Disease and Health Surveillance in Companion Animals Using Artificial Intelligence and Machine Learning

Peter-John Mäntylä Noble, BSc, BVM&S, PhD[a,*],
Sean Oliver Farrell, BSc, PhD[b]

KEYWORDS

• Electronic health records • Disease surveillance • Language models

KEY POINTS

- Large clinical record repositories and new text-processing methods present exciting opportunities for veterinary health and disease surveillance.
- Methods ranging from word-search to transformer-based language models such as PetBERT and VetBERT enable efficient automated disease and syndrome coding.
- Domain-specific pretraining of language models using veterinary text substantially improves performance over general-purpose models.
- Critical gaps remain in model generalization, standardized evaluation frameworks, and privacy-preserving deidentification methods for veterinary records.
- Generative models offer interesting opportunities but with potential environmental costs.

BACKGROUND

Historically, disease and health surveillance in companion animals has, understandably, taken second place to public health surveillance and closely linked food animal disease surveillance. Statutory surveillance in the latter areas has been well-funded and deeply embedded in the medical and food animal production industries and their regulatory bodies. In companion animals, disease and health surveillance have largely been left to academia and specific industries such as insurance, pharma, and pet food manufacturers.

[a] Department of Small Animal Clinical Science, University of Liverpool, Leahurst Campus, Chester High Road, Neston, Wirral, UK; [b] School of Medicine, Stanford University, Stanford, CA, USA
* Correspondng author.
E-mail address: rtnorle@liverpool.ac.uk

Vet Clin Small Anim 56 (2026) 1151–1164
https://doi.org/10.1016/j.cvsm.2026.03.016
vetsmall.theclinics.com
0195-5616/26/© 2026 Published by Elsevier Inc.

Abbreviations	
AI	artificial intelligence
CNN	convolutional neural networks
LDA	latent dirichlet allocation
LSTM	long short-term memory
NLP	natural language processing
RNN	recurrent neural networks
VEHR	veterinary electronic health records

Pet populations are burgeoning with 60% of UK households (17.2 million) owning pets,[1] creating a significant contribution to the economy (estimated at over £20 billion annually in the United Kingdom).[2] In the United States, 71% of households (94 million) own pets[3] contributing $312 billion to the economy in 2024.[4] Additionally, the recent pandemic illustrated the anthroponotic transmission of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) virus infecting dogs and cats.[5,6] Furthermore, as zoonotic diseases and antimicrobial resistance emerge as twin threats to global health security, it becomes critical to understand disease patterns and resistance mechanisms at the human-animal interface.[7–11] Veterinary electronic health records (vEHR) have also been used to support pharmacoepidemiology through monitoring of reporting of adverse events, improving drug safety and informing evidence-based prescribing practices across diverse animal populations.[12,13] Finally, the regular use of drugs that may have environmental impact (such as parasiticides) requires an understanding of disease prevalence to appropriately manage use of these drugs highlighting the key One Health perspective[14] that might be informed by companion animal disease surveillance.

Given resource constraints in companion animal surveillance, there is a clear need to optimize any approaches to minimize time and therefore cost for this task. Two relatively recent developments have the potential to revolutionize and accelerate companion animal health and disease surveillance. Firstly, the move by almost all veterinary practices and laboratories to vEHRs enables ready collation of records at scale, creating large analyzable datasets.[15–18] Secondly, the massive reduction in cost of computer power allows the use of a variety of artificial intelligence (AI) tools for data analysis. However, the predominance of free-text clinical narratives in vEHRs presents unique challenges, including nonstandardized terminology, inconsistent acronym use, typographical errors, negations, and context-dependent language. These characteristics necessitate sophisticated natural language processing (NLP) approaches to extract meaningful surveillance signals from unstructured clinical documentation.[19,20]

The use of large scale vEHR collections and AI to augment disease and health surveillance provides a unique opportunity for practicing veterinary surgeons to both enable research by contributing to large scale data collation initiatives and to benefit from the population-level epidemiologic insights and early outbreak warnings that these systems generate by analyzing millions of records simultaneously. In the following sections, we will focus on challenges associated with analyzing vEHR and how AI methods are accelerating this task.

UNIQUE CHALLENGES OF VETERINARY CLINICAL TEXT

As mentioned earlier, clinical free text can suffer from poor punctuation, context-specific and overloaded acronyms and negation and word truncation. **Fig. 1** presents an example of a free-text vEHRs, highlighting both the richness of detail and the potential challenges involved in extracting structured information. The narrative contains

URT signs. Started about 3 weeks ago. not off his food especially. Goes out so no history re U/F. Owner noticed noisy breathing especially in the evenings when he is sleeping. she played a recording - sounds nasal. PE: Mild weight loss. not keen on oral inspection - colour good, most of his teeth OK but inflamed gingiva UR arcade - possible neck lesion - they've noticed halistosis - not obvious in consult. Colour good. Hr = 180, lungs sound quiet - ? chest too quiet? URT noise. SMLNs normal. Abdo nAD. PLAN:Treat as URT infection. Unable tp tablet so given convenia injection. also bisolvon and steam inhalations. R/V 2 weeks, sooner if concer. INI will need further investigation - bloods/Xrays etc.

