Language Models in Veterinary Clinical Practice

Applications, Risks, and Practical Guidance

Nathan Bollig, DVM, MS[a,*],
Jonathan L. Lustgarten, MS, PhD, VMD, CSPO[b],
Elizabeth Venit, VMD, MPH[c]

KEYWORDS

• Artificial intelligence • Large language models • Natural language processing
• Clinical decision support • Responsible AI • AI in veterinary medicine

KEY POINTS

- Large language models can support client communications, medical record documentation, clinical decision support, and clinical practice reviews.
- Advantages must be weighed against concerns such as accuracy, data privacy and security, introduction of biases, and clinical deskilling.
- When using any artificial intelligence tool, veterinarians must ensure continued compliance with jurisdictional regulations, obtain appropriate informed consent from the client, and maintain accountability, transparency, and human oversight.

INTRODUCTION

Computer systems that interpret or generate language are increasingly used in veterinary and human health care for client communication, medical record documentation, clinical decision support (CDS), and practice reviews. However, these tools raise concerns regarding data security, informed consent, transparency, bias mitigation, hallucination risks, and the need for continuous human oversight. Individual tools require additional precautions and considerations for prudent use. This article surveys tools that support veterinary clinical practice using artificial intelligence (AI) language models for language generation. Potential risks, considerations, and recommendations for each application are provided.

All authors contributed equally.
[a] Association for Veterinary Informatics, Madison, WI, USA; [b] Veterinary AI, Vetcove, Miami, FL, USA; [c] American Association of Veterinary State Boards, Overland Park, KS, USA
* Corresponding author. Association for Veterinary Informatics, 500 Westover Drive 7548, Sanford, NC 27330.
E-mail address: nbollig@avinformatics.org

Vet Clin Small Anim 56 (2026) 1087–1104
https://doi.org/10.1016/j.cvsm.2026.03.014 **vetsmall.theclinics.com**
0195-5616/26/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Abbreviations	
AI	Artificial Intelligence
CDS	Clinical Decision Support
LLM	Large Language Model
NLP	Natural Language Processing
RAG	Retrieval-Augmented Generation
SOAP	Subjective, Objective, Assessment, and Plan

BACKGROUND

Introduction to Language Models

Manipulation or generation of language by a computer is called *natural language processing (NLP)*, and the underlying technology for representing or processing language is called a *language model.*[1] While some of the earliest NLP systems employed symbolic grammars to represent language and mostly relied on pre-defined rules to emulate language understanding,[2–4] since the late 1980s methods have shifted away from pre-defined rules toward statistical methods such as *machine learning*, in which language understanding can be emulated by learning from a text dataset, that is, *corpus.*[5–7] Language models have since co-evolved in tandem with a steady increase in access to computing power and large text datasets.[8,9]

Current NLP models are primarily based on *neural networks*, which follow a structure analogous to biological neurons. Each *unit* (eg, neuron) integrates information from previous units, transforms the data, and passes a modified signal forward to subsequent units until an input is transformed into a final output.[10,11]

The *architecture* of a neural network describes how units within that network are connected to each other and determines how information flows through and is represented by the model. A neural network is often organized into "layers" of units, with information flowing successively from layer to layer. Arguably, the greatest innovations in language modeling have been new architectures that expand the range of information a model can represent. The *transformer* architecture published by Vaswani and colleagues in 2017 is a prominent driver of AI language model popularity.[12] This article focuses on *generative language models*, that is, models that generate language as their output, many of which use the transformer architecture.[13–15]

Large Language Models

A *large language model (LLM)* is a deep neural network of comparatively large size, which today roughly refers to a model with at least 100 billion model parameters (ie, connections between the units).[16] The size makes this technology largely dependent on companies who operate datacenters and offer LLM services via Internet platforms.[17]

LLMs are trained on billions or trillions of words of text with a singular goal: predict a word in a sentence based on the surrounding words (**Fig. 1**A). This training can be conducted on a massive scale by hiding words in the training corpus and having the model use the surrounding text to predict the missing word, ultimately creating models capable of emulating generalized, multimodal language understanding.[18] Typically, a user or software program will interact with an LLM by submitting a text query called a *prompt* and receiving the LLM's response as a body of text. The prompt flows through a large neural network, and each word of the response is selected probabilistically based on the model's prediction of what word will best come next in the sequence. Workflows in which the LLM's output is reviewed by a human prior to further action have a *human in the loop* (**Fig. 1**B). While users might interact directly

with commercially available LLM platforms, they are increasingly likely to experience LLMs indirectly through third-party software that collects data, formulates prompts, and sends queries to a commercially hosted LLM service (Fig. 1C).

Hallucinations

Clinicians should recognize that an LLM only *emulates* language understanding through statistical modeling. This process can also be biased through incorrect training information,[19] leading to concerns regarding its validity when applied in medical settings.[20]

Incorrect, nonsensical, or harmful results from an LLM are known as *hallucinations*. A wide range of distinct types of hallucinations are possible, including incorrect facts, fabricated facts, contextual inconsistencies, logical inconsistencies, and instruction inconsistencies.[21] See Table 1 for examples of hallucinations that an LLM could produce while creating a veterinary Subjective, Objective, Assessment, and Plan (SOAP) note.

