:

Purpose of the Study: This survey is part of a research study evaluating VetBot, ^{*} an AI-driven veterinary chatbot that provides triage assessment, first aid guidance, and general care recommendations for dogs and cats. Your feedback will help assess the system's usability and overall user experience.

What You Will Do: You will answer questions about your experience using the VetBot system. The survey includes the standard 10-item System Usability Scale (SUS) and follow-up questions about your experience.

Eligibility: This survey is for pet owners (dog, cat, or both) who have interacted with the VetBot system prototype during the usability testing session.

Voluntary Participation: Your participation is completely voluntary. You may skip questions or stop at any time.

Confidentiality: Your responses are confidential. Data will be analyzed in aggregate and no personally identifiable information will be reported. This study adheres to the Data Privacy Act of 2012 (RA 10173).

No Compensation: There is no monetary compensation for participating in this survey.

Contact Information: If you have questions about this survey, please contact jamesrobert.dangbis@lorma.edu.

"I have read and understood the information above, and I voluntarily agree to participate in this survey."

Back

Next

Clear form

SECTION 2: RESPONDENT INFORMATION

Type of pet *

- Cat
- Dog
- Both

How long have you been a pet owner? *

- Less than 6 months
- 6 months to 1 year
- 1 to 3 years
- 3 to 5 years
- More than 5 years

Back

Next

Clear form

SECTION 3: SUS QUESTIONNAIRE

Please rate each statement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree) based on your experience using VetBot.

*

	1 - Strongly Disagree	2 - Disagree	3 - Neutral	4 - Agree	5 - Strongly Agree
I think that I would like to use VetBot frequently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found VetBot unnecessarily complex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought VetBot was easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that I would need the support of a technical person to be able to use VetBot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the various functions in VetBot were well integrated.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought there was too much inconsistency in VetBot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would imagine that most people would learn to use VetBot very quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found VetBot very cumbersome to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt very confident using VetBot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I needed to learn a lot of things before I could get going with VetBot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Back

Submit

Clear form

This form was created inside The LORMA Schools, Medical Center & Foundation. - [Contact form owner](#)

Does this form look suspicious? [Report](#)

Google Forms

10. Author(s) Biodata

Mr. Edward Andrew P. Alverde is a Bachelor of Science in Computer Science candidate at Lorma Colleges, College of Computer Studies and Engineering. He is from Urayong, Bauang, La Union. He has been a consistent scholar from his freshman through junior years and served as the CCSE Student Body Organization 3rd Year Peace Officer for the 2024–2025 academic year. He earned the ITPEC Information Technology Passport (IP) certification in April 2025 and has attended training and conferences in Natural Language Processing, robotics and AI, the INNOVATHON Region 1 Startup Sprint Hackathon 2025, and the Agentblazer Champion Workshop by SmartBridge in collaboration with Salesforce. His research involvement includes adaptation studies on online learning, internet productivity among senior high school students, and gender minority stress, in addition to the present study on VetBot.

Mr. James Robert F. Dangbis is a Bachelor of Science in Computer Science candidate at Lorma Colleges, College of Computer Studies and Engineering. He is from Naguilian, La Union, and has been a consistent scholar across his college years. He placed second in the Coder's Cup during CCSE Week 2023 and was a Web Design Finalist in the 2nd CITEDH-R1 Regional Research Conference and IT Skills Competition (CRRCISC) 2024. He earned the ITPEC Information Technology Passport (IP) certification in April 2025 and has attended training workshops and conferences in Natural Language Processing, robotics and AI, Wadhvani Ignite, the INNOVATHON Region 1 Startup Sprint Hackathon 2025, and Agentblazer Champion training. His prior research involvement includes work on water filtration using natural resources and on gender equality in educational institutions, alongside the present study on VetBot.

Mr. Voke Michael Oghenekaro is a Bachelor of Science in Computer Science candidate at Lorma Colleges, College of Computer Studies and Engineering. Originally from Abuja, Nigeria, he completed his earlier education through schools including Bristol Academy and Government Science Secondary School Pyakasa, where he placed among the highest UTME and WAEC scorers in his cohort and earned distinctions in WAEC 2017. He earned the ITPEC Information Technology Passport (IP) certification in April 2025. His prior research work includes a study on gender equality in educational institutions, alongside the present study on VetBot.

Mr. Paul Emmanuelle Quimpo is a Bachelor of Science in Computer Science candidate at Lorma Colleges, College of Computer Studies and Engineering. He is from Bacnotan, La Union, and was a Salutatorian in his senior high school class at Lorma Colleges Senior High, where he received Outstanding Student awards in Robotics, Web Development, and Java, and was a finalist in the City of San Fernando La Union ICT Summit Hackathon 2019. He earned the ITPEC Information Technology Passport (IP) certification in November 2024 and completed TryHackMe cybersecurity training. He has attended training in Natural Language Processing and the Global Conference on Robotics and Artificial Intelligence Technologies. His prior research involvement includes Geosense, an Arduino-based earthquake detection alert system, alongside the present study on VetBot.

Mr. Johnny F. Verzola, MTS, served as the thesis project adviser. As a member of the faculty of the College of Computer Studies and Engineering at Lorma Colleges, he supervised the research project until completion and contributed to the methodological direction of the study.