Narrative Review Article

Advances in AI-Assisted Decoding of Sperm Whale Vocalisations and Their Implications for Animal Communication

Alice Chavarot

Francis Holland School Sloane Square, 39 Graham Terrace, London, SW1W 8JF, United Kingdom

ABSTRACT

Sperm whales (Physeter Macrocephalus) are an endangered species with a highly complex vocalization system. This review summarizes widespread research on how artificial intelligence (AI) can be utilized to examine whether their vocalizations exhibit patterns of meaning compatible with human language. Sperm whales produce highly structured and varying click sequences that prompt debate about the extent of their cognitive and linguistic capacities, but technological and methodological limitations have hindered our efforts to examine these communications at scale. AI, particularly natural language processing within machine learning, enables researchers to identify acoustic patterns, classify call types, and map the complex contextual features of whale communication within their society. This paper evaluates current approaches, namely key researchers and initiatives such as Project CETI (Cetacean Translation Initiative), that apply techniques such as tagging, aerial drones, and tethered buoy arrays to further this research. This review also addresses the critical challenges within the field, namely, limited available datasets, inadequate contextual understanding, and anthropocentric bias. Through the evaluation of both the progress and constraints of AI-led bioacoustics, this paper highlights the potential for interspecies communication and the potential drawbacks it may have, such as specialized fishing and further disruption of sperm whale communities. Ultimately, the goal is not only to investigate the potential linguistic structure of sperm whale vocalizations and explore the possibility of interspecies communication, but also to consider how such findings could influence scientific understanding of animal intelligence, cognition, and human–animal relationships.

Keywords: Sperm Whale; Animal Cognition; Linguistics; Artificial Intelligence; Conservation, Biotechnology

INTRODUCTION

In the vast, blue depths of the ocean, sperm whales produce a cacophony of clicks never to be heard at the surface. Though scientists have been studying them since they were first named in the 18th century, we still have much to learn about these reclusive creatures. This species is not very easy to study; most of their lives are spent hunting, and they can dive to depths of over 10,000 feet (1). Their intricate vocalizations prompt debate about the extent of their cognitive and linguistic capacities. Now, with AI at their side, scientists are diving into the mysteries of whale communication, unraveling the secrets that are spun beneath the waves.

Though it is usually thought that humans are the most intelligent species, sperm whales have a biological advantage. With the largest brains of any animal on the planet (up to 20 pounds), their brains are six times the

Corresponding author: Alice Chavarot, E-mail: alicechavarot@icloud.com.

Accepted June 22, 2026
https://doi.org/10.70251/HYJR2348.43678686

size of the average human's. Furthermore, they are a highly social species, spending their lives in matriarchal groups of about 10 whales. They are renowned for their high level of communication in the form of codas, which also help in locating prey through echolocation. All of these factors contribute to the fascinating question of the whales' intelligence. Biologically, they have complex cognitive structures and highly developed social systems. Some studies have even shown differentiation in codas from different areas, like regional accents. For example, Caribbean and Pacific whales have different ranges of codas, but the "Five Regular" call — a regular pattern of 5 evenly spaced clicks — has a universal function as it is used by sperm whales worldwide. Additionally, there is evidence to suggest that certain coda types function as markers of clan identity, much like ethnic markers in humans (2). These features make sperm whales a useful model for computational approaches to studying animal communication.

This research could also lead to significant developments in conservation efforts. Whale populations are constantly under threat from habitat destruction, climate change, ship noise, and hunting. In fact, due to the spermaceti oil carried in their heads, sperm whales were the main target of commercial whalers in the 19th and 20th centuries (3). Between 1800 and 1987, approximately 1,000,000 sperm whales were killed worldwide by whalers. Today, it is estimated that there are around 200,000 and 1,500,000 around the world (4). Despite the 1986 ban against whaling for profit, countries like Japan, Iceland, and Norway continue the practice, and have killed almost 40,000 whales since. Their methods are cruel; whales are shot with exploding harpoons that burst when flesh is penetrated. Not all whales are immediately killed — some suffer for up to 30 minutes before death, as the harpoons cannot help but be imprecise due to the whale's size (5). With more information on their innate intelligence, measures can be placed to address current welfare concerns.