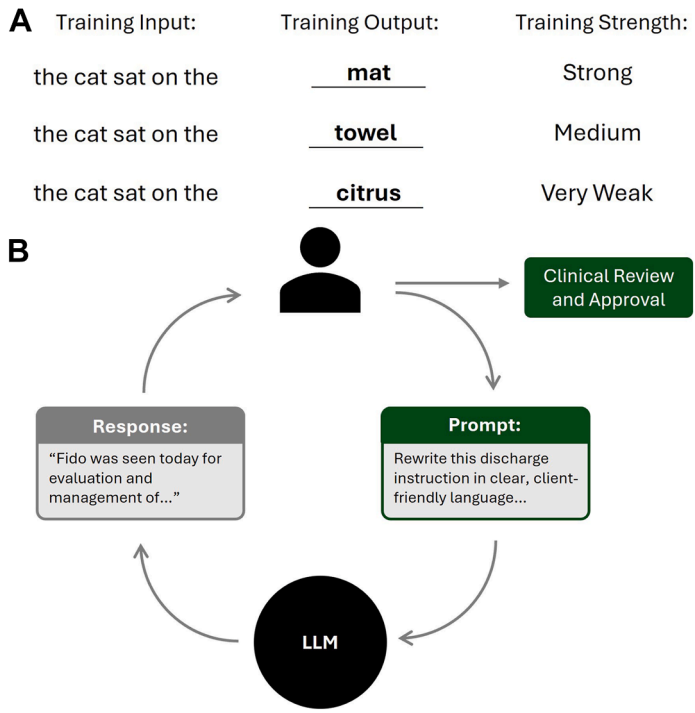


Fig. 1. Large Language Models (LLMs). (*A*) LLMs are trained to predict the next word or missing word in a sentence (training output), based on surrounding text (training input). Natural text is used to construct these training instances, providing a massive training set that reflects word usage across a vast array of contexts (training strength). (*B*) Users interact with an LLM by submitting a prompt and receiving a text response from the model, forming a human-in-the-loop AI system. (*C*) Users may interact with LLMs through third-party applications that submit prompts containing medical details over the internet. (*D*) The interaction between the user and the LLM may be managed by software that performs additional functions such as retrieval-augmented generation (RAG). This retrieves additional information (eg, medical records, public databases) and adds additional content to the LLM prompt.

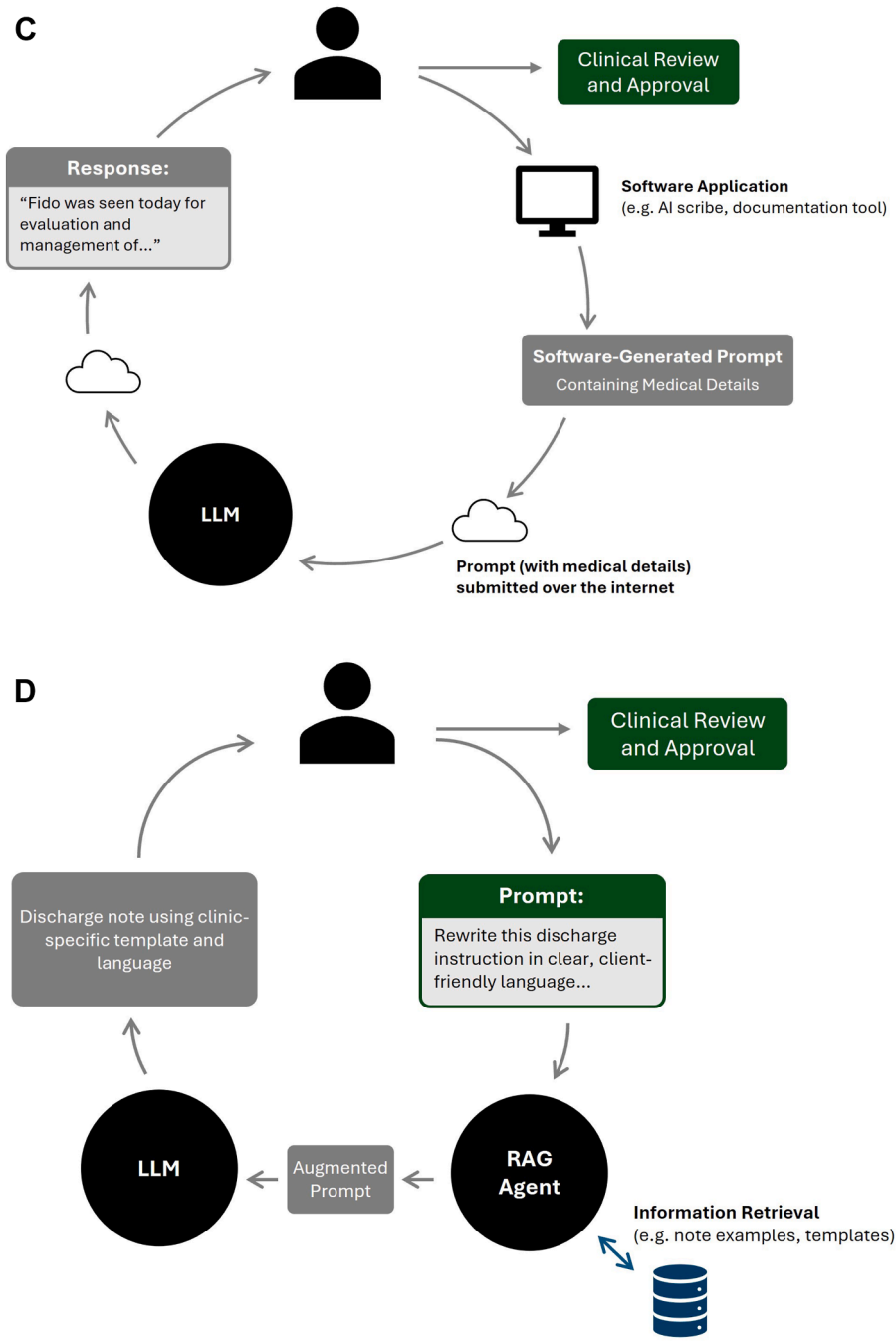


Fig. 1. (*continued*).

