



The Animal Whisperer Project

Oliver Bendel
FHNW School of Business
Windisch, Aargau, CH
oliver.bendel@fhnw.ch

Nick Zbinden
FHNW School of Business
Windisch, Aargau, CH
zbindennick@gmail.com

Abstract

Generative AI has become widespread since 2022. Technical advancements have resulted in multimodal large language models and other AI models that generate, analyze, and evaluate texts, images, and sounds. Such capabilities can be helpful in encounters between humans and animals. For example, apps with generative AI on a smartphone can be used to assess the body language and behavior of animals – e.g., during a walk or hike – and provide a recommendation for human behavior. It is often useful to take into account the animal’s environment and situation. The apps can help people to avert approaches and attacks, and thus also protect animals. In “The Animal Whisperer Project”, three apps were developed as prototypes based on the multimodal large language model GPT-4 from OpenAI from the beginning to mid-2024. Three specific GPTs resulted: the Cow Whisperer, the Horse Whisperer, and the Dog Whisperer. All three showed impressive capabilities after the first prompt engineering. These were improved by implementing information from expert interviews and adding labeled images of animals and other materials. AI-based apps for interpreting body language, behavior, and the overall situation can apparently be created today, without much effort, in a low-budget project. However, turning them into products would certainly raise questions, such as liability in the event of accidents.

CCS Concepts

• Social and professional topics; • Security and privacy; • Computing methodologies;

Keywords

Artificial Intelligence, Generative AI, Multimodal Large Language Model, GPT, Animal-Computer Interaction, Animal Body Language

ACM Reference Format:

Oliver Bendel and Nick Zbinden. 2024. The Animal Whisperer Project. In *The International Conference on Animal-Computer Interaction (ACI 2024)*, December 02–05, 2024, Glasgow, United Kingdom. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3702336.3702347>

1 INTRODUCTION

Generative AI has become widespread since the launch of ChatGPT in 2022. The global technical development resulted in multimodal

large language models (LLMs) and other AI models that can generate, analyze, and evaluate texts, images, and sounds. An early application from mid-2023 was the Be My Eyes app with the Be My AI function (www.bemyeyes.com). Blind and visually impaired people can use it to have their surroundings described to them. The images generated with the smartphone camera are analyzed and evaluated by GPT-4 [4]. The app also takes into account people’s posture and the atmosphere of rooms, and it recognizes and translates texts. Still images are to be followed by moving images so that moving scenes can also be described. Ultimately, disabled persons can find their way around their environment better without having to rely on human assistance.

It makes sense to envisage multimodal LLMs for other areas of application, such as animal-computer interaction, animal-machine interaction, and related areas [6, 7, 13, 14]. Encounters between humans and animals in which the parties involved are injured are a recurring problem. The reason for this often lies in careless human behavior. One example is accidents on cow pastures, where the mother animals defend their calves and attack the intruder. Other incidents are caused by domestic horses lunging or running away, or dogs biting. In these cases, the animals in question are livestock or pets. Wild animals can also become dangerous when encountering humans (or with livestock and pets) and put themselves in danger. If multimodal large language models have the above-mentioned capabilities, they may be able to interpret the body language and behavior of animals, according to the first author in early 2024. This could help both humans and animals.

To investigate this potential, the first author initiated a project entitled “The Animal Whisperer Project” at his university in February 2024. The second author was recruited as a collaborator. At this point, the multimodal large language model GPT-4 was already available. GPT-4o, which can also analyze video and audio with certain restrictions, was developed during the course of the project. It was not included due to time constraints and because the application itself was not yet practicable. Since the end of 2023, it has been possible to build so-called GPTs, “custom versions of ChatGPT”, as OpenAI calls them [3, 4]. These are in turn based on GPT-4 and their great advantage is that they are easy to create and customize while remaining powerful. They seemed ideal for developing and testing several prototypes. The idea was to investigate whether GPT-4 could interpret the body language and behavior of cows or cattle, (domestic) horses, and dogs, as well as assessing the overall situation, based on images to make appropriate and useful suggestions for the behavior of humans. As accidents involving livestock and pets are more common in Central Europe, where the project was based, the focus was placed on these animals. The team decided to implement a Cow Whisperer, a Horse Whisperer, and a Dog Whisperer as GPTs. The implementation and the expert interviews



This work is licensed under a Creative Commons Attribution International 4.0 License.

ACI 2024, December 02–05, 2024, Glasgow, United Kingdom

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1175-6/24/12

<https://doi.org/10.1145/3702336.3702347>

were documented in the project collaborator’s final thesis, which was part of the project [23].

In section 2, the two authors first develop the basics of multimodal LLMs. They explain why they chose to focus on cows or cattle, horses, and dogs, and outline the communication of these animals. They also list related projects. In the third section, the authors describe the implementation of the project. In particular, they discuss the prompt engineering and improvement of the apps with the help of labeled photos and other materials. The fourth section highlights the opportunities and risks of the project, also with a view to possible product development. At the end, in section 5, a summary is provided, and an outlook is given, particularly regarding the multimodality of the large language models. Overall, examples of all three apps are shown to provide an insight into each of them.

2 BASICS OF THE PROJECT

In this section, multimodal large language models are discussed, the focus of the project on cows or cattle, horses, and dogs is explained, and the communication and behavior of these animals is briefly explained. Related projects are also listed, without any claim to completeness.