Furthermore, understanding their vocalisations can help indicate stress signals, which usually manifest as critical changes in behavioral patterns such as open mouths, agitation of pectoral fins, or side fluking (whales roll onto their back and vigorously slap the surface). However, these signals are highly variable; whales may react differently to the same situation due to individual personalities. (6). Thus, it may not be the most effective method for interspecies understanding, whereas vocalisations are constant and reliable.

Interpretation is likely to play an important role for designing effective conservation strategies, such as establishing marine protected areas and implementing noise reduction measures (7). By better understanding their needs and social behaviors, we can shift human perspectives from exploitation to coexistence, inspiring greater care for marine ecosystems.

However, despite rapid advances in bioacoustics and AI (artificial intelligence), significant challenges remain in interpreting sperm whale communication. While researchers have made substantial progress in analyzing vocal patterns, many questions regarding the function and meaning of these signals remain unresolved. This review therefore examines how AI methods originally developed for human language are being adapted to sperm whale vocalisations, and how their application expands across different species and ecological contexts before critically assessing the methodological limits of interpretation.

AI APPLICATIONS IN LINGUISTICS

Over the last decade, AI has made incredible advances in decoding and predicting linguistic patterns. Current AI systems can categorize basic phonetic details into phonemes and words, recognize grammatical structures, and relate words to visual perception (7). This means that AI can model statistical relationships between language and visual input, indicating pattern-based language modelling rather than simple replication. Though this basic understanding is relatively new, it has the potential to improve AI replication of vocalisations, though whether this extends to understanding of meaning remains unverified.

A foundational example of this approach is Large Language Models (LLMs), a type of neural network (computational models designed to mimic the function of the human brain), that can analyze vast collections of languages and reliably predict the next word. LLMs are extremely popular and encompass AI models such as ChatGPT, which generate human-like dialogue. However, these models manipulate the elements of a language without ever understanding its meaning. Therefore, though this introduces the possibility of reproducing 'sentences' which could be played back to whales, it would not give us an insight into what they are actually saying (8).

AI Applications in Sperm Whale Vocalisations

Although AI models are still developing, they have already been applied directly to sperm whale vocalisations In one experiment led by Shane Gero, the biology lead for Project CETI (Cetacean Translation Initiative) and founder of the Dominica Sperm Whale Project, neural networks successfully identified individual sperm whales based on their unique codas with 99% accuracy. Algorithms like Convolutional Neural Networks, which detect patterns in spectrograms (visualizations of whale sounds) and Recurrent Neural Networks, which track sequences over a span of data, were utilized for this project (9).

Further advancing this work, Gašper Beguš from the University of California, Berkeley, developed a deep learning model for human language called fiwGAN, whose purpose was to imitate codas and embed information into them. While fiwGAN was initially designed for modeling human speech, Beguš adapted it to analyze sperm whale vocalisations. He was then able to analyze a dataset of 3948 sperm whale codas, which proved that the whales consistently produced codas with specific frequency patterns analogous to human vowels. Furthermore, these codas either had a single frequency peak (called a formant) or two. The team termed these "vowels," with single-formant codas being a-vowels and double-formant codas being i-vowels, drawing parallels to human language in which vowels also differ in their formant frequencies. They also found rising and falling frequency patterns in the codas, comparing them to diphthongs (vowel sounds which glide from one to another, e.g., "coin") in human speech (10).

Cross-Species Communication Models

Beyond sperm whales, AI systems have also been applied to other species to explore the structure of their communications. In 2023, the Earth Species Project began testing playback experiments, playing artificially generated calls to zebra finches and observing how the birds respond. The goal is to pass the 'Turing test', the point at which humans, or in this case, birds, wouldn't be able to tell the difference between the machine and their own kind (9). This work builds on earlier non-AI interaction studies, such as a 2021 experiment involving an Alaskan humpback whale named Twain, where researchers manually played recorded contact calls and observed that the whale adjusted her call timing in response, suggesting responsive, temporally coordinated communication (12).

AI has also enabled significant advances in our understanding of pet cognition. Con Slobochikoff, author of *Chasing Doctor Dolittle: Learning the Language of Animals*, began his career studying prairie dogs and quickly gained an appreciation for the complex communicative ability of their alarm calls. He then worked as a behavioral consultant for disobedient dogs, and found that many of his clients had completely misinterpreted what their dogs wanted to convey. Our pets use multimodal signals, combining vocalisations with body posture, yet Slobochikoff notes that we fixate on the sound and disregard any other cues, such as tail position/movement, or ear position. He has been working to create an AI model which can translate pet behaviors into human language, but this still involves a lot of manual training, and the model is not yet fully autonomous. This means that researchers must actively teach the AI how to recognize and interpret different pet behaviors (11).