Table 1
Example hallucinations in generating a subjective, objective, assessment, and plan note

Category	Type	Example
Factuality Hallucination	Factual Contradiction	A medical conclusion that is not possible or appropriate after considering medical knowledge • For example, AI concludes "heartworm negative" despite positive SNAP test
	Factual Fabrication	A medical conclusion made with an exaggerated level of confidence or referring to a non-existent disease or other medical information • For example, AI cites an AVMA guideline that does not exist
Faithfulness Hallucination	Instruction Inconsistency	The SOAP note does not meet the instructed format • For example, AI generates client-facing language instead of a medical SOAP
	Context Inconsistency	A medical conclusion or omission inconsistent with clinical data provided to the LLM • For example, AI recommends NSAID therapy for a cat with documented chronic kidney disease
	Local Inconsistency	Information conflicts with other information provided in the output, such as Plan section not aligned with Assessment • For example, Assessment lists "suspected bacterial pyoderma" but Plan omits relevant therapy

This table illustrates example hallucinations that an LLM could produce when creating a veterinary Subjective, Objective, Assessment, and Plan (SOAP) note, following the taxonomy of Huang et al.[21]

A range of techniques have been created to leverage LLMs more effectively and reduce the likelihood of undesirable responses. After pre-training on a large text corpus, for example, to predict the next word in a sequence, a model can be *fine-tuned*, modifying the LLM parameters, by providing a specific set of (query, response) pairs relevant to a specific domain of knowledge or task, and this may reduce the frequency of hallucinations.[22,23]

Further improvement can be achieved by systems that employ *retrieval-augmented generation (RAG),*[24,25] processes where relevant information from outside sources is gathered and combined with the user's query to the LLM. When retrieval is used correctly, this supplies additional context, such as a patient's prior medical history, that would not have been available when training or fine-tuning the LLM prior to its deployment (**Fig. 1**D).

A range of opportunities exists to improve a model's output quality without modifying the model itself through *prompt engineering.*[26] This involves providing task-specific instructions or context with a query to guide the model's responses. One common form of prompt engineering, *few-shot prompting*, involves providing several complete examples of similar queries and desired outputs directly within the prompt itself.[27] Another common strategy is *chain-of-thought prompting*, which directs the LLM to follow a specific step-by-step reasoning process when formulating its response.[28] Numerous additional tactics are commonly employed in commercial LLM systems.[29]

The most important mitigation tactic for clinicians relying on LLM tools is *education*. LLMs should be approached with caution because any LLM can hallucinate at any time, even if the model has given correct responses to similar queries in the past. There are no universal guardrails preventing LLMs from producing incorrect or harmful responses. When misbehaving, LLMs often sound confident.[21] These concerns are especially salient in systems not rigorously and continually assessed in the clinical context.

DISCUSSION

This section reviews 4 common applications of language models: Client Communications, Medical Records, CDS, and Clinical Practice Review. This is followed by general professional and ethical guidance. Recommendations for the usage of AI language models in veterinary clinical practice are summarized in **Table 2**.

CLIENT COMMUNICATIONS

Drafting Communications

LLMs are commonly used to draft and review communications, such as post-visit summaries, discharge instructions, emails, and appointment reminders, using provided clinical background.[30,31] They can suggest clearer, more empathetic, or simplified phrasing for clients.[30] However, hallucination risks necessitate careful oversight.[32] Due to limited published evidence on veterinary-specific LLM performance, error profiles, and safety guardrails, their reliability in practice cannot yet be fully evaluated.

Automated Responses and Chatbots

More recently, AI-enabled chatbots have been deployed to answer routine client questions and assist with appointment scheduling.[30,33] In some scenarios, these advanced systems conduct preliminary symptom triage by guiding pet owners through structured question pathways.[34] While not yet broadly available in veterinary medicine, patients in human medicine can interact with an automated system, or chatbot, before receiving medical attention.[35] In veterinary medicine, published work and public evaluation of AI chatbots are limited. Despite the relative lack of evidence of accuracy, these tools are marketed for veterinary consumer-facing applications, where users can ask complex questions of their uploaded veterinary record.

Considerations and Recommendations

LLM applications improve efficiency and accessibility, but they carry risks. While medical condition summaries and discharge instructions warrant the most scrutiny, LLMs can insert errors into any communication. Users must review all client communications. Users should consider the impact an error, an incorrect tone, or a misunderstood linguistic style would have on their relationship with the client. Chatbots may provide incomplete information or inaccurate responses, oversimplify complex conditions, or provide outdated information.[36] If the Chatbot provides information that contradicts that of the veterinarian, this may lead to conflict or loss of trust in the veterinarian's professional judgment.[36]

MEDICAL RECORDS

Record Summarization

Because most veterinary records are unstructured and billing codes are inconsistent, there is growing interest in using NLP and LLMs to extract structured insights from clinical data.[37,38] LLMs can distill and summarize complex medical records and

Table 2
Recommendations for usage of generative language models in veterinary practice

Client Communications	• Apply disclaimers to clearly label AI-generated content • Use AI for drafting communications rather than final delivery • Do not use AI chatbots to deliver real-time medical advice to clients • Ensure all AI-generated content is reviewed by a human • Ensure private information transmitted or stored by the AI system is secure
Medical Records	• Require attending clinician review of all AI-generated material before inclusion into the medical record • Regularly audit AI tool outputs for accuracy and tone (eg, inferring diagnoses, editorializing) • Verbally narrate physical examination findings and key clinical information to ensure accurate documentation • Base the decision to retain or not retain recorded audio on applicable regulations and client consent • Ensure private information transmitted or stored by the AI system is secure • Establish informed consent appropriate to the application's risk level and all applicable laws
Clinical Decision Support (CDS)	• Use AI tools to augment, not replace, clinician professional judgment, and critically review all outputs • Document use of AI-based decision support tools in the medical record • Avoid using CDS tools to create first drafts; instead, use them to evaluate and/or supplement clinician-drafted content • Use applications trained and validated on veterinary data relevant to the intended species and scope of practice • Define procedures in advance for resolving conflicts between AI recommendations and clinical judgment (eg, seek an independent clinician's opinion) • Regularly audit AI tool outputs for accuracy • Ensure private information transmitted or stored by the AI system is secure
Clinical Practice Review	• Ensure that AI-based provider review is for professional development, not surveillance • Involve clinical and administrative staff in tool selection, configuration, and governance • Provide users with full access to their own results • Limit users to only accessing their own results • Require human review of all AI-generated results and supplement with human feedback before taking any action • Do not use AI-generated scores in compensation or disciplinary decisions • Ensure private information transmitted or stored within the AI system is secure

unstructured clinical narratives into structured outputs such as summaries or key data points.[39] A common use of this technology is to summarize referral or medical records from other clinics. However, multiple studies indicate that performance can be variable and requires oversight and review.[39,40] Another study demonstrated that ChatGPT accurately identified overweight body condition scores from free-text records with high precision.[41] There is increasing evidence that LLMs may be able to

identify diagnoses, treatments, and outcomes from medical records.[42,43] Despite the increasing use of LLMs in this manner, there is little published information available outside these specific use cases regarding their accuracy and hallucination rates.