Fig. 1. An example of a free-text vEHR, illustrating the high level of clinical detail and the inherent difficulty in extracting structured information.

domain specific abbreviations (*URT*, *U/F*, *PE*, *UR*, *Hr*, *SMLN*, *nAD*, and *INI*) including abbreviations that are used with different meanings in other records (*R/V* here is revisit but in other records would be referring vet), poor punctuation and grammar, poor spelling (*halistosis*, *tp*, *concer*), as well as negated findings (*not off his food*, *no history re U/F*). Until recently the limited availability of NLP tools in veterinary medicine very much limited record analysis at scale. When combined with data siloing and lack of data standards, the landscape for companion animal disease surveillance was challenging.[21]

APPROACHES TO CLINICAL TEXT ANALYSIS

The extraction of structured information from veterinary clinical narratives has evolved through 3 distinct methodological paradigms, each offering different trade-offs between interpretability, scalability, and performance (**Table 1**).

Rule-based approaches employ manually crafted patterns such as regular expressions or keyword dictionaries to identify clinical entities.[22] *Learning-based approaches using traditional neural architectures* (convolutional neural networks [CNNs], recurrent neural networks [RNNs], long short-term memory networks [LSTMs]) instead learn text meaning from training data, using static word embeddings (high-dimensional vectors that encode word meaning created with methods like Word2Vec[23]) to capture document semantics. *Transformers*, complex neural network models, fundamentally altered clinical NLP employing self-attention mechanisms. These compute relationships between all words in a document simultaneously rather than sequentially.[24] Transformers generate contextualized embeddings where word representations adapt dynamically based on surrounding text. This contextual sensitivity proves essential in clinical text: *discharge* in *the wound had a nasty discharge* refers to suppuration, while *discharge the patient* refers to release from care—distinctions that static embeddings cannot capture.

Transformers are now commonly deployed as encoder-only models (eg, Bidirectional Encoder Representations from Transformers—BERT,[25] typically 100 to 350 million parameters) optimized for text understanding, or decoder-only generative models (eg, GPT[26] and LLaMA[27]), often billions of parameters in size that generate text based on a prompt. The pretraining-fine-tuning paradigm, first demonstrated in ULMFiT,[28] allows encoder models pretrained on millions of unlabeled documents to capture language meaning to be subsequently fine-tuned to specific tasks such as disease coding using much smaller labeled datasets[25,29] This approach is particularly valuable in veterinary medicine where expert-annotated data remains scarce.[21]

Table 1
Comparison of methodological approaches to veterinary clinical text analysis

Approach	Benefits	Limitations	Veterinary Examples
Rule-based (regex/ keyword)	Highly interpretable; no training data required; fast to deploy for narrow tasks	Brittle to spelling variation and novel terminology; requires domain expertise; manual evaluation reduces useable scale from millions to thousands of records	Myxomatosis[32]; flea/tick surveillance[33]
Traditional neural architectures (CNN, RNN, and LSTM)	Data-driven generalization beyond exact pattern matches; learns semantic relationships via word embeddings	Sequential processing loses long-range dependencies; static word vectors conflate multiple word meanings	DeepTag[35]; FasTag[69]
Encoder transformers (BERT-based)	Contextualized embeddings; captures long-range dependencies; efficient fine-tuning on small labeled datasets; lower compute than generative models	Requires pretraining data and compute; performance degrades on out-of-distribution data	VetTag[36]; PetBERT[40]; VetBERT[39]
Generative models (GPT, LLaMA, etc)	Few/zero-shot capability; flexible natural language interaction; large context windows accommodate clinical dictionaries	High computational and environmental cost; may underperform domain-adapted encoders on well-defined tasks	GPT-4 clinical extraction[44]; SynSUM[45]

The subsequent sections examine how these methodological approaches have been applied to specific challenges in health monitoring and disease surveillance.

DISEASE CODING AND VETERINARY ELECTRONIC HEALTH RECORDS

The coding or annotation of records for diseases is fundamental to enabling effective epidemiologic surveillance and large-scale research.[20] Early work has explored keyword-based searching as a tool for extracting specific conditions from clinical narratives, such as in the study of diabetes mellitus in dogs.[30] However, the effectiveness of keyword approaches is undermined by the frequent presence of negations in first-opinion clinical texts (eg, *not cancer*), with studies showing that negated terms appear in approximately 11% of veterinary free-text sentences.[31] This limitation led to the development of complex rule-based systems, which have found utility in investigations into diseases like myxomatosis,[32] flea and tick activity.[33,34] However, rule-based frameworks are inherently limited by assumptions of their designers regarding what word patterns will be present, and their performance may degrade as the language used in EHR evolves.