Similarly, in a study of pig vocalisations, an AI model was trained on how to detect positive or negative emotion. They found that short grunts indicated positive emotions, and long grunts often signaled discomfort. High-frequency sounds, such as squeals, occurred when the pigs were stressed. The algorithm demonstrated that pigs raised on free-range, humane farms produced fewer stress signals than others. The researchers believe this could be positively impactful on the farming industry, using AI to categorize farms so consumers can make informed decisions (13). This research suggests that AI may assist in identifying emotional states in animals through sound, a method that could be applied to whales and other species to better understand their well-being and communication.

AI in Ecological Monitoring and Interpretation Limits

Beyond linguistic modelling, AI is also widely used in ecological monitoring, where its role shifts from analyzing structured communication to identifying species and behavioral signals in large-scale environmental datasets. But, while AI excels at identifying intricate patterns within data, this does not necessarily translate into interpretable meaning in biological contexts. The decision-making processes of some advanced AI algorithms can be opaque, making it difficult to fully understand the reasoning behind their outputs. This lack of transparency can affect our confidence in the conclusions drawn from AI analysis, especially when dealing with nuanced and untested forms of communication. This highlights a broader methodological issue: interpreting AI outputs in biological contexts requires understanding the mechanisms by which predictions are generated, not just their accuracy.

This interpretability challenge becomes especially clear in applied systems such as automated species identification. A clear example of this is Benjamin Hoffman's work with the Cornell Lab of Ornithology, where he helped develop a unique app called Merlin Sound ID, which identifies bird species based on their call. The AI model converts the sound into a spectrogram, which is then compared to Cornell's audio library. The model then makes a guess on what species it is and is then checked with eBird, Cornell's global database, to ensure that the supposed species would be present in the user's location. This method has reliably identified over 1000 bird calls (9). However, Hoffman notes that Merlin was designed for species identification rather than cognitive research, highlighting the specialized nature of current AI tools.

Taken together, these examples demonstrate the breadth of AI applications across species and contexts, while also underscoring its current inability to infer intent or meaning.

ASPECTS OF WHALE SPEECH

In order to understand AI's current limitations, it is essential to consider the structure of sperm whale vocalisations.

Physeter macrocephalus, the sperm whale's scientific name, is referred to as an "animal of extremes." They are the largest of the toothed whales, can dive the deepest, and are found almost globally. Their colossal nose is home to the world's most advanced sonar system and controlled by the world's largest brain. Since they live in matriarchal pods, these whales experience a more prolonged parental care, allowing for greater transmission of learned information via communication (7). Whale communication, particularly among highly intelligent species like sperm whales, is incredibly structured and learned in a manner similar to human language acquisition. Young whales develop their codas by imitating the vocalisations of adult whales, much like how toddlers learn to speak by mimicking their parents (9).

Furthermore, though whale communication might seem quite arbitrary on the surface, researchers have discovered that whales modulate the tempo of these codas by speeding up or slowing down their delivery. Some whales also add extra clicks at the end of codas, a feature likened to musical ornamentation. This ornamentation has been observed to be more likely to appear at the beginning or end of a coda, and usually precedes a change in multi-whale communication, such as ceasing vocalisations or a change in vocalization partner (14). Broadly, this suggests that whales communicate not just to express basic needs or emotions, but to coordinate social interactions, maintain relationships, and possibly share information within their group. These subtle variations suggest that whale vocalisations may carry a more complex and nuanced system of meaning than previously believed, specific to individual personalities and reliant on contextual clues.

In May 2024, Project CETI were able to use AI to identify a sperm whale phonetic alphabet. It was previously thought that sperm whales had just 21 coda types. However, after studying almost 9,000 sperm whale communication recordings, the CETI researchers identified 156 distinct codas. They also observed that the building blocks of these codas function much like phonemes—the basic units of sound in human language that combine to form words (1). For example, the difference between *"hat"* and *"cat"* lies in a single phoneme, yet this small change completely alters the meaning. These findings provide a structural description of codas, but they do not by themselves resolve how such structures function within whale social systems, which remains an open question.

