A Whisker of Truth: A Multimodal Interdisciplinary Machine Learning Approach to Vocal, Visual, and Tactile Signals in the Domestic Cat

Astrid van Toor
ML Lead, blueOASIS
Ericeira, Portugal 2655-321
avtoor@blueoasis.pt
astridvantoor@gmail.com

Susanne Schötz[*]
Division of Logopedics, Phoniatrics,
and Audiology & Humanities Lab
Lund University, Sweden
susanne.schotz@med.lu.se

Elin Hirsch
Department of Applied Animal Science and Welfare
Swedish University of Agricultural Sciences, Skara, Sweden
elin.hirsch@med.lu.se

Abstract

We propose a multimodal deep learning framework for automated analysis of cat–human communication, integrating acoustic, visual, and tactile signals through transformer-based fusion. Using the largest expert-annotated dataset of its kind and interdisciplinary collaboration, we combine semi-supervised learning with ethological and phonetic expertise to detect subtle behavioural and phonetic cues, enable early welfare assessment, and establish species-generalisable methods.

1 Background

Humans and cats have had a mutual tolerance for millenia [1] and with an estimated 600 million cats worldwide [2] they are one of the most common companion animals. Despite this, the communication habits of our feline companions remain poorly understood. Studies in recent years have made progress in deciphering cats' ability to recognise individual humans [3] and their own names [4], as well as their use of multimodal signals in interactions with humans [5, 6]. Yet, automated behaviour analysis using machine learning techniques remains under-explored or lack expert involvement [7]. Others raise bias concerns due to absence of appropriate subject grouping mechanism across train and test sets [8].

Cats communicate through a rich repertoire of visual, behavioural, and vocal signals. They use visual cues to convey intentions during interactions with humans [9], and behavioural signals can influence adoption outcomes in shelters [10]. They also demonstrate social awareness, adapting their communication to human attentiveness [11]. Their vocal repertoire is extensive and variable, shaped by learning to better express needs and emotions to humans [12, 13]. Additionally we know that vocal prosody of the domestic cat varies with physical context [14]. Humans remain poor interpreters of these signals - even experienced cat owners achieve only modest accuracy when classifying vocalisations [15], and subtle visual expressions such as jaw drop or upper lip raise are often overlooked [16]. Misinterpretation of these signals risks compromising feline welfare, as unmet needs can manifest in undesired behaviours, abandonment, or relinquishment.

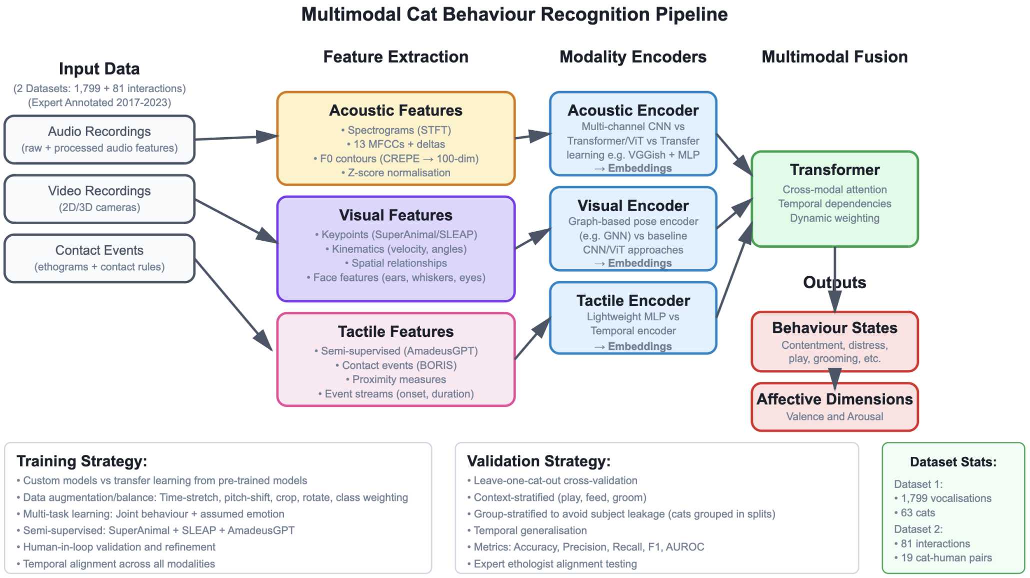
[*] https://meowsic.se/

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Deep learning has been effective in recognising specific cat vocalisations [17] and age related vocalisation changes [18], with emerging applications such as automated pain detection through facial analysis [19] and behavioural classification using accelerometry through wearable sensors [20]. While these studies demonstrate the potential of individual modalities, there is a notable gap in research combining acoustic, visual, and behavioural data for holistic understanding of feline communication and welfare assessment. Current approaches treat each modality in isolation, missing the rich interplay between vocalisations, facial expressions, and body language that characterises natural feline communication. Additionally, tactile data is still lacking in the literature - marking a promising frontier.

