Use of LLMs to Enhance the Capture of User Input to Animal Health Reporting Systems

Gokul Krishna N Vijayakumaran Asari

Introduction & Background

Veterinary Health systems are essential for the well-being of animals, food safety, and public health, but conventional methods face several challenges. These challenges include Manual Data Entry, Limited Data Handling, Dependence on Experts, Resource-Intensive and Slow Processes, Delayed Diagnosis.

Manual Data Entry
Time-consuming, prone to errors, and inefficient.

Limited Data Handling
Restricted scope for managing and processing data.

Dependence on Experts
Relies on experts, limiting scalability and increasing inconsistencies.

Resource-Intensive and Slow Processes
Delays in decision-making and increased resource consumption.

Delayed Diagnosis
Leads to slower responses and missed early interventions.

As AI and Large Language Models (LLMs) advance, there is an opportunity to improve these systems by automating many of the processes to address these challenges. This study explores the integration of LLMs to overcome these challenges by examining how effectively LLMs can capture and record the necessary symptoms from patient descriptions to assist existing ML diagnostic tools.

Animal Disease Diagnostic Tool

The Animal Diagnostic Tool [1] helps diagnose animal diseases by letting users input species and symptoms data. A knowledgeable person identifies symptoms from a farmer's conversation and enters them into the tool, which then calculates disease probabilities.

Benefits of LLM Integration:

- Eliminates the need for veterinary experts, making diagnosis accessible to non-experts.
- Extracts symptoms automatically from conversations.
- Potential to improve efficiency, accuracy, and speed in diagnosis.
- Can be available via mobile apps and internet applications.

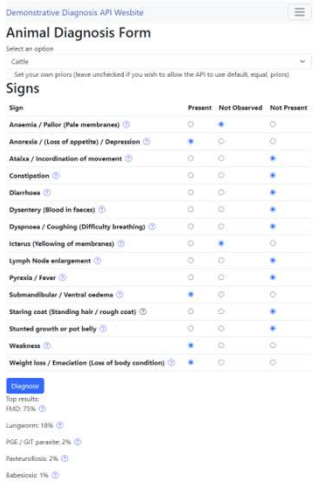

Fig.1 Diagnostic Tool UI [1]

Large Language Model (LLM)

Large Language Models (LLMs) are deep learning models based on transformer architecture, using an encoder and decoder with self-attention to understand text and word relationships. They are pre-trained on enormous text corpus and contain billions of parameters. [2]

Models Used in This Project:

	GPT 3.5 TURBO	EleutherAI /pythia-410m
Use :	General purpose	Research [4]
Params :	~ 175 billion	410 million [4]
Compute Need :	High	Moderate [4]
Capability :	Highly versatile	Task-specific tuning [4]

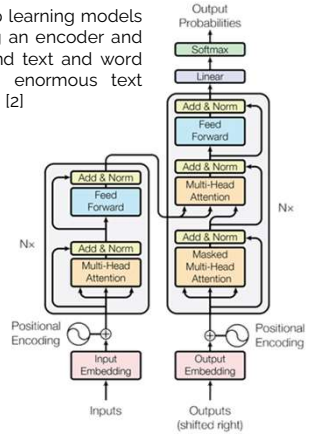

Fig.2 The Transformer - model architecture [3]

Study

These **experiments** evaluate the ability of Pythia and GPT-3.5 Turbo models in generating JSON outputs from symptom descriptions.

	Input Type	Output Type	Context Training	Models	Epochs
Exp 1	Simple description	JSON object	-	Pythia, GPT-3.5	3, 10, 20 (Pythia) 3, 10, 20 (GPT-3.5)
Exp 2	Detailed description	JSON object	-	Pythia, GPT-3.5	3, 10, 20 (Pythia) 3, 10 (GPT-3.5)
Exp 3	Simple description	JSON object	Yes (Pythia)	Pythia, GPT-3.5	3, 10, 20 (Pythia) 3, 10 (GPT-3.5)
Exp 4	Detailed description	JSON object	Yes (Pythia)	Pythia	3, 10, 20 (Pythia)
Exp 5	Detailed description	Existing symptoms +	-	Pythia	3, 10, 20 (Pythia)
Exp 6	Detailed description	Existing symptoms +	Yes (Pythia)	Pythia	3, 10, 20 (Pythia)

Table .1 DOE Matrix

Sample Symptom Description:

Hello, Doctor. I'm calling about one of my calves. It's been eating less than others. The calf is also having trouble breathing and isn't growing as well as the others. I've noticed it has a swollen belly and the skin around its jaw is puffy. Its gums and other membranes are turning yellow. These issues have been going on for a few days, and I'm quite worried. Can you help me figure out what might be wrong?

0 : not observed , 1 : present , -1 : absent

JSON object:

```
{   "animal": "Cattle",
    "signs": { "Anae": -1,
               "Anrx": 1,
               "Atax": -1,
               "Const": -1,
               "Diarr": -1,
               "Dysnt": -1,
               "Dyspn": 1,
               "Lymph": -1,
               "Pyrx": -1,
               "Stare": -1,
               "Stunt": 1,
               "Weak": -1,
               "Wght_L": -1,
               "Icter": 1,
               "SV_Oedm": 1
             }
}
```

Results

The results of the experiments were measured quantitatively using the following metrics: Recall, Precision, F1 Score, **Accuracy**, Exact Match, Animal Match, Number of Extra Signs, and Number of Missing Signs.

- **Experiment 2** seems to produce the best results, as it shows the highest accuracy in symptom signs.

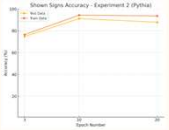

Fig.3 Pythia Symptom Accuracy

- The test accuracy at epoch 10 for Pythia's Experiment 2 is approximately **90.0%**
- Accuracy jumps sharply between epochs 3 and 10 and drops slightly after epoch 10.
- The model may benefit from early stopping to prevent overfitting and ensure generalization.

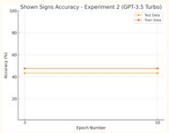

Fig.4 GPT3.5 Turbo Symptom Accuracy

- The test accuracy at epoch 10 is consistently around **43.0%**.
- There is no noticeable improvement across epochs, indicating that the model's learning has plateaued early on.
- This suggests a need for better optimization strategies

Challenges, Ethical Considerations & Future Directions

Data Formatting Issues: Improper JSON output formatting.

Hallucinations: Generation of unrelated or incorrect text.

Model-Specific Limitations: Pythia prone to more hallucinations.

Ethical Concerns: Misdiagnosis, liability, data security, and informed consent.

Future Work:

- Explore alternative models like Claude and Gemini etc.
- Enhance data quality and quantity.
- Implement advanced training techniques like RAG.
- Refine prompting strategies (e.g., few-shot prompting).
- Develop efficient scaling methods for public deployment.
- Adaptation to local languages.

References

[1] https://fraser.website/diagnose
[2] https://aws.amazon.com/what-is/large-language-model/#:~:text=Large%20language%20models%2C%20also%20known,decoder%20with%20self%2Dattention%20capabilities.
\bibitem[Vaswani2017]
[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, "Attention Is All You Need," arXiv preprint arXiv:1706.03762, submitted on 12 Jun 2017, last revised on 2 Aug 2023 (this version, v7), doi:10.48550/arXiv.1706.03762
[4] https://huggingface.co/EleutherAI/pythia-410m