2.1 Multimodal Large Language Models

A language model models the elements in a sequence (such as letters or words in a text). Stochastic processes predominantly play a role here. A large language model can understand and generate language for general purposes. It uses large amounts of data to learn billions or trillions of parameters during training. Multimodal large language models can process and output images, audio, and video in addition to text, for example. They belong to generative AI, i.e., a group of artificial intelligence (AI) applications that can generate content [3]. In the case of the apps in the project, it is essential that images can be analyzed and evaluated, conclusions can be drawn and, if necessary, illustrations can be output. In principle, LLMs such as GPT-4o can also analyze videos, i.e., moving images. However, tests in the project revealed considerable weaknesses. For example, videos were broken down into low-quality images that were difficult to analyze. In addition, videos could only be retrieved and analyzed via URLs, not via the local media library like the still images. The GPTs can be accessed via a browser and are therefore independent of devices and app stores. They can also be opened in the OpenAI app (see Figure 1).

2.2 Focusing on Cows, Horses, and Dogs

The authors’ decision to focus on the body language and behavior of cows or cattle, (domestic) horses, and dogs in the project is based on several criteria [23]. All three animal species are livestock or pets that have been influenced by humans through many years of domestication and breeding [9]. These procedures have not only changed their genetic structure and behavioral patterns, but also increased their presence. Cows or cattle, horses, and dogs are among the most common animal species in human settlements worldwide, apart from ants and other insect species. Although they are more or less animals that are used to humans and depend on humans, accidents occur time and again, sometimes with fatal

outcomes. This is due not just to the size and mass of the animals, but also to their ability to bite and lash out. Against this background, apps with multimodal LLMs that can assess animal behavior and provide recommendations for humans are of particular importance.

2.3 Communication and Behavior of Cows, Horses, and Dogs

Cows or cattle, horses, and dogs communicate in a variety of ways, with body language, acoustic signals, and olfactory markings playing a central role, as they do in all mammals [23]. In addition, there is the behavior, which manifests itself in movements, food procurement and intake, sleep phases, reproductive activities, etc. Different breeds of cows and cattle exhibit different behaviors [15]. The same applies to horses and dogs [17]. All three species were bred for very different purposes. This is perhaps most evident in dogs, which come in all sizes, shapes, and colors, and in some breeds have become very different from wolves. But horses also show considerable differences. The types of communication differ within the species. This is perhaps most evident in dogs, from lapdogs to guard dogs. The project focused on common breeds of cows and horses in Switzerland, as well as on dogs, as these are likely to have the most unpleasant encounters with serious consequences. The apps with multimodal LLMs can currently process visual information during encounters between humans and animals. This excludes acoustic signals and olfactory markings. Tactile aspects also play no role.

2.4 Related Projects

In some projects, AI is used to monitor wild animals in zoos and in the wild, for example, in terms of position, behavior, and health [8, 16]. Facial recognition has also been proposed in this context for tracking individuals [2]. Emotion recognition of pets is also being investigated [12]. Researchers in Switzerland have presented a novel deep learning architecture for classifying individual and social animal behavior – even in complex environments – directly from raw video images, which requires no human intervention after initial monitoring [16]. Other scientists have developed the multi-animal tracking system TRex that uses tagging techniques to precisely track animal behavior [21]. Machine vision has been used for years in the livestock industry, for example to improve the assessment of the condition of cows and pigs [18].

Overall, there are hardly any projects that incorporate generative AI to interpret animal body language. One reason for this is that this technology was hardly familiar to researchers outside the discipline until 2020. There are several projects on acoustic signals, i.e., decoding animal language. The best known is probably the Earth Species Project (www.earthspecies.org), which is dedicated to non-human primates, whales, and birds, among others. Apps for decoding cat language have also been on the market for years, such as MeowTalk [20]. Meowing is a special feature, as it is mostly a special language that animals use towards humans (and not towards each other). How well the apps work is controversial.

3 IMPLEMENTATION OF THE PROJECT

This section describes the project preparation, the creation of the GPTs, and pre-tests. The expert interviews are then outlined, as is the labeling of the photos for the knowledge database and the

adaptation of the GPTs. Finally, the systematic tests in preview mode and under real conditions are discussed.

3.1 Project Preparation, Creation of GPTs, and Pre-tests

“The Animal Whisperer Project” lasted from February to August 2024. The following research question was posed: “Can multi-modal large language models such as GPT-4 interpret the body language and behavior of cows, horses, and dogs, as well as the overall situation, based on images and make suggestions for human behavior to avoid serious confrontations?” In addition to cows, cattle were also relevant. Horses were understood as domestic horses. First of all, the authors developed the necessary foundations for behavioral research in these animals [1], also in connection with new technologies [19]. They already had experience in using artificial intelligence, particularly generative AI, and in creating GPTs, including sub-areas such as prompt engineering and adding documents to expand the knowledge base (in this context one can also speak of Retrieval-Augmented Generation, or RAG for short). In this context, access to ChatGPT Plus was set up, the prerequisite for being able to create and test GPTs.