One of the pioneering efforts in automated veterinary disease coding using neural network architectures was DeepTag, a bidirectional LSTM trained to predict 42 top-

level SNOMED-CT diagnosis codes.[35] DeepTag's limitation to a relatively small set of broad diagnostic categories meant it could not be directly applied to predict the thousands of fine-grained codes that could comprise veterinary disease classification.

Building upon DeepTag's foundation, VetTag (an encoder-based transformer model), represented a significant advancement by extending diagnostic capability to 4577 fine-grained SNOMED-CT codes.[36] The system employed a 2-stage training approach: unsupervised pretraining on over 1 million unlabeled notes from a private specialty veterinary group, followed by supervised fine-tuning on 112,557 expert-labeled notes from Colorado State University. This training strategy enabled VetTag to leverage abundant unlabeled clinical text while requiring only a moderate amount of expensive expert annotations.

Despite these advances, both DeepTag and VetTag were trained and tested on data from the same contributing hospitals, leaving uncertainty about their performance on out-of-distribution veterinary data, such as primary care vEHRs. Large language models (LLMs) address the critical need for models that can generalize across diverse clinical contexts, from tertiary referral centers to primary care clinics.

BERT-BASED LANGUAGE MODELS AND VETERINARY ELECTRONIC HEALTH RECORDS

The pretraining and fine-tuning paradigm has enabled excellent performance of transformers in tasks using medical text. Many domain-specific BERT models have been created. For example, BioBERT, trained on PubMed abstracts and PubMed Central full-text articles[37] and BioClinicalBERT, trained on diverse clinical corpora including clinical notes.[38] The first vEHR adapted BERT-model, VetBERT, was derived from ClinicalBERT with further pretraining on approximately 15 million clinical notes from the VetCompass Australia corpus. This adaptation enabled a 12% improvement over BERT-base in classifying antimicrobial prescription rationales, underscoring the importance of aligning pretraining data with domain-specific linguistic and clinical characteristics.[39] More recently, PetBERT,[40] a BERT-based model trained on over 500 million tokens (words and word-pieces) from the Small Animal Veterinary Surveillance Network (SAVSNET) corpus,[41] achieved state-of-the-art performance on the case study of disease outbreak detection from first opinion veterinary clinical records.

Recent comparative analyses of clinical language models across vEHRs have demonstrated that domain-specific models, such as PetBERT and VetBERT, can either outperform human clinical models of similar architectural complexity or achieve comparable performance to substantially larger models.[42] These improvements demonstrate the value of domain-adapting language models for specific veterinary contexts (first opinion vs referral, species-specific applications, among others).

GENERATIVE LANGUAGE MODELS IN VETERINARY CLINICAL TEXT

While encoder-only transformers like BERT excel at understanding and classifying text, generative language models introduce fundamentally different capabilities for veterinary informatics. These decoder-based architectures, exemplified by generative pretrained transformers (GPTs) and large language model meta AI (LLama), and commercial systems like Claude and Gemini, are trained to generate coherent text by predicting subsequent tokens.

Generative language model prompts can now include very large spans of text (context window), which can include pertinent datasets such as dictionaries of diseases and clinical signs that enable generation of useful record annotations and clinical insight.[43]

Large generative models demonstrate impressive few-shot and zero-shot capabilities (inclusion of one or a few examples of desired output as part of their prompt), potentially reducing the need for task-specific fine-tuning. Preliminary applications in veterinary medicine include using general-purpose models like GPT-4 for extracting structured clinical information from free-text notes[44] and generating synthetic veterinary clinical data for model development.[45]

For operational veterinary surveillance systems performing well-defined repetitive tasks-such as syndromic classification or disease coding-smaller domain-adapted encoder models like PetBERT often achieve comparable or superior performance to large generative models while consuming dramatically less computational resources.[40,42]

VETERINARY ELECTRONIC HEALTH RECORDS FOR SYNDROMIC SURVEILLANCE

Syndromic surveillance uses prediagnostic health data to detect potential disease outbreaks.[46] This approach is particularly valuable in veterinary medicine, where fewer than 29% of health problems are followed by diagnostic testing,[47] meaning surveillance relying on laboratory reports samples only a highly selective subset of clinical cases. Syndromes (eg, respiratory signs) appear in clinical notes before definitive diagnoses, making them valuable targets for early outbreak detection.[48,49] The 2007 melamine-adulterated pet food outbreak demonstrated this potential: retrospective analysis showed a novel syndromic surveillance system[50] could have enabled earlier food recall, though clinical signs did not trigger alerts as rapidly as laboratory data changes.[51]

The challenge lies in accurately identifying syndromes given that text-terms like *coughing* or *nasal discharge* are far more varied than specific labels like *pneumonia* Transformer-based language models offer a solution through their ability to capture contextual meaning. PetBERT-ICD, based on the PetBERT model and fine-tuned on the ICD-11 hierarchy, successfully identified the 2020 UK canine enteric disease outbreak[52] with approximately 3 weeks' lead time over traditional clinician-assigned coding, demonstrating that scalable, accurate syndromic classification from free text can unlock vital early warning capabilities for animal health surveillance.