This gap highlights the urgent need for systematic, automated approaches that integrate expertise from phonetics, ethology, and machine learning that our proposal aims to address. AI-driven multimodal systems hold particular promise in bridging the gap between feline communication and human understanding, enabling new tools to support both research and everyday welfare monitoring. Previous AI systems (e.g., MeowTalk) have yielded disappointing results, in part due to a lack of integration with linguistic and behavioural expertise. By contrast, our team brings together internationally recognised expertise in feline phonetics (Susanne Schötz) and ethology (Elin Hirsch), and specialism in machine learning for bioacoustics analysis (Astrid van Toor). Importantly, our methods are strictly non-invasive, relying on naturalistic recordings and expert annotation to avoid stress to the animals. While our immediate focus is on cats, the methodology generalises to other species, addressing a broader challenge in AI for animal communication.

2 Proposed approach


Multimodal Cat Behaviour Recognition Pipeline
Input Data
(2 Datasets: 1,799 + 81 interactions)
(Expert Annotated 2017-2023)
Audio Recordings
(raw + processed audio features)
Video Recordings
(2D/3D cameras)
Contact Events
(ethograms + contact rules)
Feature Extraction
Acoustic Features
• Spectrograms (STFT)
• 13 MFCCs + deltas
• F0 contours (CREPE → 100-dim)
• Z-score normalisation
Visual Features
• Keypoints (SuperAnimal/SLEAP)
• Kinematics (velocity, angles)
• Spatial relationships
• Face features (ears, whiskers, eyes)
Tactile Features
• Semi-supervised (AmadeusGPT)
• Contact events (BORIS)
• Proximity measures
• Event streams (onset, duration)
Modality Encoders
Acoustic Encoder
Multi-channel CNN vs Transformer/ViT vs Transfer learning e.g. VGGish + MLP
→ Embeddings
Visual Encoder
Graph-based pose encoder (e.g. GNN) vs baseline CNN/ViT approaches
→ Embeddings
Tactile Encoder
Lightweight MLP vs Temporal encoder
→ Embeddings
Multimodal Fusion
Transformer
Cross-modal attention
Temporal dependencies
Dynamic weighting
Outputs
Behaviour States
Contentment, distress, play, grooming, etc.
Affective Dimensions
Valence and Arousal

Training Strategy:

- Custom models vs transfer learning from pre-trained models
- Data augmentation/balance: Time-stretch, pitch-shift, crop, rotate, class weighting
- Multi-task learning: Joint behaviour + assumed emotion
- Semi-supervised: SuperAnimal + SLEAP + AmadeusGPT
- Human-in-loop validation and refinement
- Temporal alignment across all modalities

Validation Strategy:

- Leave-one-cat-out cross-validation
- Context-stratified (play, feed, groom)
- Group-stratified to avoid subject leakage (cats grouped in splits)
- Temporal generalisation
- Metrics: Accuracy, Precision, Recall, F1, AUROC
- Expert ethologist alignment testing

Dataset Stats:

Dataset 1:
- 1,799 vocalisations
- 63 cats

Dataset 2:
- 81 interactions
- 19 cat-human pairs

Figure 1: Complete machine learning workflow

2.1 Data

We leverage the largest expert-annotated multimodal datasets of cat-human interactions to date, comprising synchronised audio, video, and behavioural annotations from naturalistic home environments between 2017-2023. All interactions were recorded with two cameras (one head-mounted, one filming the whole scene). Meowsic [21] comprises of 1799 cat vocalisations and interactions from 63 individual cats and 36 humans, of which the acoustics have been expertly annotated by Schötz - our specialist in cat phonetics - and her prior research team [21]. Extensive acoustic features have been pre-extracted and behaviours annotated through a double-blind review process. These cats were observed vocalising at each other as well as with humans. Call types were classified based on phonetic