In March 2024, the project collaborator created three GPTs on openai.com with the names Cow Whisperer, Horse Whisperer, and Dog Whisperer via the “Configure” tab. He defined the short description of the chatbot in addition to the name. A comic-style avatar was also created using the integrated DALL-E 3. In the case of Dog Whisperer, the short description was “Behavior analyst for the body language, facial expressions, and behavior of dogs” (here and in the following, the first author’s own translation from German), to give just one example. In the “Instructions” field, complex prompts were created for each GPT (these are developer prompts, which are to be distinguished from user prompts). The project collaborator based the structure and wording on best practices. There was no template for this type of GPT in the context of animal-computer interaction.

Each prompt included 13 categories, namely: (1) conversational tone, (2) expert knowledge, (3) analysis of images, (4) feedback quality, (5) environmental context, (6) response accuracy and speed, (7) privacy and security, (8) interaction with other tools, (9) feedback and customization, (10) special requests, (11) continuous learning readiness, (12) interactive elements, and (13) multimodal content. The purpose was stated in advance, for example “Helping users to better understand dog behavior and body language to promote a healthy and positive relationship between humans and dogs” in the case of the Dog Whisperer. The maximum number of 8000 characters was not fully utilized in this phase with 6768 characters. “Online search” and “DALL-E image generation” were activated for the “Functions”. The GPTs can be accessed for test purposes via the following URLs: chatgpt.com/g/g-m892ehlEy-cow-whisperer (Cow Whisperer), chatgpt.com/g/g-WOFIYkvK-horse-whisperer (Horse Whisperer), and chatgpt.com/g/g-mv19RVDmS-dog-whisperer (Dog Whisperer). ChatGPT Plus access is required.

Initial, non-systematic tests showed that powerful GPTs were already available with the capabilities of the multimodal LLM and the prompt engineering that had taken place. The team members uploaded images from the computer or smartphone of various

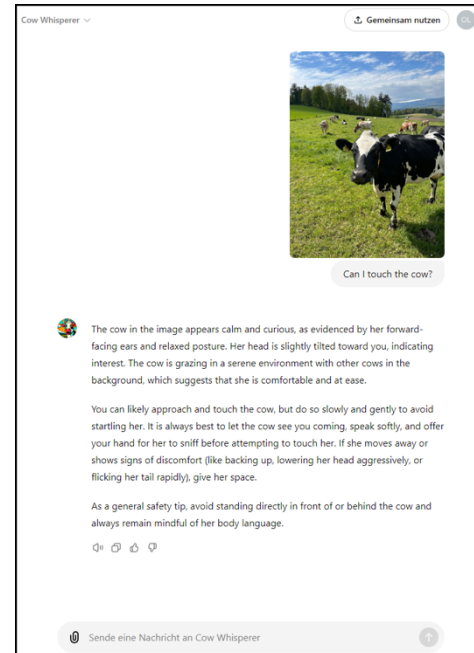


Figure 1: The Cow Whisperer.

animals in different situations. They also took pictures with the smartphone. In all cases, the GPTs interpreted the body language and behavior, if it was recognizable in the still images, and took into account the situation, such as whether it was a single animal or a group or herd and provided detailed recommendations for the user. As the team members had acquired basic knowledge about the behavior of the animal species and the appropriate behavior of humans, they were able to assess the tests to some extent and rate them as successful overall. In particular, the tips on human behavior were always effective.

3.2 Expert Interviews, Labeling of Photos, and Adaptation of GPTs

From May to June 2024, the project collaborator conducted expert interviews with five Swiss experts (also referred to here as “animal whisperers”) in the field of the behavior of cows and cattle, horses (like leisure horses and sport horses), and dogs, as well as with farmers and animal owners [23]. The interviews were recorded with a dictaphone and then transcribed and analyzed qualitatively. The overall purpose was to learn more about the animals’ body language and behavior, to understand how they are kept outdoors and in stables, and to obtain opinions on the design and use of such apps.

For the Cow Whisperer, the project collaborator took all the pictures himself using an SLR camera, in different environments and situations and at different times of day, in order to get as comprehensive a picture as possible of the animals’ body language and behavior, as well as their environment [23]. He labeled and described the images and then had them reviewed by the two experts. They also provided materials such as “Health and welfare

check points” – one for a healthy cow and one for a sick cow – and “Kuhsignale sehen und verstehen” (“Seeing and understanding cow signals”) and “Jungviehsignale sehen und verstehen” (“Seeing and understanding young cattle signals”). The project collaborator also created most of the images for the Horse Whisperer himself – he received some from the Weiss family, the images of the horse races and those from breeding in Ireland – and had them checked by the relevant experts. The pictures for the Dog Whisperer came from various sources. Some of them were taken by himself, some were provided by owners, and some were provided by the dog expert, who then examined them all.

Based on the feedback from the experts, the project team adapted the prompts in the “Instructions” field. There were minimal problems with the recommendations for user behavior. In terms of focus, there was room for improvement. The 13 categories were retained, but the explanations were made more precise (for example, a scheme for the answer was specified, so that each of them was constructed in the same way). In addition, the role was added, the purpose (as before), and now also the main tasks (analysis and interpretation, advice and information, support in solving problems) of the GPT were included, as well as the target group. The maximum number of characters had now been almost exhausted at 7767.