UNSUPERVISED DISCOVERY OF CLINICAL PATTERNS THROUGH TOPIC MODELING

While supervised approaches like disease coding and syndromic surveillance require labeled training data, fully unsupervised methods can discover latent patterns in clinical text without prior annotation. Topic modeling algorithms, including latent dirichlet allocation (LDA),[53] automatically identify clusters of co-occurring words that represent coherent *topics* across large document collections. In clinical contexts, discovered topics may correspond to diseases, syndromes or treatment protocols.[54]

The primary advantage of topic modeling is that it does not require expensive manual annotation and can identify patterns not anticipated by researchers. However, topic quality depends on careful configuration of the model. Additionally, discovered topics may not always correspond to clinically meaningful categories, and in the case of LDA, the bag-of-words assumption ignores word order and context.[53]

In the veterinary context, a study applying LDA to 200,000 vEHR from the SAVSNET database demonstrated that discovered topics aligned with clinical syndromes and exposed a temporal pattern in one topic (gastroenteric signs) that corresponded to the national outbreak of gastroenteric disease in 2020.[55]

Recent advances have extended classical topic modeling through neural architectures. For instance, a transformer based approach, BERTopic, leverages contextual embeddings to discover topics that account for word context.[56] This addresses

some limitations of classical LDA, particularly regarding contextual word meaning, though it introduces additional computational costs.

A SAVSNET study applying BERTopic topic modeling to vEHR from 1 million dogs, demonstrated population-wide exploration of health-related phenotypes[57] enabling evaluation of breed, age, sex, and temporal trends. This demonstrated that topic modeling can scale to large veterinary datasets while maintaining interpretability.

Topic modeling serves as a complementary tool to supervised methods in veterinary clinical informatics. It is particularly suited for exploratory analysis when the full range of clinical presentations is unknown; for initial characterization of large EHR datasets; and for hypothesis generation to inform subsequent focused studies.

PRIVACY, ETHICS, AND GOVERNANCE IN VETERINARY NATURAL LANGUAGE PROCESSING

Analysis of unstructured free-text data is constrained by the challenge of balancing data utility with patient privacy protection.[58,59] While structured medical data can be relatively straightforward to deidentify through the removal or masking of specific fields, unstructured clinical notes present unique challenges due to their narrative nature.[60] These free-text records may contain indirect identifiers embedded within complex clinical narratives, making traditional anonymization approaches insufficient for ensuring patient confidentiality while preserving data utility.

While traditional rule-based approaches to anonymization have been widely implemented, learning-based methods have emerged as a more sophisticated solution, leveraging machine learning[61,62] and deep learning techniques[63–67] to create scalable deidentification pipelines. These automated approaches offer promising alternatives to the resource-intensive manual anonymisation process, demonstrating impressive performance in controlled settings. However, their practical implementation faces a challenge: performance instability when encountering domain shifts between training data and real-world applications.[63,65] This limitation is particularly problematic in health care environments, where data heterogeneity is inherent due to variations across medical institutions, specialities, and geographic regions. Most notably, these models are typically developed and evaluated using limited datasets that fail to capture the full spectrum of variations present in EHRs. As a result, when deployed on out-of-distribution EHRs data, these systems often experience substantial degradation in performance.[63,64,67,68]

There currently is no specific standard for how vEHRs should be used and shared. Although animal data is not classified as sensitive under US, UK, or EU regulations on medical information, veterinary clinical notes can contain identifiable information, such as owner or veterinarian names, addresses, and contact details. Furthermore, animal-specific data, such as unique pet names combined with geospatial details, could potentially reveal the identity of the owner. Importantly, as private businesses, veterinary practices may record proprietary data or insights into client behavior that, if mishandled, could lead to unfair competition or reputational harm. Consequently, careful attention must be paid to record deidentification not only for privacy protection but also in order that owners and practitioners continue to contribute to data sharing initiatives without fear of negative repercussions.

EVALUATION FRAMEWORKS IN CLINICAL NATURAL LANGUAGE PROCESSING

The evaluation of clinical NLP models is essential for building confidence in health care AI systems and uncovering potential biases in model development; the sensitive nature of clinical data creates substantial barriers to open validation practices.

The difficulty in sharing clinical records due to privacy concerns has led researchers to explore alternative approaches, including the generation of synthetic datasets for medical research, which theoretically enables safe data sharing without reidentification risks.[69,70] However, models trained on synthetic data are susceptible to model collapse (where observed degradation in model performance occurs because it overfits to synthetic patterns, leading to reduced generalization and diversity in outputs),[71] and validation performed exclusively on synthetic datasets lacks the credibility of real-world data validation.[72]

In contrast to synthetic approaches, the MIMIC datasets represent a gold standard for clinical data sharing, demonstrating how careful deidentification and proper governance can enable widespread research use.[73,74] MIMIC-IV encompasses data from over 65,000 ICU patients and more than 200,000 emergency department patients from Beth Israel Deaconess Medical Center, with the combined MIMIC-III and MIMIC-IV datasets generating over 10,000 citations, illustrating the transformative impact of accessible clinical datasets on health care research.