Artificial Intelligence Scribes

Speech-to-text systems are increasingly used to transcribe veterinary notes or populate templates, particularly in radiology, using voice-trained or veterinary-specific language models.[44] However, performance varies based on microphone quality, recording environment, and ambient noise. Although specialized hardware such as unidirectional microphones and noise suppression can improve accuracy, the lack of standardized benchmarking prevents the establishment of definitive best practices.

Ambient microphones extend this technology by recording natural veterinarian–client conversations. While challenging, requiring accurate speaker differentiation, noise filtration, and strict adherence to privacy protocols, early adopters report significant benefits, including reduced documentation burden and fewer informational omissions.[44–48]

Medical Note Generation

Many LLM systems today can take text or audio inputs and convert them into medical notes. A common approach is to use an AI scribe to create a physical examination or SOAP note. However, it is also possible to use such tools to draft procedure notes or diagnostic reports, such as in pathology or radiology. One published demonstration showed accurate summarization and plan generation for a canine albuterol toxicosis scenario.[32] These systems draft medical notes using varying combinations of fine-tuning, prompt engineering, and RAG, with some of them integrating decision-support content from curated veterinary databases.[46,47] Their accuracy and hallucination rate depend on the clarity of the input, the context and details of the templates they use, and the availability of data to ground the LLM's response within factual information.[45,48–50] Hallucination risk remains a significant concern, especially within the Assessment and Plan sections of SOAP notes, which have direct clinical implications.[51]

Considerations and Recommendations

Although these uses may reduce time spent reading and writing medical records, their outputs must still be reviewed, edited for accuracy and completeness, and approved; veterinarians are the final medical and legal authority on the medical record. Clients must be informed of any ambient listening tools and provided with the option to opt out. Before adopting an LLM tool, veterinary managers should ask developers how models are trained and updated, and how to alter the medical record writing style or template.

Errors can arise from multiple steps in LLM-assisted medical records. Ambient listening systems may miss critical details if not spoken aloud, and if not corrected, errors can inform future outputs if the AI device builds on past output. Transcription accuracy may vary depending on the model used.[52] LLMs may produce outputs that result in overly verbose or inaccurate medical records, which must be revised or corrected. In addition, patient-specific context may be lost if the system lacks access to the patient's complete medical history,[53] increasing the risk of bias and the generation of inaccurate or inappropriate recommendations.

Individual regulatory boards may have differing determinations regarding whether the recording used to create the transcription is also considered part of the medical record. If so, it would be subject to the same data retention policies as written records.

As such, veterinary managers should ensure compliance with each board's requirements. AI systems should also disable voice data retention without explicit consent.

CLINICAL DECISION SUPPORT

Literature Retrieval and Evidence Summarization

AI systems for clinical information retrieval commonly use RAG to combine semantic search with controlled veterinary literature sources.[46,47] When a clinician poses a question, the system retrieves relevant passages from textbooks or journals and synthesizes them using an LLM. Because responses are constrained to vetted sources, hallucination rates may be lower than those seen in general-purpose LLMs. Reliability depends on the quality of the underlying database and the RAG method.[54]

Use of Artificial Intelligence for Education and Diagnostic Reasoning

Newer LLMs show improved performance on veterinary examinations, with GPT-4 achieving about 77% correctness versus 55% for GPT-3.5, and frequently including correct final diagnoses in differential lists.[32] Similar patterns are seen in human medicine, leading to training human clinicians to increasingly rely on LLMs for personalized education.[55] Despite this knowledge representation, LLMs lack true pathophysiologic reasoning and remain prone to overconfident, erroneous outputs that could mislead trainees. While general model performance is improving,[56] no published evaluations have assessed the latest models' impact on veterinary education.

Considerations and Recommendations

CDS tools are among the highest-risk applications of LLMs in medicine. Risks include hallucinations, lack of validation for veterinary-specific contexts, and deskilling of the veterinarian through loss of independent professional thought.[52,57–59] These tools are especially prone to hidden biases from inadequate or unaligned training data.[60] LLMs have been shown to struggle particularly with complex scenarios or those requiring special empathetic or ethical considerations,[61] and may inhibit the shared decision-making process between the veterinarian and the client. Many of these tools have been developed for human medicine and adapted for veterinary applications, and therefore lack real-world validation. Finally, there are no universal benchmarking data for evaluating the performance of individual tools in veterinary medicine.[62]

These risks may culminate in the deterioration of the veterinarian's professional judgment, resulting in substandard care. Rather than using these tools to create a first draft, the prudent clinician would complete their thought process and then run the medical record through the CDS tool to identify omissions, such as missed differential diagnoses.[52] These tools should be used to augment and support the veterinarian's professional judgment rather than replace it.

PROVIDER AND PRACTICE-LEVEL REVIEW

Language models can identify missed information and deviations from expected clinical patterns by detecting outliers in diagnostic frequency, treatment selection, or patient outcomes.[63–65] LLMs have also been used to assess adherence to the Consolidated Standards of Reporting Trials in clinical trials.[66] In veterinary medicine, applications primarily focus on herd health and epidemiology to identify early indicators of outbreaks.[67–71] Similar approaches could support companion-animal practice quality improvement by reviewing provider and overall practice adherence to standards, though human expertise is required, as statistical anomalies might reflect case-mix differences rather than actual quality issues.