With the help of the five experts, extensive documents with labeled and unlabeled, but otherwise described and explained images were created for all three GPTs. The project collaborator created or collected 10 – 15 images per animal species and labeled them. He uploaded the documents to the system in the “Knowledge” section. He also included the materials mentioned, where available. In the “Instructions”, he drew the GPTs’ attention to the new materials and their significance for the analysis of animal behavior and body language. In the “Functions”, “Online search”, “DALL-E 3 image generation”, and “Code interpreter and data analysis” were activated. The latter option is important so that the attached documents can be analyzed and integrated.

The project collaborator tagged each photo with specific labels documenting different body regions with corresponding aspects of the animals’ body language and behavior [23]. These labels were used to create a structured dataset that served as the basis for the model’s fitting. The objects in the images were labeled as follows (with additional options for the cows):

- Red frame (ears): Marks the head area of the animal, especially the ears.
- Orange frame (muzzle): Highlights the muzzle, nostrils, and mouth area.
- Light red frame (eye): Places the focus on the eye to determine eye contact and direction of gaze.
- Green frame (animal): Encompasses the entire body of the animal and is used to analyze general posture and movements.
- Blue frame (legs): Focuses on the animal’s legs to examine gait and stride.
- Light blue frame (tail/tail): Marks the tail area to analyze tail and tail movements.
- Pink frame (horns): In cows, the horns are also specially marked.

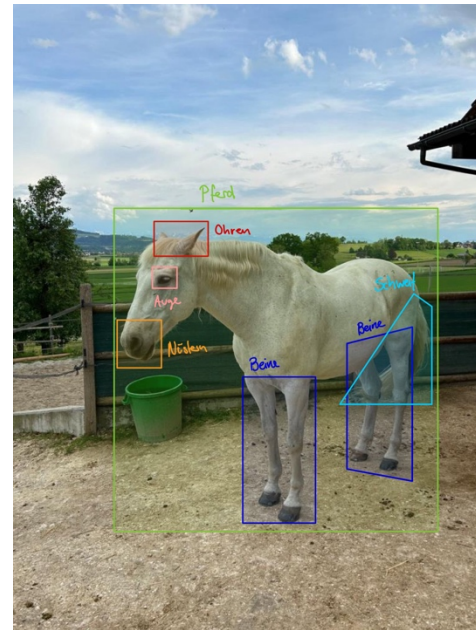


Figure 2: Labeled horse picture [23].

- Purple frame (warning triangle): In cows, the “warning triangle” (“starvation pit”), which – seen from behind – is located on the left side of the body and says a lot about the cow’s condition, is additionally marked.

The descriptions were linked to these markings [23]. These are based on the authors’ own knowledge and observations as well as on the feedback from the five “animal whisperers”. The aspects described included (in generic terms):

- Posture: Position and alignment of the body
- Movement: Dynamic actions such as walking, standing, or lying down
- Expression: Facial expressions and gestures, including ear and tail movements
- Interactions: Behavior in relation to other animals or people
- Context: Environment and situation in which the photo was taken
- Behavioral analysis: interpretation of the behavior shown in relation to typical behavioral patterns of the respective animal species
- Body language: Specific indications of the animal’s body language, such as tension, relaxation, or specific communication signals

Figure 2 shows one of 12 images created and labeled for the Horse Whisperer. In this case, the dimensions mentioned are described as follows, using the “template” shown above [23]:

- Stance: The horse stands upright; the body is straight. It stands at a slight angle with its head slightly lowered.
- Movement: The horse stands still, but the tail moves as if the horse is wagging it. The tail movement primarily serves to ward off insects.

- **Expression:** The ears are more to the side or even to the rear, which could indicate attention or slight discomfort. The facial expression appears rather neutral to slightly tense.
- **Interactions:** No direct interactions with other animals or humans can be seen.
- **Context:** The photo was taken in a free-range pen. A green meadow and hills can be seen in the background. There are two other horses in the stable that can move around freely.
- **Behavioral analysis:** The horse shows a slightly tense posture, which is made clear by the ears pointing backwards and the lowered head. The wagging tail indicates that the horse may be trying to ward off flies or show its mood.
- **Body language:** The ears are pointed backwards, which may indicate attention, discomfort, or tension. The head is slightly lowered, which may indicate relaxation, tiredness, or lack of interest. The eyes appear calm, but there may be a slight tension. The horse stands upright but appears slightly tense, which could indicate a restless or waiting posture. The tail moves as if the horse is wagging, which may indicate fly defense or an emotional reaction.

3.3 Tests of the Customized GPTs

In addition to the aforementioned initial tests after the creation of the first version of the GPTs, the project collaborator carried out systematic tests from the beginning to mid-July 2024 after the GPTs had been adapted (revision of the prompts and uploading of the documents with the labeled photos) [23]. The prompts were created in preview mode using the TAG approach (Task, Action, Goal). This was to check whether the apps basically work, with the focus on analysis and evaluation of body language and behavior. Here is an example of the Dog Whisperer [23]:

- **TASK:** I want to understand what the dog in the picture is communicating with its body language and behavior, especially since it is looking towards something outside the picture.
- **ACTION:** Please analyze the dog's posture and facial expressions in the photo. Pay attention to details such as the posture, the orientation of its ears, the position of its tail, and the attention of its eyes.
- **GOAL:** The goal is to get a detailed interpretation of the dog's body language in order to better assess what the dog is trying to communicate to me.