The SAVSNET group introduced PetEVAL,[75] an open evaluation dataset comprising approximately 17,000 vEHRs annotated across 3 core natural language processing tasks: anonymization via named entity recognition (NER), disease extraction via NER, and syndromic surveillance framed as a multi-class classification problem. To support open release while maintaining confidentiality, each record underwent independent review by multiple annotators to verify the absence of residual identifiers, addressing a critical barrier to sharing clinical free-text data. The anonymization task enables systematic benchmarking of NER models designed to redact sensitive information from first-opinion veterinary records, while the disease extraction task facilitates scalable codification of free-text consultations into structured clinical concepts. The syndromic surveillance component further assigns consultations to the 20 top-level chapters of the ICD-11 hierarchy, enabling population-level monitoring and downstream epidemiologic analyses. PetEVAL is freely available and explicitly positioned as a community benchmark, with the authors encouraging its use as a baseline for future model development, comparative evaluation, and the extension of the dataset to additional tasks. Nevertheless, while PetEVAL represents a substantial step toward standardized and transparent evaluation in veterinary NLP, there remains a clear need for further vEHR datasets to be released by independent groups to support external validation, improve generalizability, and reduce overreliance on single-source benchmarks.

ENVIRONMENTAL CONSIDERATIONS: LARGE LANGUAGE MODELS VERSUS SMALLER TRANSFORMERS

Generative LLM and tools based on these have environmental impact. The models use substantive compute using large amounts of energy to both produce them and to use them for inference. Recent estimates suggest that training large models can consume energy equivalent to the lifetime emissions of multiple automobiles.[76,77] Additionally, data center cooling requirements require substantial water consumption, with recent analyses indicating that training a single LLM can consume millions of liters of freshwater for cooling purposes.[78] It has been demonstrated that more compact models using less compute can compete favorably with much larger counterparts.[79] The massively parallel nature of language model processing mandates use of graphics processing units (GPUs) for both training and efficient inference, these GPUs have high energy use, and while some authors argue their watts per record might be lower,[80] some authors argue that more efficient CPU hardware may also

be a route to pursue.[81] Smaller, more efficient models can often achieve comparable performance to their larger counterparts.[40,42] Smaller models also present a lower barrier to engagement with research given their requirement for cheaper computer hardware.

SUMMARY

Veterinary surveillance in companion animals has remained a blindspot in One Health disease monitoring but recent advances in the collation of vEHR into large datasets combined with advancement in computer methods has created new opportunities.

The evolution from rule-based systems through traditional neural architectures to modern transformer-based approaches demonstrates the field's rapid maturation. Domain-specific models such as PetBERT and VetBERT have proven that targeted pretraining on veterinary clinical text yields substantial performance gains while maintaining computational efficiency and environmental sustainability.

Significant challenges remain in creating nationally and internationally representative datasets for disease monitoring. Current models' generalizability beyond their training institutions requires investigation, because many systems have been developed on referral hospital data rather than primary care settings where most consultations occur. Additionally, the risk of capturing potentially identifiable information mandates robust deidentification frameworks are developed and validated.

Environmental sustainability considerations must be integrated into model selection and system design decisions, ensuring that advances in AI-powered surveillance align with broader One Health sustainability goals. Open datasets like PetEVAL provide valuable benchmarks, yet the field would benefit substantially from additional independently developed evaluation resources to support external validation and reduce overreliance on single-source benchmarks.

These are challenges that the veterinary sector can surely overcome. The collaborative nature of the veterinary informatics community, combined with increasing recognition of companion animals' importance in One Health frameworks, provides a strong foundation for addressing these obstacles. As computational methods continue to advance and international collaboration strengthens, AI-powered surveillance systems hold immense promise for transforming our ability to monitor and respond to emerging health threats in companion animal populations. By maintaining focus on both technical excellence and practical sustainability, the veterinary informatics community can develop surveillance capabilities that truly serve the interconnected health of animals, humans, and the environment.

CLINICS CARE POINTS

- *Your records are critical raw material.* Contributing notes to a data science initiative directly enables researchers to generate population-level health insights.
- *Annotation methods affect accuracy.* Simple word-pattern matching is brittle—negated terms, spelling variants, and inconsistent abbreviations can all cause missed or false-positive detections. Language models capture word context more effectively.
- *Training data context affects applicability.* Models developed on referral hospital records may not reflect first-opinion practice patterns.
- *Syndromic surveillance detects trends, not diagnoses.* Outbreak detection may be based on clinical signs, not confirmed diagnoses, thus a population-level alert requires clinical follow-up to explore causation.

- *General-purpose AI tools are not validated for veterinary use.* Generative LLMs have not been systematically evaluated on veterinary clinical text, therefore, domain-adapted models should be preferred where available for clinical or surveillance applications.