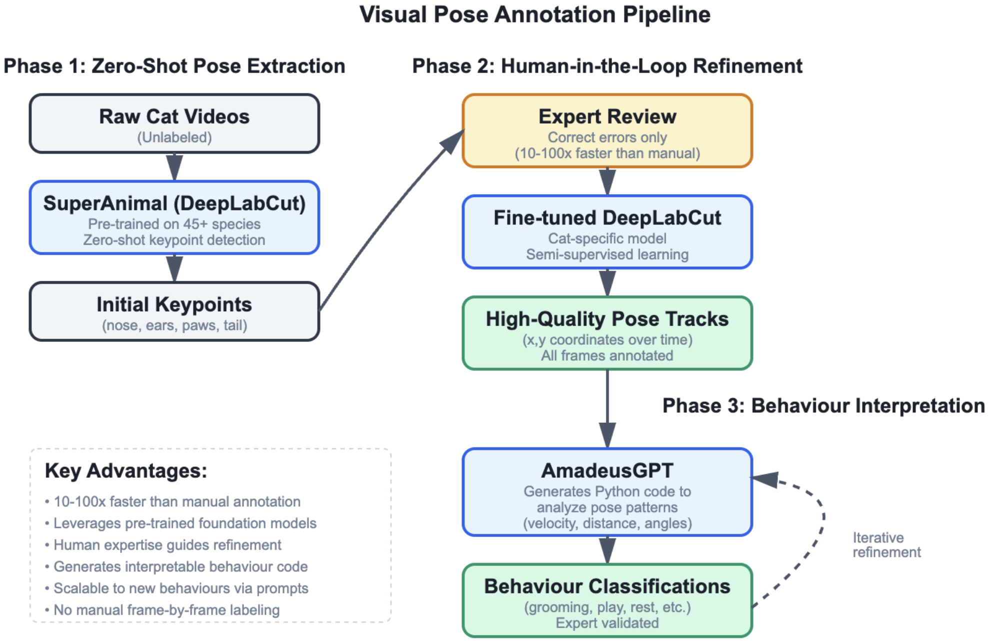

Visual Pose Annotation Pipeline
Phase 1: Zero-Shot Pose Extraction
Phase 2: Human-in-the-Loop Refinement
Raw Cat Videos
(Unlabeled)
SuperAnimal (DeepLabCut)
Pre-trained on 45+ species
Zero-shot keypoint detection
Initial Keypoints
(nose, ears, paws, tail)
Expert Review
Correct errors only
(10-100x faster than manual)
Fine-tuned DeepLabCut
Cat-specific model
Semi-supervised learning
High-Quality Pose Tracks
(x,y coordinates over time)
All frames annotated
Phase 3: Behaviour Interpretation
Key Advantages:
• 10-100x faster than manual annotation
• Leverages pre-trained foundation models
• Human expertise guides refinement
• Generates interpretable behaviour code
• Scalable to new behaviours via prompts
• No manual frame-by-frame labeling
AmadeusGPT
Generates Python code to
analyze pose patterns
(velocity, distance, angles)
Behaviour Classifications
(grooming, play, rest, etc.)
Expert validated
Iterative
refinement

Figure 2: Pose annotation workflow

features according to [22]. Further visual and tactile signals will be annotated as per the pipeline outlined in Figure 2 and elaborated on in the following paragraph. The cat-human-communication (CHC) dataset [5] captures 81 readily annotated interactions across 19 cat-human pairs with frame-level annotations for acoustic events (vocalisations, pitch contours), visual behaviours (ear positions, facial expressions, posture), and tactile interactions (contact initiation, stroking patterns). Only healthy cats were included in the trials. Metadata about the cats' and owners' experience, familiarity, and individual characteristics were collected via online questionnaires based on [21]. Owners also provided information about their cats' personality.

We are aware of the importance of scalability and have outlined plans for expanding data collection; given that factors such as breed, age, sex, housing conditions, location, and seasonal changes are known to influence behaviour [20], we will continue expanding our data collection efforts to build upon our existing foundation and achieve more comprehensive representation across these key variables leveraging the previous framework and recording equipment [21, 10] (see also Appendix B1). Our validation strategy (Section 2.6) furthers ensures generalisation to unseen individuals, and the modular architecture supports adaptation to larger datasets and new contexts.

As for breeds, in the Meowsic dataset more than half of the cats were housecats or mixed breeds, whereas the others constituted a variety of different pure breeds including Maine Coon, Bengal and Siamese [21]. The CHC dataset also includes mainly housecats and mixed breeds (including Persian-Norwegian Forest Cat and Persian-Maine Coon).

Due to the absence of comparable expert-annotated multimodal cat-human interaction datasets, traditional state-of-the-art comparisons are not feasible. Our work establishes the first comprehensive benchmark in this domain, pioneering a foundation for future comparative studies.

2.2 Annotation

To scale annotation beyond manual labelling we employ a semi-supervised pipeline combining tools such as SuperAnimal [23] and SLEAP [24] for automated keypoint extraction with human-in-the-loop refinement. This approach reduces annotation time by 10-100x [23] while maintaining expert-level accuracy. Importantly, annotation is an ongoing, iterative process central to this project. We are actively developing comprehensive ethograms based on the latest research [25–27] and our emerging findings, rather than relying solely on existing classifications. This iterative approach allows us to identify context-dependent and relationship-specific behavioural patterns (e.g., how tail elevation differs between cats with varying human familiarity) that static ethograms may overlook. Semi-supervised tactile identification is obtained by feeding time-series pose data to AmadeusGPT [28] to generate hypotheses like: "*Cat ear angle changed immediately after contact onset → tactile influence on behaviour*", which are then validated and refined by our expert team. While AmadeusGPT is emerging, it has demonstrated robustness across multiple benchmarks including the MABe 2022 challenge, cross-species validation (mice, horses), and stress-testing with out-of-distribution prompts achieving 88% consistency. The tool has been validated with a random sample of real user queries showing practical reliability.

Remaining multimodal communicative signals will be coded using the event-logging software BORIS [29] and ethograms developed in earlier studies [25, 26, 30]. All modalities will be temporally aligned to common frame and sample rates.