Although today the sperm whale's complex bioacoustic system is most famously used for communication, it evolved as a powerful tool for echolocation to allow these creatures to find prey and navigate the oceans. Echolocation clicks differ immensely from vocalization clicks, with each having an intense first pulse followed by a few pulses of declining amplitude. Echolocation clicks are short, broadband pulses used primarily for navigation and hunting. The whale emits these clicks and listens for the echoes that bounce back from objects or prey in their environment, allowing them to determine the distance, size, shape, and density of what is around them. In contrast, vocalization clicks (codas) are patterned sequences of clicks used for social communication among sperm whales. These are distinct from the more regular clicks used in echolocation, with ornamentation and possible regional dialect variations. (15). This multi-level click is due to the complex structure of the whale's *spermaceti* organ (Figure 1).

Sound generation begins when air is forced through the phonic lips at the front of the nasal complex, creating vibrations that travel through the *spermaceti*, a cavity filled with almost 2000 liters of wax. These vibrations are then reflected off the frontal sac and pass through the junk, a fatty tissue structure that focuses and projects the clicks as directional sound waves (7). This composite system allows for significant variation in sounds, producing the complex diversity in vocalisations that whales exhibit.

This anatomical capacity for precise sound production helps explain why recent studies have identified a far greater diversity of sperm whale codas than was previously recognized. However, while the whales' bioacoustic system clearly allows for large variation in vocal output, there remains a significant gap in linking specific vocal patterns to behavioral outcomes. Most existing datasets lack depth and individual specification to establish whether variations in codas correspond to changes in social role, emotional state, or environmental context. Addressing this gap requires datasets that integrate vocalisations with the intricate behavioral, spatial, and social data collected over extended time periods.

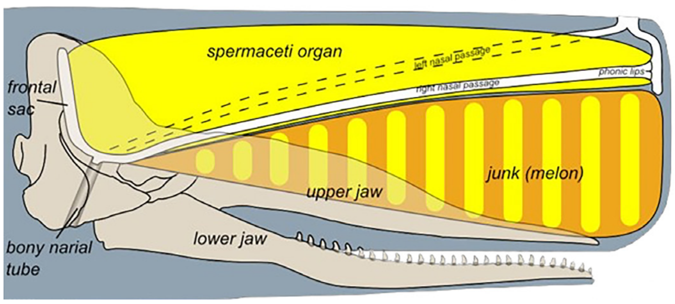

Figure 1. *Diagram of sperm whale head anatomy (cutaway view), illustrating the major structures involved in sound production and echolocation. The phonic lips located in the nasal complex generate vibrations when air is forced through them. These vibrations pass through the spermaceti organ, a large wax-filled structure that functions as an acoustic medium for transmitting sound. The surrounding fatty tissues, including the junk, help to focus and project the sound waves into the surrounding water as highly directional clicks used for both echolocation and communication. The frontal sac reflects sound internally, contributing to signal amplification and directionality (7).*

WHAT ARE THE BENEFITS OF UNDERSTANDING THEM?

From a scientific perspective, decoding sperm whale codas could provide valuable insights into the evolution of complex communication systems, potentially deepening our understanding of language, intelligence, and social structures in non-human species, which can be carried forward to other animals in the future. This research could also enhance machine learning and AI models designed for pattern recognition, benefiting all fields in the development of AI.

Ecologically, a deeper understanding of sperm whale behavior can improve conservation efforts by identifying distress signals, migration patterns, and the impacts of human activities such as ship noise and climate change. Hopefully, with a scientific understanding of whale emotions, rather than anthropomorphic interpretations, action can be taken to improve the relationship between humans and animals. For sperm whales, recognizing meaning behind communication could foster a greater sense of empathy and cause commercial whalers to rethink the exploitation of these animals for pharmaceutical purposes, pet food, or tourism as a rare dish. Unlocking whale communication challenges our anthropocentric views on intelligence and could reshape the way we relate to other species. In the long term, human-whale communication could put an end to the thousands of whales killed annually and lead to more effective marine policy or coexistence strategies.