Approximately 5 tests were carried out on each animal. The first test of the Horse Whisperer included a photo of a horse in the pasture. This was taken in advance, as were the other photos in this series of tests. The Horse Whisperer responded to the standardized prompt: "The horse in the picture shows several interesting behavioral characteristics and body language that I would like to analyze in detail here." It provided the following analysis [23]:

- **Ear position:** The horse's ears are pointed forward. This indicates that the horse is attentive and interested. It seems that something in the environment has caught its attention. The pricked ears usually signal that the horse is curious and alert.
- **Eyes:** The horse's eyes appear calm and alert. There are no signs of stress or fear, such as wide-open or narrowed eyes.

Calm eyes indicate that the horse is neither stressed nor anxious at this moment, but relaxed and focused.

- **Nostrils:** The horse's nostrils are relaxed and not flared. Dilated nostrils could indicate excitement, anxiety or stress, while relaxed nostrils are a sign of calm and contentment. The horse is probably breathing calmly and evenly.
- **Posture:** The horse is standing in a relaxed posture in the pasture. It has a balanced weight distribution on all four legs, which indicates that it feels comfortable. The general body tension is relaxed, which is another sign of relaxation and contentment.
- **Environment:** The horse is in a green meadow, surrounded by trees and mountains in the background. This natural and calm environment probably contributes to the horse's relaxed attitude. Horses generally feel more comfortable in familiar and safe environments.

The app summarized the results as follows: "Overall, the horse in the picture shows signs of attention, curiosity, and relaxation. The pricked ears and calm eye expression indicate that it is interested in its surroundings without showing signs of stress or fear. The relaxed posture and calm surroundings support this impression. It appears that the horse is in a familiar and safe environment and feels comfortable. This interpretation of body language can help you to better assess the horse's well-being and adjust your interactions accordingly. When a horse is alert and relaxed, you can approach or work with it with confidence as it appears to be in a positive emotional state."

Compared to the first tests – an example of a conversation with the Cow Whisperer was shown above – the analysis is more detailed and precise. This finding is confirmed by the other tests. However, there are also misjudgments. For example, the Horse Whisperer was shown a picture of an arm with a hand touching the horse. The app said: "The interaction with the human appears to be positive. The human appears to be soothing or stroking the horse, which the horse accepts and possibly enjoys." This was followed by a corrective prompt from the user, in this case the project collaborator: "Unfortunately, that's not quite correct. Please analyze the interaction carefully. I'll help you understand it better. The human provokes the horse with his action and the horse reacts in an irritated manner. That's why the ears are turned backwards. However, it is not fear or stress, but annoyance on the part of the horse. You are welcome to compare the picture with what you know so that you have more information." The Horse Whisperer then corrected its analysis as desired.

After these systematic tests in preview mode and final adjustments to the prompts, the project collaborator carried out further tests from a user perspective from mid to late July 2024 [23]. He used the OpenAI app on his smartphone (iPhone 14 Plus). No corrective instructions were given. Rather, the application was simply asked to analyze and evaluate one photo at a time and make recommendations. This approach was intended to provide results to determine accuracy and reliability under real-world conditions. The tests took place in different environments that correspond to the animals' natural or familiar habitats and everyday activities.

The specific environments were chosen to be familiar to the animals so that the results would not be distorted by unusual external influences.

- The project collaborator carried out the tests with the Cow Whisperer mainly in the pasture in order to record the cows' natural behavior in an open environment. The pasture offers the cows enough space for movement and makes it possible to observe their social behavior and their reactions to various stimuli. In addition, he carried out tests during his walks. The focus here was on the applicability of the app under real conditions. The aim of these additional tests was to recreate situations in which tourists encounter cows while hiking in the mountains in order to analyze their reactions.
- The project worker carried out the tests with the Horse Whisperer on a horse farm. This environment offers optimal conditions for observing the animals' behavior and reactions in a familiar environment. The horse farm in question is equipped with all the necessary facilities, including stables, grazing areas, and riding arenas. The tests included various tasks designed to reflect the horses' natural behavior in their daily environment. These included reactions to unfamiliar objects, behavioral observations during grazing, and reactions to human interactions.
- The tests with the Dog Whisperer took place during walks by the project collaborator, both in urban and rural environments. These test environments were chosen to observe the dogs' behavior in different social and physical or structural contexts. The walks made it possible to study the dogs' reactions to different stimuli such as other animals, people, and traffic.

A total of 75 tests (25 per animal) were performed with all three applications. A correspondingly large number of photos and conversation histories were produced [23]. The method used in the tests consisted of visually checking the pictures and comparing them with the written assessments of the GPTs. One point was awarded for each match between the observation of the animal's body language or behavior and the description, and none for discrepancies. Finally, the results of the individual tests were summarized to determine the percentage of correct assessments and to obtain an overall accuracy of the behavioral analysis. Table 1 shows the results of the individual tests, and Table 2 shows the results of all the tests, sorted by attribute.