2.3 Multimodal Feature Extraction

Our approach extracts complementary features from each modality, designed to capture the full spectrum of feline communicative signals:

- **Acoustic**: A triple feature stack representation of (1) spectrograms via STFT, (2) 13 MFCCs plus deltas encoding perceptual characteristics, and (3) fundamental frequency (f0) contours extracted via CREPE [31] resampled to 100-dimensional vectors for multi-channel neural network input. All features undergo z-score normalisation computed on the training set.
- **Visual**: Keypoint coordinates from SuperAnimal [23] / SLEAP [24] are augmented with derived kinematic features (velocities, accelerations, inter-joint angles). We encode spatial relationships between key body parts (ear-to-ear distance, tail curvature, head-to-body angle) that may correlate with affective states. Face-specific features include ear angle relative to head, whisker position, and pupil dilation where visible.
- **Tactile**: Annotated human–cat contact events through semi-supervised analysis using AmadeusGPT [28] and BORIS [29] (touch, stroking, play) for pose-based features (proximity of cat and human limbs). These signals are encoded as event streams (onset, duration, type of contact), which are passed through a lightweight MLP or temporal encoder. This allows the model to capture the role of tactile interaction in shaping cat communicative behaviour, despite the absence of direct tactile sensors.

2.4 Architecture and training strategy

Each signal will employ a modality-specific neural network to act as feature extractor. These embeddings are then processed by a transformer network that learns cross-modal attention patterns and temporal dependencies from time-series data. We hypothesise that this architecture allows the model to dynamically weigh the importance of different modalities over time. For example prioritising acoustic features during vocalisations, visual cues during silent displays, or analysing their combinations when signals co-occur, such as a meow that shifts meaning depending on accompanying posture or tactile context. For the individual modalities we will explore a number of feature extraction methods such as; custom multi-channelled convolutional neural networks (section 2.3), embeddings from prominent pre-trained neural network feature extractors (e.g. VGGish [32], AmadeusGPT [28]) followed by downstream multilayer perceptron processing, transformer extractors, graph neural networks for skeletal structure modelling of pose sequences, and transfer learning by fine-tuning robust pre-existing models. Consequently this work will present a comparative review of SOTA machine learning techniques in digital bioacoustics and computer vision for animal pose. We will include a variety of data augmentation techniques such as time-stretching and pitch-shifting for audio, and random cropping and rotation for video, with class weighting to address potential behavioural imbalance.

Our end-to-end pipeline processes raw video input through sequential stages: (1) pose estimation using SLEAP/SuperAnimal for keypoint detection, (2) multimodal feature extraction including kinematic, acoustic, and tactile signals, (3) behavioural pattern learning and deviation detection, and (4) automated report generation with optional expert validation for continuous improvement. The pipeline is illustrated in Fig. 1.

2.5 Error Mitigation

We acknowledge that multi-stage pipelines can accumulate errors. To address this we will: (1) validate each component independently (pose estimation accuracy, behavioural annotation consistency, multimodal fusion performance), (2) implement confidence scoring at each stage to flag uncertain outputs, and (3) use expert-in-the-loop validation to catch and correct propagated errors.

Modern pose estimation frameworks have demonstrated high accuracy that minimise error propagation in downstream tasks. SLEAP achieves a mAP score of 0.927 and DeepLabCut (DLC) of 0.928, with SLEAP being faster than DLC at 2,194 versus 458 FPS [24, 33]. SuperAnimal, built on DeepLabCut, further enhances data efficiency by requiring 10-100× less labelled data than standard transfer-learning approaches while maintaining high accuracy [23]. These foundation models can perform zero-shot inference across 45+ species including cats, demonstrating robust generalisation to diverse imaging conditions and animal morphologies. For challenging scenarios with occlusions or ambiguous poses, SLEAP's confidence scoring system allows uncertainty quantification at the keypoint level, which we will leverage to flag low-confidence detections for manual review.

Rather than relying on fixed threshold-based rules, our approach emphasises learned behavioural patterns from annotated data and deviation detection from individual baselines. We acknowledge that concrete behavioural thresholds (e.g., specific ear angles indicating discrete emotional states) remain an area of active research. Instead, our system learns to identify general indicators of potential concern—such as tail positions suggesting stress, changes in movement patterns, or altered response patterns to familiar humans—which flag cases for further investigation rather than making definitive classifications. AmadeusGPT serves as a complementary tool for discovering complex spatio-temporal patterns and enabling natural language behavioural queries. Critically, we leverage the rich contextual metadata in our datasets (cat-human familiarity, interaction history, environmental setting) to develop context-aware models that account for relationship-specific behavioural nuances.

2.6 Evaluation

To prevent subject leakage that may have been compromised in previous studies [8], and to ensure scientific rigour, we will employ robust and industry-standard validation strategies namely leave-one-cat out cross-validation to assess generalisation to new individuals, cross-validated hold-out validation, and context- and group-stratified validation ensuring all contexts and individual cats are represented equally and separately across train and test sets. We measure the model's ability across context through accuracy, precision, recall, f1-score, and AUROC measures. In addition to standard machine learning metrics, we will evaluate the system's potential for real-world usability by testing how well the models align with expert ethologists' annotations and whether predictions generalise across home environments.