Considerations and Recommendations

AI support for provider review and education should still be conservatively approached. These tools may, for example, attempt to rate a health care provider on their bedside manner; while developers can approximate empathic responses with markers such as time spent in "comfortable" silence,[72] AI cannot fully grasp all nuances. Audio-only LLM models may miss nonverbal cues, such as facial expressions and gestures, that can convey a different tone than voice alone.

These tools may also be used to grade a health provider's medical knowledge. However, this application is highly susceptible to bias, both from the training data used to develop it and from the developers' potential biases. For example, if the training data were biased toward more advanced, expensive diagnostic options, the provider may be penalized for recommending less advanced yet tailored treatment options for that specific client and patient context. Similarly, an AI tool may provide responses that reflect unexpected judgments about optimal care, such as along the quantity-versus-quality-of-life dimension.[73]

Given these concerns, provider review applications should not be taken at face value. A human, ideally another veterinarian, must be in the loop to provide a trustworthy evaluation of a human-to-human interaction.

PROFESSIONAL AND ETHICAL GUIDANCE

LLMs can reduce time spent on administrative tasks such as client communication, medical record creation, and CDS.[74] Physician studies demonstrate AI administrative assistance may increase time spent with patients, improve diagnostic accuracy, and support decision-making by reducing documentation burden.[52,75,76] Veterinarians may derive similar benefits.

Veterinarians should, however, weigh potential advantages against ethical and practical concerns.[77,78] When using any AI device, veterinarians must ensure continued compliance with jurisdictional regulations, obtain appropriate informed consent from the client, and maintain accountability, transparency, and human oversight.

Data Security and Privacy

Veterinarians must ensure compliance with federal and state or provincial laws and regulations requiring adequate data security and confidentiality.[79] These laws apply to all records, including paper records, electronic medical systems, and the use of all tools, including AI-based applications.

When choosing an AI system that may generate content for the medical record, several best practices are recommended (**Box 1**). The terms of use and privacy policies must be closely examined. Although veterinary medicine lacks uniform security standards, software may demonstrate commitment to data security through SOC 2 or ISO 27001 certifications, which involve external audits. Many current systems in the veterinary medical field use open AI systems that may use submitted data to retrain their models or evaluate system performance. There must be safeguards in place to prevent insecure transfer of data between the device and external parties, such as the company that developed the device or the open AI system. If data are to be transferred to or stored within servers accessible by third parties, all third parties should demonstrate adherence to data privacy and security standards, such as anonymization and encryption. Data security laws and regulations continue to evolve around AI. Users of AI should ensure a current and accurate understanding of the applicable laws within their country and jurisdiction.

Box 1
General recommendations for data security and privacy in clinical use of generative language models

Data Security and Privacy Recommendations
- Review terms of use and privacy policy for AI tools
- Understand how the tool stores, transmits, and protects private client data (eg, anonymization, encryption)
- Review available security certifications (eg, SOC 2, ISO 27001)
- Establish a clinic-level AI policy defining acceptable security standards
- Monitor access controls, breach detection, and security alerts[53]

Informed Consent

Informed consent is a critical regulatory requirement that achieves 2 goals. The first is supporting the veterinarian's ethical obligation to the client to ensure there is understanding and alignment in how care is delivered to the patient. The second is more practical: should the veterinarian's use of AI devices lead to a civil lawsuit or a regulatory board complaint, documentation and demonstration of obtaining informed client consent may help protect the veterinarian. However, it does not absolve the veterinarian from poor professional judgment.

Currently, there are no standardized guidelines for when and how to obtain informed client consent for the use of AI in veterinary medicine, although several factors should be considered (**Box 2**). Veterinarians must use their professional judgment to determine the appropriate level of consent for each application, client, and patient. The level of consent may vary depending on the application's risk. Low-risk administrative uses, such as appointment reminders, may not require explicit consent. For AI scribes and ambient listening devices, implied consent may be sufficient in some localities (eg, posting signage indicating the device's use and the option to opt out); however, veterinarians are advised to investigate state-specific laws regarding two-party consent for recording. Higher risk diagnostic or clinical applications should necessitate informing the client of the risks and benefits of using that device, followed by documentation of the client's consent.[77] Current laws, regulations, policies, and guidelines should be consulted to inform the level of expected consent.

To obtain informed consent, the veterinarian should be able to provide the client with basic knowledge of how the device arrives at the output, and disclose their experience with the tool.[80]

Trustworthiness

Bias can be introduced into the AI device during model development or by the veterinarian themselves. System performance varies widely depending on how the model was developed and whether its training data accurately represent the target clinical population.[80] Biases introduced by training methodology and data may be impossible

Box 2
General recommendations for informed consent in clinical use of generative language models

Informed Consent Recommendations
- Base consent requirements on a risk assessment of the tool, including data security, clinical impact, clinician knowledge of how the tool arrives at its output, clinician comfort, and applicable laws
- Provide clients with opt-out options for AI tools whenever feasible

Box 3
Factors associated with automation bias

Factors that increase the likelihood of automation bias include
- Increased difficulty of the task and heightened workload
- Insufficient experience of the veterinarian
- Increased ambiguity of AI algorithms
- Lack of awareness regarding the possibility or harmfulness of AI hallucinations
- Bias in training data[57,82]

for the veterinarian to fully evaluate and thus create uncertainty regarding the trustworthiness of AI-generated outputs. Bias can arise from training datasets that overrepresent or underrepresent certain species, breeds, or conditions, or from models optimized with human data without veterinary-specific validation. Studies in the medical field suggest that bias can be introduced if the AI device is trained only on one socioeconomic population,[60] or it may display bias toward a preferred course of treatment if it is trained on data from tertiary care hospitals rather than community care clinics. Solutions employed by developers of AI systems to reduce training set bias include dataset diversification, algorithmic fairness techniques, bias audits and evaluation metrics, synthetic data augmentation, post-deployment monitoring, stakeholder engagement, and ethical review.[52,60]

Automation bias occurs when users excessively trust AI outputs, bypassing professional judgment. This can lead to deskilling, medical errors of omission or commission, degraded care, and gradual deterioration of AI performance as incorrect outputs remain uncorrected.[57,81] Because automation bias is based on human behavior, the most critical step in correcting it is to know that everyone can experience it. **Box 3** lists factors that commonly increase the likelihood of automation bias.