The results can be summarized as follows:

- The three apps were able to provide a meaningful interpretation in practically every normal situation with common breeds and under average conditions. The accuracy of all GPTs was over 80 percent (cows: 86.38; horses: 81.25; dogs: 81.21).
- The apps obviously used the technical vocabulary of the documents uploaded to the knowledge base. For example, the Horse Whisperer spoke of the pointed ears of a horse. Theoretically, it could also have used this terminology from its data material or from an online search – but there were other similarities of this kind.
- The apps also obviously benefited from the content of the labeled photos and other materials. This is also obvious

Table 1: Individual tests Cow Whisperer

Test Number	Correct	Possible	Wrong	Percent
1	14	19	5	73.68
2	4	4	0	100
3	5	6	1	83.33
4	8	8	0	100
5	9	11	2	81.81
6	7	8	1	87.5
7	4	5	1	80
8	11	11	0	100
9	5	5	0	100
10	8	9	1	88.89
11	4	4	0	100
12	6	8	2	75
13	4	5	1	80
14	6	11	5	54.54
15	22	25	3	88
16	8	10	2	80
17	6	9	3	66.67
18	11	11	0	100
19	9	9	0	100
20	11	11	0	100
21	9	9	0	100
22	12	16	4	75
23	9	9	0	100
24	6	7	1	85.714
25	24	27	3	88.89
Total:	222	257	35	86.38

because the team built the knowledge base specifically for the requirements of the virtual animal whisperers. Moreover, this is precisely the principle of GPTs – they obtain their power through their customization by means of prompt engineering and the development of the knowledge base.

- The content of the apps' recommendations was correct and appropriate for the situation in question. They generally resulted in defensive behavior on the part of the user. The recommendations are easy to understand and easy to implement. They are also not too long, so that the necessary actions can be implemented immediately.
- Cow Whisperer and Horse Whisperer were the most convincing. One explanation is that a particularly large number of labeled photos and other materials were fed into these. Another explanation is that it is relatively difficult to interpret the body language of dogs uniformly on a visual basis, for example because they have very different ears. These challenges hardly exist with cows and domestic horses.
- If the animals were dark, such as dark brown or black, or if they had long hair, the apps struggled. It is particularly difficult for them to come to a conclusion in intermediate positions (such as the ears). In addition, the angle of the images, the quality of the photos, and the lighting conditions on site are decisive for the quality of the interpretation. If the animal is moving a lot, high-quality images are difficult.

Table 2: All tests Cow Whisperer with attributes

Attribute	Correct	Possible	Wrong	Percent
Ears	34	46	12	73.91
Eyes	14	19	5	73.68
Head	27	34	7	79.41
Body	46	50	4	92
Tail	4	4	0	100
Surroundings	15	15	0	100
Total	3	3	0	100
Interaction	3	3	0	100
Behavior	11	12	1	91.67
Fur/Skin	2	2	0	100
Gaze	10	14	4	71.43
Horns	2	2	0	100
Group	1	1	0	100
Health	2	2	0	100
Muzzle	3	3	0	100

- The apps described all animals as living beings that should be treated with respect. They did not call for violence towards them, not even self-defense. When in doubt, the advice was always to keep your distance, stay calm, etc. No distinction was made between farm animals and pets. This shows that possible biases do not have to occur in a specific app with a specific purpose [11].

It should be noted that tests with generative AI can only provide snapshots, as these are stochastic systems. The results are based on probabilities and therefore to a certain extent on coincidences. Nevertheless, trends can be identified. In addition, a complex knowledge base has been developed that guides the system in certain directions.

4 OPPORTUNITIES AND RISKS OF THE APPS

This section discusses the opportunities and risks of the three apps in “The Animal Whisperer Project”. The authors’ own findings from the implementations and tests as well as their own considerations and statements from the experts play a role here. The opportunities are mentioned first.

- The three apps are able to analyze and evaluate the body language and behavior of animals and the overall situation and make recommendations for the user’s behavior on this basis. This provides laypeople with important information on walks and hikes or other encounters with animals that could potentially protect their health and save their lives.
- The correct behavior of app users also protects the health and lives of animals. When animals approach, attack, or defend themselves, they often put themselves in danger and become stressed – along with other animals present. In some cases, they are euthanized or slaughtered after the attacks [10].
- Almost everyone has a smartphone and carries it around with them on walks and hikes – a small, mobile gadget that is closely connected to our everyday lives. The apps can be accessed via the browser and are therefore independent of

the operating system and device. Alternatively, they can be accessed via the OpenAI app. No additional device is required for the apps, and they are always ready for use.

- The presence and use of the apps can increase users’ interest in animal body language and behavior. They are also able to build their own GPTs that may be more suited to their own needs and experience (in fact, there are already several Dog Whisperers in the GPT Store). However, they should read up on labeling data and take advantage of it.