DECLARATION OF ARTIFICIAL INTELLIGENCE AND ARTIFICIAL INTELLIGENCE-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used generative AI. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

DISCLOSURES

The authors declare no conflicts of interest.

REFERENCES

1. UK pet food. UK pet population 2024. 2024. Available at: https://www.ukpetfood.org/industry-hub/data-statistics-/uk-pet-population-.html. Accessed January 12, 2025.
2. Hepper. 19 incredible UK pet spending statistics & facts: updated in 2025. 2025. Available at: https://articles.hepper.com/pet-spending-statistics-uk/. Accessed January 12, 2025.
3. Insurance Information Institute. Facts + statistics: pet ownership and insurance. 2024. Available at: https://www.iii.org/fact-statistic/facts-statistics-pet-ownership-and-insurance. Accessed January 12, 2025.
4. Pet advocacy network. 2024 economic impact of the US pet sector. Pet advocacy network. 2024. Available at: https://petadvocacy.org/economics. Accessed January 12, 2025.
5. Bae J, Ro C, Kang Y, et al. Human-to-animal transmission of SARS-CoV-2, South Korea, 2021. Emerg Infect Dis 2023;29(5):1087–9.
6. Smith SL, Anderson ER, Cansado-Utrilla C, et al. SARS-CoV-2 neutralising antibodies in dogs and cats in the United Kingdom. Current Research in Virological Science 2021;2:100011.
7. Bidaisee S, Macpherson CNL. Zoonoses and one health: a review of the literature. J Parasitol Res 2014;2014. https://doi.org/10.1155/2014/874345.
8. McEwen SA, Collignon PJ. Antimicrobial resistance: a one health perspective. Microbiol Spectr 2018;6(2). https://doi.org/10.1128/MICROBIOLSPEC.ARBA-0009-2017.
9. Kol A, Arzi B, Athanasiou KA, et al. Companion animals: translational scientist's new best friends. Sci Transl Med 2015;7(308). https://doi.org/10.1126/SCITRANSLMED.AAA9116.
10. Robertson ID, Irwin PJ, Lymbery AJ, et al. The role of companion animals in the emergence of parasitic zoonoses. Int J Parasitol 2000;30(12–13):1369–77.
11. Van Duijkeren E, Wolfhagen MJHM, Box ATA, et al. Human-to-Dog transmission of methicillin-resistant Staphylococcus aureus. Emerg Infect Dis 2004;10(12):2235.
12. Davies H, Blackwell E, Fins IS, et al. Recording of suspected adverse drug reaction reporting in veterinary free-text clinical narratives. J Small Anim Pract 2024;65(6):361–7.

13. Davies H, Noble PJ, Fins IS, et al. Developing electronic health records as a source of real-world data for veterinary pharmacoepidemiology. Front Vet Sci 2025;12:2025.
14. World Health Organization. One health. 2017. Available at: https://www.who.int/health-topics/one-health.
15. Small animal veterinary surveillance network. SAVSNET: small animal veterinary surveillance network. 2024. Available at: https://www.liverpool.ac.uk/savsnet/. Accessed January 12, 2025.
16. Companion animal veterinary surveillance network. CAVSNET: companion animal veterinary surveillance network. 2024. Available at: https://cavsnet.umn.edu/. Accessed January 12, 2025.
17. O'Neill DG, Church DB, McGreevy PD, et al. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. PLoS One 2014;9(3):e90501.
18. Banfield pet hospital. State of pet health 2016 report. Banfield pet hospital. 2016. Available at: https://www.banfield.com/state-of-pet-health.
19. Paynter AN, Dunbar MD, Creevy KE, et al. Veterinary big data: when data goes to the dogs. Animals 2021;11(7). https://doi.org/10.3390/ANI11071872.
20. 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;11–2024. https://doi.org/10.3389/fvets.2024.1352239.
21. Lustgarten J, Zehnder AM, Shipman W, et al. Veterinary informatics: forging the future between veterinary medicine, human medicine, and one health initiatives—a joint paper by the association for veterinary informatics (AVI) and the CTSA one health alliance (COHA). JAMIA Open 2020;3:306–17.
22. Turchin A, Kolatkar NS, Grant RW, et al. Using regular expressions to abstract blood pressure and treatment intensification information from the text of physician notes. J Am Med Inform Assoc 2006;13(6):691–5.
23. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst 2013;26. https://doi.org/10.48550/arXiv.1310.4546.
24. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Neural Information Processing Systems 2017;31:6000–10.
25.. Devlin J, Chang MW, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. Mineapolis (MN): Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2019. p. 4171–86.
26. Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners. OpenAI Blog 2019;1(8):9. Available at: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
27. Touvron H, Martin L, Stone K, et al. Llama 2: open foundation and fine-tuned chat models. 2023. Available at: https://arxiv.org/pdf/2307.09288.
28. Howard J, Ruder S. Universal language model fine-tuning for text classification. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers) 2018;1:328–39.
29. Liu Y, Ott M, Goyal N, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv preprint arXiv 2019. 190711692.
30. Heeley AM, O'Neill DG, Davison LJ, et al. Diabetes mellitus in dogs attending UK primary-care practices: frequency, risk factors and survival. Canine Medicine and Genetics 2020;7(1):1–19.