Given the challenges affecting AI system trustworthiness, including bias and hallucinations, the critical importance of human review of all AI-generated output must be emphasized.[83]

Human in the Loop

Veterinarians are the final authority and ultimate responsible party for all clinical decisions and actions related to the practice of veterinary medicine, regardless of whether language models are employed. A human should actively review the output of any AI tool, each time it is used, prior to any decision or action. Importantly, that human must have the expertise required to detect and correct errors. This recommendation applies even to lower risk uses such as generating client emails or other communications.

SUMMARY

This article reviewed computer systems that support veterinary clinical practice using AI language models for language interpretation and generation. LLMs are statistical models that emulate generalized language understanding when responding to user-provided prompts.

AI tools can provide meaningful benefits in clinical settings but require careful use due to ethical and practical considerations. This article presented recommendations to support high-quality outcomes. Veterinarians should review the best practices for any tool, as, for example, the results of ambient AI scribes may vary widely depending on factors such as recording quality and explicit verbalization of findings. Veterinarians should thoughtfully review terms of use and privacy policies, solicit informed consent

at a level appropriate to the risk and all applicable laws, recognize human drivers of automation bias, and document their use of CDS tools in the medical record.

No current language model can be guaranteed to avoid incorrect, misleading, or harmful outputs; therefore, all AI-generated clinical content should be reviewed by appropriately authorized veterinary staff who would be qualified to perform the same task without AI.

CLINICS CARE POINTS

- In the clinic, LLMs have been used to support client communications, medical record transcription and generation, CDS, and clinical practice reviews.
- These systems can generate confident yet incorrect or inappropriate content, including clinically meaningful errors; all AI-generated material must be reviewed by a veterinarian before use.
- Anyone can experience automation bias, the tendency to over-trust automated systems, especially under certain conditions, such as heightened workload.
- When using any AI tool, veterinarians must ensure continued compliance with jurisdictional regulations, obtain appropriate informed consent from the client, and maintain accountability and transparency.
- Regardless of AI involvement, the veterinarian remains the final medical and legal authority for all clinical decisions and medical record content.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used GPT-5.2 from OpenAI and Claude Sonnet 4.6 from Anthropic to analyze and critique drafted language. After using this tool, the authors reviewed and edited article language and take full responsibility for the publication's content.

DISCLOSURES

The authors declare no financial conflicts of interest.

REFERENCES

1. Eisenstein J. Introduction to natural language processing. Cambridge (MA): MIT Press; 2019.
2. Chomsky N. Three models for the description of language. IEEE Trans Inf Theor 1956;2(3):113–24.
3. Hutchins WJ. The Georgetown-IBM experiment demonstrated in January 1954. In: Frederking RE, Taylor KB, editors. Machine translation: from real users to research. Berlin, Heidelberg: Springer; 2004. p. 102–14.
4. Rajaraman V. From ELIZA to ChatGPT. Resonance 2023;28(6):889–905.
5. Manning C, Schütze H. Foundations of statistical natural language processing. Cambridge (MA): MIT Press; 1999.
6. Berger AL, Brown PF, Della Pietra SA, et al. The candide system for machine translation. Plainsboro (NJ): Association for Computational Linguistics; 1994. p. 157–62.
7. Brown PF, Della Pietra VJ, Della Pietra SA, et al. The mathematics of statistical machine translation: parameter estimation. Comput Linguist 1993;19(2):263–311.

8. Mack CA. Fifty years of moore's law. IEEE Trans Semicond Manuf 2011;24(2):202–7.
9. De Luzi F. History of large language models. In: De Luzi F, Monti F, Mecella M, editors. Engineering information systems with large language models. Nature Switzerland: Springer; 2025. p. 13–31.
10. Krogh A. What are artificial neural networks? Nat Biotechnol 2008;26(2):195–7.
11. Mitchell TM. Machine learning. New York (NY): WCB/McGraw-Hill; 1997. p. 414.
12. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. presented at: 31st international conference on neural information processing systems. 2017. Long Beach, (CA). Available at: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.
13. Feuerriegel S, Hartmann J, Janiesch C, et al. Generative AI. Business & Information Systems Engineering 2024;66(1):111–26.
14. Nah F-H, Zheng R, Cai J, et al. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research 2023;25(3):277–304.
15. Bengesi S, El-Sayed H, Sarker MK, et al. Advancements in generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. IEEE Access 2024;12:69812–37.
16. Naveed H, Khan AU, Qiu S, et al. A comprehensive overview of large language models. ACM Trans Intell Syst Technol 2025;16(5):106–72.
17. Luitse D, Denkena W. The great transformer: examining the role of large language models in the political economy of AI. Big Data & Society 2021;8(2).
18. Wang T, Roberts A, Hesslow D, et al. What language model architecture and pretraining objective works best for zero-shot generalization? 2022:22964–22984. Available at: https://proceedings.mlr.press/v162/wang22u.html.
19. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digit Health 2024;3(11):e0000651.
20. Sallam M. ChatGPT utility in healthcare education, research, and practice. Healthcare 2023;11(6):887.
21. Huang L, Yu W, Ma W, et al. A survey on hallucination in large language models. ACM Trans Inf Syst 2025;43(2):42–55.
22. Zhang S, Dong L, Li X, et al. Instruction tuning for large language models: a survey. ACM Comput Surv 2025;58(7):1–36.
23. Hu Z, Wang L, Lan Y, et al. LLM-Adapters: an adapter family for parameter-efficient fine-tuning of large language models. 2023:5254–5276. Available at: https://aclanthology.org/2023.emnlp-main.319/.
24. Guu K, Lee K, Tung Z, et al. Retrieval augmented language model pre-training. 2020:3929–3938. Available at: https://proceedings.mlr.press/v119/guu20a.html.
25. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. 2020. Available at: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.
26. Sahoo P, Singh AK, Saha S, et al. A systematic survey of prompt engineering in large language models: techniques and applications. https://doi.org/10.48550/arXiv.2402.07927.
27. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. Adv Neural Inf Process Syst 2020;33:1877–901.
28. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. Adv Neural Inf Process Syst 2022;35:24824–37.