There are also some problems and risks associated with using the apps:

- As the first tests and then the user tests have already revealed, it depends on the quality or the section of the image. Figure 1 shows a cow at close range. The Cow Whisperer concluded in one of the conversations that the cow is alone in the pasture. But this is not necessarily the case. Prompt engineering could solve this problem to a certain extent by asking the app to take a photo from a greater distance.
- The recommendations of the three apps can be incorrect, for example, due to misinterpretation of body language, behavior, and the overall situation. Although the user tests did not show any total failures, these cannot be ruled out. This can cause all kinds of consequences, including collisions between humans and animals, resulting in injuries and fatalities. Misunderstandings can also arise from the user prompts.
- This goes hand in hand with responsibility and liability issues that are familiar from other apps, for example, in the healthcare sector. As with their other projects, the team members deliberately created prototypes and did not publish them worldwide, but only via links. When it comes to product development, there are a few problems to solve, not least from a legal perspective.
- If apps of this kind are widespread, there is a risk that users will rely on the recommendations and trust their own findings less. They may be less willing to have bad and good experiences and gain knowledge themselves. This danger was also pointed out by one of the five “animal whisperers” [23].
- The three apps cannot process moving images or acoustic, olfactory, and tactile signals. However, this means that important information for the overall assessment is missing. The multimodality of the large language models will offer further possibilities in this respect in the future [3]. The analysis of moving images is particularly interesting in this context.
- As a rule, the right to one’s own image is not infringed when recording animals, unless people are present, which may be the case with dog owners and horse riders, for example. However, personal objects and land or buildings may be in the picture, and people – such as other walkers – may feel harassed.
- ChatGPT Plus access is required to use the apps. This is subject to a fee and is not available to everyone. In addition, you are tying yourself to a provider that has great market power and is quite controversial, for example, because of its

use of training data and the remuneration of employees in the process of reinforcement learning from human feedback.

- You also need a mobile phone (or a computer with internet access) to use the apps. Although you have good coverage in Europe and most other parts of the world, it can happen that you end up in a dead zone. In this rare case, the applications are of no value.
- As already mentioned, generative AI is a stochastic system. This not only has implications for the tests, but also for the practice of the applications. However, it is precisely in these situations that one should always be able to rely on the results. This was solved to a certain extent by developing a complex knowledge base.

This only covers a few points. The opportunities and risks of apps that analyze and evaluate images are also listed in the paper on *Be My Eyes* [4]. There must be further work in which a detailed ethical, social, technical, and economic discussion is conducted.

5 SUMMARY AND OUTLOOK

Three GPTs were developed in “The Animal Whisperer Project”, which already had certain capabilities in the first version. They reminded of the *Be My AI* function, which inspired the present project. It was shown how the capabilities of the LLM can increase the quality of the apps through clever prompt engineering and the addition of suitable documents with labeled and described images and other materials. The research question was answered positively.

The project had a low budget. Apart from the working time, less than 250 dollars were spent, mainly on ChatGPT Plus access and travel expenses. Three powerful apps were created and documented within half a year with around 400 hours of work. Future projects can benefit from this. The effort involved could be significantly reduced if the prompts and information are made freely available. Approaches such as the *SuperAnimal* model can also be considered. This uses pre-trained pose estimation models that can be applied to over 45 animal species [22]. Alternatively, fine-tuning can be done for each species, but this can be very time-consuming and beyond the capabilities of non-experts. In addition, the language model can also be pushed in an undesirable direction.

In any case, multimodality is a key to success. In this context, analyzing videos means that movements of the animal and its ears, tail, extremities, etc., can be interpreted even more extensively. However, it will not be easy to obtain rights-free material to expand the video base. But, as in the present project, this could also be produced in-house. The analysis of audio opens up further dimensions, whereby the language model is currently conditioned primarily on the human voice and manner of speech, not on animal sounds.

Finally, the principle of analysis and evaluation with the help of multimodal large language models can be transferred to numerous other areas of application. It is also possible to combine solutions and approaches. For example, *Cow Whisperer*, *Horse Whisperer*, and *Dog Whisperer* could be of benefit to the blind and visually impaired, and their specific capabilities could be integrated into the *Be My Eyes* app. Here, analyzing the visual is exactly what the blind or visually impaired person is missing.

Lastly, the efforts that can be located in the field of animal-computer interaction or at least in neighboring fields can be extended to the field of animal-machine interaction in the narrower sense. This is because multimodal large language models are already changing robotics. The machines equipped with them perceive their environment and can react to requests and prompts in a way that was previously only conceivable in the context of scenario building and training in real space. This brings visions of the use of robots and drones in nature reserves for monitoring and protecting animals within reach [5, 6].

Acknowledgments

The authors would like to thank the animal whisperers, namely Christian Manser and Thyas Künzle (cows), Miro and Ricardo Weiss (horses), and André Kännel (dogs). They would also like to thank all the other interviewees. Without them, this project could not have taken place in this form.