3 Expected outcomes and impact

Our project provides the first comprehensive, multimodal architecture for domestic cat vocal, visual, and tactile signals, providing open-source benchmarks and datasets to advance the field of computational animal behaviour. By establishing ground-truth ethograms and healthy behavioural baselines, we aim to automate welfare assessments such as subtle facial expression changes or shifts in vocal prosody or touch that humans struggle to perceive. This reduces observer bias and enables earlier detection of stress, pain, or illness. While our focus is on domestic cats, the proposed framework generalises readily to other species. The modular multimodal pipeline (acoustic, visual, tactile) can be adapted to diverse animal communication contexts, supporting comparative ethology and cross-species welfare studies.

The design supports a fully automated, personalised pipeline through modular integration of pose estimation, feature extraction, and behavioural analysis. A critical component is personalised baseline modelling: the general model, pre-trained on our diverse multi-cat dataset, learns broad communication patterns, then fine-tunes on individual user recordings of their cat during healthy periods. In production, the system operates through a two-stage process: (1) users establish an individualised baseline through multiple healthy-period recordings, and (2) subsequent recordings are analysed for deviations that flag potential welfare concerns. This deviation-based approach addresses individual variation in communication style, breed differences, and unique cat-human relationship dynamics—for example, tail elevation patterns may differ between cats with strong versus weak human attachment, or play behaviour may manifest differently across breeds and age groups. Rather than diagnosing specific conditions, the system identifies general indicators (altered movement patterns, changes in responsiveness, shifts in typical communication) warranting closer examination. Optional expert review services can provide interpretation of flagged cases, with anonymised production data feeding back into model refinement through active learning to continuously improve detection capabilities.

3.1 Acknowledgements

The authors gratefully acknowledge Lund University Humanities Lab, the Marcus and Amalia Wallenberg Foundation, and Agria och Svenska Kennelklubbens Forskningsfond (the Agria and SKK Research Foundation).

References

[1] Leslie A. Lyons and Jennifer Dawn Kurushima. A Short Natural History of the Cat and Its Relationship with Humans. In *The Cat*, pages 1254–1262. Elsevier, 2012. ISBN 978-1-4377-0660-4. doi: 10.1016/B978-1-4377-0660-4.00042-9. URL `https://linkinghub.elsevier.com/retrieve/pii/B9781437706604000429`.

[2] Carlos A. Driscoll, Juliet Clutton-Brock, Andrew C. Kitchener, and Stephen J. O'Brien. The Taming of the Cat. *Scientific American*, 300(6):68–75, June 2009. ISSN 0036-8733. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5790555/`.

[3] Atsuko Saito and Kazutaka Shinozuka. Vocal recognition of owners by domestic cats (Felis catus). *Animal Cognition*, 16(4):685–690, July 2013. ISSN 1435-9456. doi: 10.1007/s10071-013-0620-4. URL `https://doi.org/10.1007/s10071-013-0620-4`.

[4] Atsuko Saito, Kazutaka Shinozuka, Yuki Ito, and Toshikazu Hasegawa. Domestic cats (Felis catus) discriminate their names from other words. *Scientific Reports*, 9(1):5394, April 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-40616-4. URL `https://www.nature.com/articles/s41598-019-40616-4`. Publisher: Nature Publishing Group.

[5] Elin N Hirsch, Joost van de Weijer, and Susanne Schötz. Vocal, visual, and tactile signals in cat–human communication: A pilot study. In *Proceedings of the 4th International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR)*, Kos, Greece, September 2024. 6 and 9 Sep 2024.

[6] Charlotte de Mouzon and Gérard Leboucher. Multimodal Communication in the Human–Cat Relationship: A Pilot Study. *Animals*, 13(9):1528, January 2023. ISSN 2076-2615. doi: 10.3390/ani13091528. URL `https://www.mdpi.com/2076-2615/13/9/1528`. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Stavros Ntalampiras, Danylo Kosmin, and Javier Sanchez. Acoustic classification of individual cat vocalizations in evolving environments. In *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, pages 254–258, July 2021. doi: 10.1109/TSP52935.2021.9522660. URL `https://ieeexplore.ieee.org/abstract/document/9522660`.

[8] Stavros Ntalampiras, Luca Andrea Ludovico, Giorgio Presti, Emanuela Prato Previde, Monica Battini, Simona Cannas, Clara Palestrini, and Silvana Mattiello. Automatic Classification of Cat Vocalizations Emitted in Different Contexts. *Animals*, 9(8):543, August 2019. ISSN 2076-2615. doi: 10.3390/ani9080543. URL `https://www.mdpi.com/2076-2615/9/8/543`. Publisher: Multidisciplinary Digital Publishing Institute.

[9] Tasmin Humphrey, Leanne Proops, Jemma Forman, Rebecca Spooner, and Karen McComb. The role of cat eye narrowing movements in cat–human communication. *Scientific Reports*, 10(1):16503, October 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-73426-0. URL `https://www.nature.com/articles/s41598-020-73426-0`. Publisher: Nature Publishing Group.

[10] Elin Netti Hirsch, Maria Andersson, Jenny Loberg, and Lena Maria Lidfors. Development of existing scoring systems to assess behavioural coping in shelter cats. *Applied Animal Behaviour Science*, 234: 105208, January 2021. ISSN 0168-1591. doi: 10.1016/j.applanim.2020.105208. URL `https://www.sciencedirect.com/science/article/pii/S0168159120302963`.