29. Tonmoy SMTI, Zaman SMM, Jain V, et al. A comprehensive survey of hallucination mitigation techniques in large language models. 2024. Avaailable at: http://arxiv.org/abs/2401.01313.
30. Tancredi W. The role of artificial intelligence chatbots in veterinary medicine. Today's veterinary practice. Gainesville (FL): North American Veterinary Community; 2023. p. 12–4.
31. Rumple S. Using ChatGPT to create efficiencies in your veterinary practice. IDEXX. 2025. Available at: https://www.practicelife.com/en/latest/using-chatgpt-to-create-efficiencies-in-your-veterinary-practice/. Accessed December 12, 2025.
32. Chu CP. ChatGPT in veterinary medicine: a practical guidance of generative artificial intelligence in clinics, education, and research. Front Vet Sci 2024;11:1395934.
33. Angela Beal D. The power of AI in veterinary client communication. 2025. Available at: https://software.idexx.com/resources/blog/the-power-of-ai-in-veterinary-clinics-client-communication.
34. Thompsett L. How Vet-AI is revolutionising pet care through prevention. InsurTech digital. 2025. Available at: https://insurtechdigital.com/news/how-vet-ai-is-revolutionising-pet-care-through-prevention. Accessed December 12, 2025.
35. Guo J, Ma J. Enhancing intelligent triage with large language models: a comprehensive evaluation and optimization study. Annu Int Conf IEEE Eng Med Biol Soc 2025;2025:1–6.
36. Pal A, Wangmo T, Bharadia T, et al. Generative AI/LLMs for plain language medical information for patients, caregivers and general public: opportunities, risks and ethics. Patient Prefer Adherence 2025;19. https://doi.org/10.2147/PPA.S527922.
37. Lustgarten JL, Zehnder A, Shipman W, et al. Veterinary informatics: forging the future between veterinary medicine, human medicine, and one health initiatives. JAMIA Open 2020;3(2):306–17.
38. Trice C, Donnelly AL, Ackerman L, et al. Medical management considerations. In: Ackerman L, editor. Pet-Specific Care for theVeterinary Team. 2021. Available at: https://doi.org/10.1002/9781119540687.ch9.
39. Wulcan JM, Jacques KL, Lee MA, et al. Classification performance and reproducibility of GPT-4 omni for information extraction from veterinary electronic health records. Front Vet Sci 2024;11:1490030.
40. Shroff H, Shankar A, Baron A, et al. A large language model assistant for summarizing hepatology referral documents. Am J Gastroenterol 2026. https://doi.org/10.14309/ajg.0000000000003905.
41. Fins IS, Davies H, Farrell S, et al. Evaluating ChatGPT text mining of clinical records for companion animal obesity monitoring. Vet Rec 2024;194(3):e3669.
42. Shah SV. Accuracy, consistency, and hallucination of large language models when analyzing unstructured clinical notes. JAMA Netw Open 2024;7(8).
43. Jiang Y, Irvin JA, Ng AY, et al. VetLLM: large language model for predicting diagnosis from veterinary notes. Pac Symp Biocomput 2024;29:120–33.
44. Walker S. Using veterinary dictation software: benefits and challenges. 2024. Available at: https://software.idexx.com/resources/blog/using-veterinary-dictation-software-benefits-and-challenges.
45. Palm E, Manikantan A, Mahal H, et al. Assessing the quality of AI-generated clinical notes: validated evaluation of a large language model ambient scribe. Front Artif Intell 2025;8:1691499.