References

- [1] Patrick Bateson and Kevon N. Laland. 2013. Tinbergen’s four questions: An appreciation and an update. *Trends in Ecology & Evolution*, 28(12), 712–718. <https://doi.org/10.1016/j.tree.2013.09.013>
- [2] Oliver Bendel and Ali Yürekirmaz. 2023. A Face Recognition System for Bears: Protection for Animals and Humans in the Alps. In *Ninth International Conference on Animal-Computer Interaction (ACI’22)*, December 05–08, 2022, Newcastle-upon-Tyne, United Kingdom. ACM, New York 2023. <https://dl.acm.org/doi/proceedings/10.1145/3565995>
- [3] Oliver Bendel. 2024. 300 Keywords Generative KI: Ökonomische, technische und ethische Grundlagen. Springer Gabler, Wiesbaden
- [4] Oliver Bendel. 2024. How Can Generative AI Enhance the Well-being of Blind? In: *Proceedings of the AAI 2024 Spring Symposium Series, Symposium “Impact of GenAI on Social and Individual Well-being”*. Stanford University, Stanford, California, March 25–27, 2024. The AAI Press, Washington, DC. <https://ojs.aaai.org/index.php/AAAI-SS/article/view/31232/33392>
- [5] Oliver Bendel. 2022. Passive, Active, and Proactive Systems and Machines for the Protection and Preservation of Animals and Animal Species. *Frontiers of Animal Science*, 3:834634 (2022). DOI: 10.3389/fanim.2022.834634. <https://www.frontiersin.org/articles/10.3389/fanim.2022.834634/full>
- [6] Oliver Bendel. 2018. Towards Animal-friendly Machines. *Paladyn, Journal of Behavioral Robotics*, 2018, Volume 9, Issue 1:204–213. <https://www.degruyter.com/view/journals/pjbr/9/1/article-p204.xml>
- [7] Oliver Bendel. 2013. Considerations about the Relationship between Animal and Machine Ethics. *AI & SOCIETY*, 11 December 2013. <https://link.springer.com/article/10.1007/s00146-013-0526-3>
- [8] Jenna V. Congdon, Mina Hosseini, Ezekiel F. Gading, Mahdi Masousi, Maria Franke, and Suzanne E. MacDonald. 2022. The Future of Artificial Intelligence in Monitoring Animal Identification, Health, and Behaviour. *Animals* 12, no. 13:1711. <https://doi.org/10.3390/ani12131711>
- [9] Carlos A. Driscoll, David W. Macdonald, and Stephen J. O’Brien. 2009. From wild animals to domestic pets, an evolutionary view of domestication, *PNAS*, 16 June 2009, 106 (supplement_1) 9971–9978. <https://doi.org/10.1073/pnas.0901586106>
- [10] Erwan Le Garff, Vadim Mesli, and Yann Delannoy. 2016. Domestic Predation of an Elder: A Fatal Dog Attack Case. *Journal of Forensic Sciences*, 9 June 2017. <https://doi.org/10.1111/1556-4029.13370>
- [11] Thilo Hagendorff, Leonie N. Bossert, Yip Fai Tse, and Peter Singer. 2023. Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals. *AI and Ethics* (2023) 3:717–734. <https://doi.org/10.1007/s43681-022-00199-9>
- [12] Jacek Mazurkiewicz. 2024. Artificial Intelligence Methods for Pet Emotions Recognition. In Wojciech Zamojski, Jacek Mazurkiewicz, Jaroslaw Sugier, Tomasz Walkowiak, and Janusz Kacprzyk (eds) *System Dependability – Theory and Applications. DepCoS-RELCOMEX 2024. Lecture Notes in Networks and Systems*, vol 1026. Springer, Cham. https://doi.org/10.1007/978-3-031-61857-4_16
- [13] Clara Mancini. 2011. Animal-computer interaction: a manifesto. *Interactions*, Volume 18, Issue 4:69–73. <https://dl.acm.org/doi/10.1145/1978822.1978836>
- [14] Clara Mancini and Eleonora Nannoni. 2023. Editorial: Animal-computer interaction and beyond: The benefits of animal-centered research and design. *Front. Vet. Sci.* 9:1109994. doi: 10.3389/fvets.2022.1109994
- [15] Lori Marino and Kristin Allen. 2017. The Psychology of Cows. *Animal Behavior and Cognition*, 4(4):474–498. <https://doi.org/10.26451/abc.04.04.06.2017>

- [16] Markus Marks, Qiuhan Jin, and Oliver Sturman *et al.* 2022. Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments. *Nature Machine Intelligence* 4:331–340. DOI: 10.1038/s42256-022-00477-5
- [17] Katrina Merckies and Olivia Franzin. 2021. Enhanced Understanding of Horse–Human Interactions to Optimize Welfare. *Animals*, 11(5), 1347. <https://doi.org/10.3390/ani11051347>
- [18] Abozar Nasirahmadi, Sandra A. Edwards, and Barbara Sturm. 2017. Implementation of machine vision for detecting behaviour of cattle and pigs. *Livestock Science*, Volume 202:25-38. <https://doi.org/10.1016/j.livsci.2017.05.014>
- [19] Gail L. Patricelli. 2023. Behavioral ecology: New technology enables a more holistic view of complex animal behavior. *PLOS Biology*, 21(8), e3002264. <https://doi.org/10.1371/journal.pbio.3002264>
- [20] Rostislav Skripchenko and Ivan Burlakov. 2022. The Cat’s Meow–Feline Translations. *Technology and Language* 3.3 (2022):22–37
- [21] Tristan Walter and Iain D. Couzin. 2021. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *eLife*, 10, e64000. <https://doi.org/10.7554/eLife.64000>
- [22] Shaokai Ye, Anastasiia Filippova, and Jessy Lauer *et al.* 2024. SuperAnimal pre-trained pose estimation models for behavioral analysis. *Nature Communications*, 15(1), 5165. <https://doi.org/10.1038/s41467-024-48792-2>
- [23] Nick Zbinden. 2024. Animal Whisperer: A GenAI App for Decoding Animal Body Language. Bachelor Thesis. FHNW School of Business, Olten