[11] Lingna Zhang, Katie B. Needham, Serena Juma, Xuemei Si, and François Martin. Feline communication strategies when presented with an unsolvable task: the attentional state of the person matters. *Animal Cognition*, 24(5):1109–1119, September 2021. ISSN 1435-9456. doi: 10.1007/s10071-021-01503-6.

[12] M. Moelk. Vocalizing in the house-cat; a phonetic and functional study. *The American Journal of Psychology*, 57:184–205, 1944. ISSN 1939-8298. doi: 10.2307/1416947. Place: US Publisher: Univ of Illinois Press.

[13] Susanne Schötz. *The Secret Language of Cats: How to Understand Your Cat for a Better, Happier Relationship*. Hanover Square Press, Toronto, 2018.

[14] Susanne Schötz, Joost van de Weijer, and Robert Eklund. Context effects on duration, fundamental frequency, and intonation in human-directed domestic cat meows. *Applied Animal Behaviour Science*, 270:106146, January 2024. ISSN 0168-1591. doi: 10.1016/j.applanim.2023.106146. URL `https://www.sciencedirect.com/science/article/pii/S0168159123003180`.

[15] Nicholas Nicastro and Michael J. Owren. Classification of domestic cat (Felis catus) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, 117(1): 44–52, March 2003. ISSN 0735-7036. doi: 10.1037/0735-7036.117.1.44.

[16] C. C. Caeiro, A. M Burrows, and B. M. Waller. Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions? *Applied Animal Behaviour Science*, 189:66–78, April 2017. ISSN 0168-1591. doi: 10.1016/j.applanim.2017.01.005. URL `https://www.sciencedirect.com/science/article/pii/S0168159117300102`.

[17] Weilin Sun, Vincent Lu, Aaron Truong, Hermione Bossolina, and Yuan Lu. Purrai: A Deep Neural Network based Approach to Interpret Domestic Cat Language. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 622–627, December 2021. doi: 10.1109/ICMLA52953.2021.00104. URL `https://ieeexplore.ieee.org/document/9680210`.

[18] Astrid van Toor, Nadeem Qazi, and Stefania Paladini. A deep learning pipeline for age prediction from vocalisations of the domestic feline. *Scientific Reports*, 15(1):1–18, October 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-17986-z.

[19] Marcelo Feighelstein, Ilan Shimshoni, Lauren R. Finka, Stelio P. L. Luna, Daniel S. Mills, and Anna Zamansky. Automated recognition of pain in cats. *Scientific Reports*, 12(1):9575, June 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-13348-1. URL `https://www.nature.com/articles/s41598-022-13348-1`. Publisher: Nature Publishing Group.

[20] Michelle Smit, Rene A. Corner-Thomas, Ina Draganova, Christopher J. Andrews, and David G. Thomas. How Lazy Are Pet Cats Really? Using Machine Learning and Accelerometry to Get a Glimpse into the Behaviour of Privately Owned Cats in Different Households. *Sensors*, 24(8):2623, January 2024. ISSN 1424-8220. doi: 10.3390/s24082623. URL `https://www.mdpi.com/1424-8220/24/8/2623`. Publisher: Multidisciplinary Digital Publishing Institute.

[21] Susanne Schötz, Joost van de Weijer, and Robert Eklund. Phonetic methods in cat vocalisation studies: A report from the Meowsic project, June 2019. URL `https://zenodo.org/records/3245999`. Conference Name: Fonetik 2019 Publisher: Zenodo.

[22] Susanne Schötz. Phonetic Variation in Cat–Human Communication. In M. Ramiro Pastorinho and Ana Catarina A. Sousa, editors, *Pets as Sentinels, Forecasters and Promoters of Human Health*, pages 319–347. Springer International Publishing, Cham, 2020. ISBN 978-3-030-30734-9. doi: 10.1007/978-3-030-30734-9_14. URL `https://doi.org/10.1007/978-3-030-30734-9_14`.

[23] Shaokai Ye, Anastasiia Filippova, Jessy Lauer, Steffen Schneider, Maxime Vidal, Tian Qiu, Alexander Mathis, and Mackenzie Weygandt Mathis. SuperAnimal pretrained pose estimation models for behavioral analysis. *Nature Communications*, 15(1):5165, June 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-48792-2. URL `https://www.nature.com/articles/s41467-024-48792-2`. Publisher: Nature Publishing Group.

[24] Talmo D. Pereira, Nathaniel Tabris, Arie Matsliah, David M. Turner, Junyu Li, Shruthi Ravindranath, Eleni S. Papadoyannis, Edna Normand, David S. Deutsch, Z. Yan Wang, Grace C. McKenzie-Smith, Catalin C. Mitelut, Marielisa Diez Castro, John D'Uva, Mikhail Kislin, Dan H. Sanes, Sarah D. Kocher, Samuel S.-H. Wang, Annegret L. Falkner, Joshua W. Shaevitz, and Mala Murthy. SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19(4):486–495, April 2022. ISSN 1548-7105. doi: 10.1038/s41592-022-01426-1. URL `https://www.nature.com/articles/s41592-022-01426-1`. Publisher: Nature Publishing Group.