31. Cheng K, Baldwin T, Verspoor K. Automatic negation and speculation detection in veterinary clinical text. In: proceedings of the australasian language technology association workshop 2017. 2017. Available at: http://www.rvc.ac.uk/vetcompass.
32. Farrell S, Noble PJM, Pinchbeck GL, et al. Seasonality and risk factors for myxomatosis in pet rabbits in Great Britain. Prev Vet Med 2020;176. https://doi.org/10.1016/J.PREVETMED.2020.104924.
33. Farrell S, McGarry J, Noble PJM, et al. Seasonality and other risk factors for fleas infestations in domestic dogs and cats. Med Vet Entomol 2023;37(2):359–70.
34. Tulloch JSP, McGinley L, Sánchez-Vizcaíno F, et al. The passive surveillance of ticks using companion animal electronic health records. Epidemiol Infect 2017;145(10):2020–9.
35. Nie A, Zehnder A, Page RL, et al. DeepTag: inferring diagnoses from veterinary clinical notes. npj Digit Med 2018;1(1):1–8.
36. Zhang Y, Nie A, Zehnder A, et al. VetTag: improving automated veterinary diagnosis coding via large-scale language modeling. npj Digit Med 2019;2(1):1–8.
37. Lee J, Yoon W, Kim S, et al. Data and text mining BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2019;1–7. https://doi.org/10.1093/bioinformatics/btz682.
38.. Alsentzer E, Murphy J, Boag W, et al. Publicly available clinical BERT embeddings. In: Rumshisky A, Roberts K, Bethard S, Naumann T, editors. Proceedings of the 2nd clinical natural language processing workshop. Minneapolis (MN): Association for Computational Linguistics; 2019. p. 72–8. https://doi.org/10.18653/v1/W19-1909.
39. Hur B, Baldwin T, Verspoor K, et al. Domain adaptation and instance selection for disease syndrome classification over veterinary clinical notes. In: Demner-Fushman D, Cohen KB, Ananiadou S, et al, editors. Proceedings of the 19th SIGBioMed workshop on biomedical language processing. Association for Computational Linguistics; 2020. p. 156–66. https://doi.org/10.18653/v1/2020.bionlp-1.17.
40. Farrell S, Appleton C, Noble PJM, et al. PetBERT: automated ICD-11 syndromic disease coding for outbreak detection in first opinion veterinary electronic health records. Sci Rep 2023;13(1):1–14.
41. Sánchez-Vizcaíno F, Jones PH, Menacere T, et al. Small animal disease surveillance. Vet Rec 2015;177(23):591–4.
42. Boguslav MR, Kiehl A, Kott D, et al. Fine-tuning foundational models to code diagnoses from veterinary health records. arXiv preprint 2024. https://doi.org/10.48550/arXiv.2410.15186.
43. Jin YG, Wu G, Seo JW, et al. AI veterinary assistance: enhancing clinical decision-making in animal healthcare. IEEE Access 2025;13:119292–304.
44. 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):no.
45. Rabaey P, Arno H, Heytens S, et al. SynSUM – synthetic benchmark with structured and unstructured medical records. arXiv preprint 2024. https://doi.org/10.48550/arXiv.2409.08936.
46. Hoinville LJ, Alban L, Drewe JA, et al. Proposed terms and concepts for describing and evaluating animal-health surveillance systems. Prev Vet Med 2013;112(1–2):1–12.
47. Robinson NJ, Dean RS, Cobb M, et al. Diagnostic testing in first opinion small animal consultations. Vet Rec 2014;176(7):174.
48. Anholt RM, Berezowski J, Jamal I, et al. Mining free-text medical records for companion animal enteric syndrome surveillance. Prev Vet Med 2014;113(4):417–22.