46. DVM360Staff. Two companies are integrating AI-based tools for enhanced SOAP notes. DVM 360. Available at: https://www.dvm360.com/view/two-companies-are-integrating-ai-based-tools-for-enhanced-soap-notes. Accessed December 12, 2025.
47. CoVet launches consult CoVet featuring LifeLearn sofie AI integration. CoVet; 11/10/2025. 2025. Available at: https://www.co.vet/post/covet-launches-consult-covet-featuring-lifelearn-sofie-ai-integration-for-instant-clinical-decisio.
48. Ma SP, Liang AS, Shah SJ, et al. Ambient artificial intelligence scribes: utilization and impact on documentation time. J Am Med Inform Assoc 2025;32(2):381–5.
49. Olson KD, Meeker D, Troup M, et al. Use of ambient AI scribes to reduce administrative burden and professional burnout. JAMA Netw Open 2025;8(10):e2534976.
50. Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. J Am Med Inform Assoc 2025;32(2):375–80.
51. Hatem R, Simmons B, Thornton JE. A call to address AI "Hallucinations" and how healthcare professionals can mitigate their risks. Cureus 2023;15(9):e44720.
52. Rabbani SA, El-Tanani M, Sharma S, et al. Generative artificial intelligence in healthcare: applications, implementation challenges, and future directions. BioMedInformatics 2025;5(3). https://doi.org/10.3390/biomedinformatics5030037.
53. Fareed M, Fatima M, Uddin J, et al. A systematic review of ethical considerations of large language models in healthcare and medicine. Front Digit Health 2025;7doi. https://doi.org/10.3389/fdgth.2025.1653631.
54. Xu S, Yan Z, Dai C, et al. MEGA-RAG: a retrieval-augmented generation framework. Front Public Health 2025;13.
55. Zhui L, Yhap N, Liping L, et al. Impact of large language models on medical education and teaching adaptations. JMIR Med Inform 2024;12:e55933.
56. Zorpette G. Large language models are improving exponentially. IEEE Spectrum 2025. Available at: https://spectrum.ieee.org/large-language-model-performance. Accessed December 10, 2025.
57. Sujan M, Furniss D, Grundy K, et al. Human factors challenges for the safe use of artificial intelligence in patient care. BMJ Health & Care Informatics 2019;26(1). https://doi.org/10.1136/bmjhci-2019-100081.
58. Coghlan S, Quinn T. Ethics of using artificial intelligence (AI) in veterinary medicine. AI Soc 2024;39(5):2337–48.
59. Appleby RB, Basran PS. Artificial intelligence in veterinary medicine. J Am Vet Med Assoc 2022;260(8):819–24.
60. Levartovsky A, Omar M, Nadkarni GN, et al. Sociodemographic bias in large language model–assisted gastroenterology. JAMA Netw Open 2025;8(9). https://doi.org/10.1001/jamanetworkopen.2025.32692.
61. Li J, Zhou Z, Lyu H, et al. Large language models-powered clinical decision support: enhancing or replacing human expertise? Intelligent Medicine 2025;5(1). https://doi.org/10.1016/j.imed.2025.01.001.
62. Bouchemla F, Akchurin SV, Akchurina IV, et al. Artificial intelligence feasibility in veterinary medicine: a systematic review. Vet World 2023;16(10):2143–9.
63. Zhou X, Cohen E, Zhou X, et al. Leveraging LLM to identify missed information in patient-physician communication: improving healthcare service quality. Front Med (Lausanne) 2025;12:1631565.
64. Abacha AB, Yim W-W, Fu Y, et al. MEDEC: a benchmark for medical error detection and correction in clinical notes. presented at: findings of the Association for

computational linguistics. Vienna, Austria: ACL; 2025. Available at: https://aclanthology.org/2025.findings-acl.1159/.
65. Boussina A, Krishnamoorthy R, Quintero K, et al. Large language models for more efficient reporting of hospital quality measures. NEJM AI 2024;1(11). https://doi.org/10.1056/aics2400420.
66. Srinivasan A, Berkowitz J, Friedrich NA, et al. Large language model analysis of reporting quality of randomized clinical trial articles: a systematic review. JAMA Netw Open 2025;8(8):e2529418.
67. Davies H, Nenadic G, Alfattni G, et al. Text mining for disease surveillance in veterinary clinical data: part one, the language of veterinary clinical records and searching for words. Front Vet Sci 2024;11:1352239.
68. Davies H, Nenadic G, Alfattni G, et al. Text mining for disease surveillance in veterinary clinical data: part two, training computers to identify features in clinical text. Front Vet Sci 2024;11:1352726.
69. Mukherjee D, Sagar K, Kobialka RM, et al. Filling the gap: artificial intelligence-driven one health integration to strengthen pandemic preparedness in resource-limited settings. Front Public Health 2025;13:1707306.
70. Lagua EB, Mun HS, Ampode KMB, et al. Artificial intelligence for automatic monitoring of respiratory health conditions in smart swine farming. Animals (Basel) 2023;13(11). https://doi.org/10.3390/ani13111860.
71. Wen H, Shi H, Yu J, et al. An intelligent diagnostic method for porcine gastrointestinal infectious diseases based on multimodal AI and large language model. Front Vet Sci 2025;12:1660745.
72. Bedmutha MS, Bascom E, Sladek KR, et al. Artificial intelligence-generated feedback on social signals in patient-provider communication: technical performance, feedback usability, and impact. JAMIA Open 2024;7(4):ooae106.
73. Zohny H, Allen JW, Wilkinson D, et al. Which AI doctor would you like to see? Emulating healthcare provider-patient communication models with GPT-4: proof-of-concept and ethical exploration. J Med Ethics 2025. https://doi.org/10.1136/jme-2024-110256.
74. Albergante L, O'Flynn C, Meyer GD. Artificial intelligence is beginning to create value for selected small animal veterinary applications while remaining immature for others. J Am Vet Med Assoc 2025;263(3). https://doi.org/10.2460/javma.24.09.0617.
75. Goh E. Large language model influence on diagnostic reasoning. JAMA Intern Med 2024;184(10):1138–45.
76. Tierney AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catalyst 2024;5(3).
77. American Association of Veterinary State Boards. Regulatory considerations of the use of artificial intelligence in veterinary medicine. 2025. Available at: https://www.aavsb.org/wp-content/uploads/2025/08/AAVSB-AI-Guidance-Whitepaper.pdf. Accessed January 13, 2026.
78. British Veterinary Association. BVA policy position on artificial intelligence in the veterinary profession. 2025. Available at: https://www.bva.co.uk/media/6777/bva-position-on-artificial-intelligence.pdf. Accessed January 13, 2026.
79. Babcock SL, Pfeiffer C. Laws and regulations concerning the confidentiality of veterinarian-client communication. J Am Vet Med Assoc 2006;229(3):365–9.
80. Poore TJ, Pinard CJ, Shabbir A, et al., Context matters: comparison of commercial large language tools in, veterinary medicine, 2025. Available at: https://doi.org/10.48550/arXiv.2510.01224.

81. Abdelwanis M, Alarafati HK, Tammam MMS, et al. Exploring the risks of automation bias in healthcare artificial intelligence applications: a bowtie analysis. J Saf Sci Resil 2024;5(4). https://doi.org/10.1016/j.jnlssr.2024.06.001.
82. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inf Decis Making 2023;23(1). https://doi.org/10.1186/s12911-023-02103-9.
83. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearbook of Medical Informatics 2019;28(1):128–34.