[25] Lauren Ashley Stanton, Matthew Stephen Sullivan, and Jilian Marie Fazio. A standardized ethogram for the felidae: A tool for behavioral researchers. *Applied Animal Behaviour Science*, 173:3–16, December 2015. ISSN 0168-1591. doi: 10.1016/j.applanim.2015.04.001. URL `https://www.sciencedirect.com/science/article/pii/S0168159115001008`.

[26] Bertrand L. Deputte, Estelle Jumelet, Caroline Gilbert, and Emmanuelle Titeux. Heads and Tails: An Analysis of Visual Signals in Cats, Felis catus. *Animals*, 11(9):2752, September 2021. ISSN 2076-2615. doi: 10.3390/ani11092752. URL `https://www.mdpi.com/2076-2615/11/9/2752`. Publisher: Multidisciplinary Digital Publishing Institute.

[27] Janice M. Siegford, Juan P. Steibel, Junjie Han, Madonna Benjamin, Tami Brown-Brandl, Joao R. R. Dórea, Daniel Morris, Tomas Norton, Eric Psota, and Guilherme J. M. Rosa. The quest to develop automated systems for monitoring animal behavior. *Applied Animal Behaviour Science*, 265:106000, August 2023. ISSN 0168-1591. doi: 10.1016/j.applanim.2023.106000.

[28] Shaokai Ye, Jessy Lauer, Mu Zhou, Alexander Mathis, and Mackenzie W. Mathis. AmadeusGPT: a natural language interface for interactive animal behavioral analysis, July 2023. URL `http://arxiv.org/abs/2307.04858`. arXiv:2307.04858 [cs].

[29] Olivier Friard and Marco Gamba. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11):1325–1330, 2016. ISSN 2041-210X. doi: 10.1111/2041-210X.12584. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12584`. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12584.

[30] Lauren R. Finka, Lucia Ripari, Lindsey Quinlan, Camilla Haywood, Jo Puzzo, Amelia Jordan, Jaclyn Tsui, Rachel Foreman-Worsley, Laura Dixon, and Marnie L. Brennan. Investigation of humans individual differences as predictors of their animal interaction styles, focused on the domestic cat. *Scientific Reports*, 12(1):12128, July 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-15194-7. URL `https://www.nature.com/articles/s41598-022-15194-7`. Publisher: Nature Publishing Group.

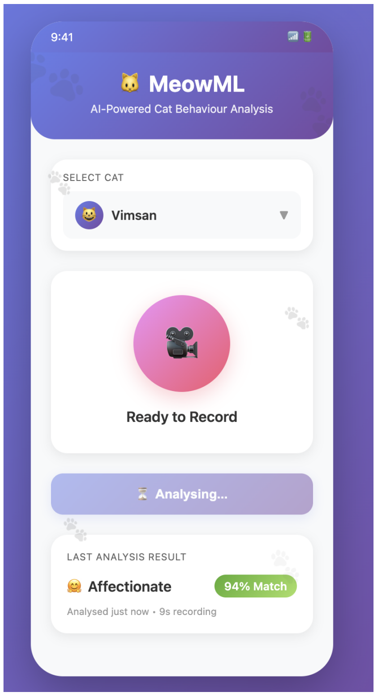
[31] Jong Wook Kim, Justin Salamon, Peter Li, and Juan Pablo Bello. CREPE: A Convolutional Representation for Pitch Estimation, February 2018. URL `http://arxiv.org/abs/1802.06182`. arXiv:1802.06182 [cs, eess, stat].

[32] models/research/audioset/vggish at master · tensorflow/models, . URL `https://github.com/tensorflow/models/tree/master/research/audioset/vggish`.

[33] Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289, September 2018. ISSN 1546-1726. doi: 10.1038/s41593-018-0209-y.

[34] Regulation - 2016/679 - EN - gdpr - EUR-Lex, . URL `https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng`. Doc ID: 32016R0679 Doc Sector: 3 Doc Title: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) Doc Type: R Usr_lan: en.

A Supplementary Figures



SELECT CAT

Vimsan

LAST ANALYSIS RESULT

Affectionate 94% Match

Analysed just now • 9s recording

Figure 3: Mobile application mock-up interface for cat behaviour analysis. In production the app records synchronised audio-visual data to assess behavioural cues.

B Methodological and Ethical Details

B.1 Addressing Limitations

Beyond the data collection variables mentioned in the main text, key limitations include cross-cultural differences in human-cat interactions patterns that may affect model transferability. Throughout our work we will continue data collection in different locales. For an eventual production-level application we would strive to include Explainable AI (XAI) technologies to explain the parameters for certain predictions.