49. Hashimoto S, Murakami Y, Taniguchi K, et al. Detection of epidemics in their early stage through infectious disease surveillance. Int J Epidemiol 2000;29(5):905–10.
50. Kass PH, Weng HY, Gaona MAL, et al. Syndromic surveillance in companion animals utilizing electronic medical records data: development and proof of concept. PeerJ 2016;2016(5):e1940.
51. Weng HY, Gaona MAL, Kass PH. Evaluation of a novel syndromic surveillance system for the detection of the 2007 melamine-related nephrotoxicosis foodborne outbreak in dogs and cats in the United States. PeerJ 2020;8:e9093.
52. Radford AD, Singleton DA, Jewell C, et al. Outbreak of severe vomiting in dogs associated with a canine enteric coronavirus, United Kingdom. Emerg Infect Dis 2021;27(2):517.
53. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research 2003;3:993–1022.
54. Chen JH, Goldstein MK, Asch SM, et al. Predicting inpatient clinical order patterns with probabilistic topic models vs conventional order sets. J Am Med Inform Assoc 2017;24(3):472–80.
55. Noble P-JM, Appleton C, Nenadic G, et al. Topic modelling to extract information from veterinary clinical records and inform surveillance in first opinion veterinary practice in the UK. PLoS One 2021;16(12):e0260402.
56. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv 2022. *220305794*.
57. Noble PJM, Farrell SO, Al-Moubayed N, et al. Comprehensive representation of health-related phenotypes in one million dogs using topic modelling of electronic health records. J Big Data 2026. https://doi.org/10.1186/s40537-026-01365-0.
58. Simon GE, Shortreed SM, Coley RY, et al. Assessing and minimizing re-identification risk in research data derived from health care records. eGEMs 2019;7(1):6.
59. Abouelmehdi K, Beni-Hssane A, Khaloufi H, et al. Big data security and privacy in healthcare: a review. Procedia Comput Sci 2017;113:73–80.
60. Dorr DA, Phillips WF, Phansalkar S, et al. Assessing the difficulty and time cost of de-identification in clinical narratives. Methods Inf Med 2006;45(3):246–52.
61. Leevy JL, Khoshgoftaar TM, Villanustre F. Survey on RNN and CRF models for de-identification of medical free text. J Big Data 2020;7(1):1–22.
62. Lee J, Jeong J, Jung S, et al. Verification of de-identification techniques for personal information using tree-based methods with shapley values. J Personalized Med 2022;12(2):190.
63. Yang X, Lyu T, Li Q, et al. A study of deep learning methods for de-identification of clinical notes in cross-institute settings. BMC Med Inf Decis Making 2019;19(5):1–9.
64. Syed M, Sexton K, Greer M, et al. DeIDNER model: a neural network NER model for use in the de-identification of clinical notes. In: Biomedical Engineering Systems and Technologies, International Joint Conference, BIOSTEC . Revised Selected Papers. BIOSTEC (Conference). Vol 5. NIH Public Access; 2022:640-647. doi:10.5220/0010884500003123.
65. Abu-El-Rub N, Urbain J, Kowalski G, et al. Natural language processing for enterprise-scale De-identification of protected health information in clinical notes. AMIA Summits on Translational Science Proceedings 2022;2022:92.
66. Ahmed T, Aziz MMA, Mohammed N. De-identification of electronic health record using neural network. Sci Rep 2020;10(1):1–11.
67. Jiang Z, Zhao C, He B, et al. De-identification of medical records using conditional random fields and long short-term memory networks. J Biomed Inf 2017;75:S43–53.

68. Lee K, Dobbins NJ, Mcinnes B, et al. Transferability of neural network clinical De-identification systems. J Am Med Inform Assoc 2021;28(12):2661–9.
69. Venkataraman GR, Pineda AL, Bear Don't Walk OJ, et al. FasTag: automatic text classification of unstructured medical narratives. PLoS One 2020;15(6):e0234647.
70. Chen RJ, Lu MY, Chen TY, et al. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng 2021;5(6):493–7.
71. Shumailov I, Shumaylov Z, Zhao Y, et al. AI models collapse when trained on recursively generated data. Nature 2024;631(8022):755–9.
72. Agyakari K, Babu R, Mulay S, et al. Position paper: building trust in synthetic data for clinical AI. arXiv preprint 2025. https://doi.org/10.48550/arXiv.2502.02076.
73. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3(1):1–9.
74. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10(1):1–9.
75. Farrell S, Radford A, Al Moubayed N, et al. PetEVAL: a veterinary free text electronic health records benchmark. In: Demner-Fushman D, Ananiadou S, Miwa M, et al, editors. Proceedings of the 24th workshop on biomedical language processing. Association for Computational Linguistics; 2025. p. 341–53. https://doi.org/10.18653/v1/2025.bionlp-1.29.
76.. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th annual meeting of the Association for computational linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 3645–50. https://doi.org/10.18653/v1/P19-1355.
77. Patterson D, Gonzalez J, Le Q, et al. Carbon emissions and large neural network training. arXiv preprint arXiv 2021. 210410350 Available at: https://arxiv.org/abs/2104.10350.
78. Li P, Yang J, Islam MA, et al. Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. arXiv preprint arXiv 2023. *230403271* Available at: https://arxiv.org/abs/2304.03271.
79. Kott D, Kiehl A, Price C, et al. Towards sustainable AI: evaluating efficiency of LLMs for coding of diagnoses from veterinary health records. In: 2025 IEEE Swiss conference on data science (SDS). 2025. p. 48–54. https://doi.org/10.1109/SDS66131.2025.00014.
80. Aditya AA. The energy efficiency paradox of AI hardware: debunking GPU power consumption myths. International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2025;11(2):21–7.
81. Yokoyama AM, Ferro M, de Paula FB, et al. Investigating hardware and software aspects in the energy consumption of machine learning: a green AI-centric analysis. Concurrency Comput Pract Ex 2023;35(24):e7825.