Our behavioural annotation framework is deliberately iterative and data-driven. We are building comprehensive ethograms that integrate established research with novel patterns emerging from our multimodal analysis. This includes capturing context-dependent behaviours (e.g., how approach behaviour differs based on prior interaction history), relationship-specific signals (e.g., vocal prosody variations between familiar versus unfamiliar humans), and subtle welfare indicators that may not appear in existing ethograms (e.g., micro-expressions or postural changes preceding overt stress signals). By combining expert ethological knowledge with data-driven pattern discovery, we aim to create more nuanced and accurate behavioural classifications than those based solely on historical categorical systems.

The new datasets will be collected with the same setup as the one used in the CHC dataset [10]. Additionally, we will expand the dataset through systematic data collection across multiple contexts and environments, including homes, clinics, and shelters. This includes high-quality multimodal recordings (audio, 2D/3D video), annotated using expert-developed ethograms and acoustic–phonetic analysis tools (e.g., Praat, BORIS). These efforts are designed to ensure diversity in breed, age, familiarity, and interaction types, thereby supporting both model training and generalisation.

Our mobile application will incorporate an optional active learning component where users can contribute anonymised data to improve model performance. With explicit user consent, videos undergo automatic privacy protection (face masking, audio filtering) before expert review, creating a continuous feedback loop for model refinement while maintaining strict privacy standards.

B.2 Computational Requirements

Training the multimodal deep learning architecture requires GPU clusters with expectedly 300-400 GPU hours for full model development across all modalities. Inference would be optimised for mobile consumer-level hardware, and computationally heavy processing would be hosted on external servers.

B.3 Ethics and Consent Procedures

All human participants provide informed consent via GDPR-compliant forms detailing data collection, storage, and sharing procedures. Consent includes permission for research use and open dataset release under CC-BY-NC license. Participants complete recording sessions following standardised protocols with full procedure disclosure. Cat welfare is monitored throughout recordings with immediate cessation protocols if stress indicators are observed.

The project adheres to all ethical and professional considerations outlined by the relevant national Ethical Review Authorities, institutional policies, informed consent following the GDPR [34] and confirms to the ARRIVE guidelines for animal research.

B.4 Data Management and Asset Documentation

New datasets include comprehensive metadata covering recording conditions, participant demographics, and annotation procedures. Data sharing agreements specify research use permissions and anonymisation protocols. Code repositories will include detailed documentation, installation instructions, and reproduction scripts. Existing asset licenses are verified and documented: SuperAnimal (Apache 2.0), SLEAP (BSD-3), BORIS (GPL-3). All derived works maintain appropriate attribution and license compatibility.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and background state the main contributions; namely proposing a multimodal framework integrating acoustic, visual, and tactile signals for cat-human communication analysis using the largest expert-annotated dataset of its kind. The scope is expanded on in the paper focusing on multi-modal signals in domestic cats with potential for species generalisation and real-world applicability.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: The body of the paper does not discuss this, however Appendix B1 includes a small section on limitations, of which mitigation techniques would be expanded in a more comprehensive proposal.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: This is a workshop proposal outlining a research framework rather than presenting theoretical results requiring formal proofs.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example

     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification: No current access is provided since this is preliminary work. The proposal outlines the intention to provide open-source benchmarks and dataset upon completion of the work.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: As this is a proposal for future work without completed experiments, specific hyperparameters, training details, and exact methodological specifications are not yet available. The proposal outlines the intended approach.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: No experimental results are presented as this is a proposal for future work.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: As this is a proposal for future work without completed experiments, no computational resources have been evaluated to reproduce experiments. We include a small section on anticipated computation requirements in Appendix B2.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The proposal explicitly states adherence to ethical guidelines by the relevant national Ethical Review Authorities, institutional policies, GDPR compliance, ARRIVE guidelines for animal research, and emphasizes non-invasive recording methods to avoid animal stress in B3 of the Appendix "Ethics and Consent Procedure".

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

Justification: The proposal discusses positive impacts including improved animal welfare monitoring, early detection of health issues, and applications for shelters, veterinarians, and pet owners. It addresses potential benefits for millions of cats worldwide and cross-species applicability.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [No]

    Justification: The proposed cat communication analysis system does not pose significant risks for misuse. The focus on welfare assessment represents beneficial rather than potentially harmful applications. However we acknowledge the potential risk for misuse in cross-species application if our system would allow for the attraction of e.g. valuable endangered wildlife or harmful practices in factory farming. This will be addressed thoroughly in an expanded version of the proposal and further work.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

Justification: The proposal properly cites existing tools and datasets (SuperAnimal, SLEAP, AmadeusGPT, BORIS, CREPE, VGGish) and references the original papers. Additional statements on license details are addressed in Appendix B4 and would be addressed further during implementation.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: Full data collection procedures are outlined in the academic publications that accompany the assets in question.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [No]

Justification: Previous literature of the proposed assets include extensive details on participant instructions and participation agreements. Given the limitations of the presented short proposal, this is omitted in the current text to leave room for methodology.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: The proposal states adherence to the relevant review bodies both in section 2.1 of the body and Appendix B3. Additionally we have ethical approvals for previous data collection efforts that can be attached after the double-blind review process.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: The proposal mentions AmadeusGPT for behavioural sequence annotation and hypothesis generation, which proposes an important component of the semi-supervised annotation pipeline.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.