WHALE VOCALISATIONS COMPARED TO HUMAN LANGUAGE

Many similarities exist between human and animal language, particularly in early vocal development. The cries of mammalian children can be so alike that white-tailed deer respond to distress calls from marmots, humans, and seals alike. Not only deer, but birds often respond to the alarm calls of other bird species, even those they are not closely related to. The sharp, sudden onset of these calls triggers vigilance or escape behaviors. Additionally, vocal learning follows similar patterns across species—harbor seal pups, like human infants, adjust their pitch to capture a parent's attention. Additionally, some bat species have distinct "contact calls" that individuals use to identify themselves to group members. These calls can vary geographically, suggesting a form of vocal learning and "dialect" acquisition, much like regional accents in human language. (9).

Whales and humans have also been discovered to have specific similarities. Scientists have identified a foundational set of distinct sound units, phonemes. These phonemes are not used in isolation but are combined in various sequences and patterns. This combinatorial capacity allows sperm whales to construct a diverse repertoire of unique vocalizations, suggesting a potentially complex system for conveying information. The presence of a combinative coding system is seen as a prerequisite to the 'duality of patterning' previously thought to be unique to the human language (1). This phenomenon describes how the fundamental units of language, called words or morphemes, are composed of smaller, meaningless units called phonemes that serve to differentiate meaningful elements from one another (17).

However, when attempting to interpret these apparent structural similarities using computational tools, a key limitation emerges in the assumptions built into current AI systems. Most natural language processing options, including LLMs, are optimized for human language and assume properties such as compositional syntax, semantic grounding, and a linear structure. These assumptions risk mischaracterizing whale vocalisations, which may rely more heavily on rhythm, timing, repetition, and social context than on linear syntactic structure. Consequently, successful pattern prediction should not be confused with semantic understanding. This represents a non-obvious but critical flaw in current AI-based approaches: models may accurately reproduce vocal patterns while misunderstanding what those patterns mean.

However, it can be argued that animal communication systems cannot be compared to human ones. Human languages are expressed through the variety of structure features such as displacement (the capability of language to express things which are not immediately present), productivity (the ability to create and understand new words and sentences using existing language rules), reflexivity (the ability of language to be used to talk about itself), and recursion (occurs when phrases or clauses are embedded within others, creating infinitely complex sentences). Some argue that such a rich construction is not yet present in the systems of animal languages, and therefore, interspecies comparison cannot exist (7).

CURRENT METHODOLOGIES FOR DATA COLLECTION

A key factor in AI's success has been the availability of large datasets. The more extensive and diverse the dataset, the better AI can analyze the full range of a species' communication. Today, many technologies are available to collect this data. One method is tethered buoy arrays, which can record background bioacoustic signals. These systems typically consist of sensors spaced several hundred meters apart, extending from the surface down to approximately 1200 meters—the depth at which sperm whales hunt. By equipping each mooring with multiple sensors and deploying multiple moorings, these arrays can help localize and track whale movements. Their main advantage lies in their reliability and ability to continuously record acoustic signals over a wide area.

Other methods include tags, which are recording devices attached to the whales, able to measure not only acoustics, but also pressure, temperature, and movement patterns. However, these tags have limitations regarding the reliability of the data recorded and the effectiveness of their adhesion onto the whales. Some of these issues can include sensor malfunction, calibration issues, or battery life limitations.

Aquatic drones can be used as a passive method of obtaining audio and video recordings of the ocean ecosystem. These drones can range in technological advancements, from free-floating cameras to autonomous, self-propelled navigation robots, like small submarines. They offer the possibility of capturing essential data in a non-invasive manner.

Finally, aerial drones can survey vast areas to monitor whale populations and have a load weight of several kilograms to enable tag deployment. These drones are capable of landing and taking off on the surface, and flying for up to 30 minutes to deploy tags to remote areas to facilitate tracking (7). Each method has its strengths and weaknesses, but the most comprehensive datasets can be found by combining them, and the best methods depend on what researchers are specifically looking for (*e.g.* tagging for monitoring individual whale's movements, aerial drones for studying population distribution).

However, a limitation of this data collection is what is known as the "cocktail party problem" (named by British scientist Colin Cherry in 1953). This defines the challenge of isolating individual vocalisations from an environment, and it is an issue which has significantly opposed developments in the industry. In the past, biologists have been forced to disregard masses of data due to overlapping vocalisations which made it impossible to study a single voice. In 2021, the Earth Species Project was able to create a neural network called BioCCPNet to separate overlapping sounds into individual tracks and filter out any background noise. It achieves this by translating the audio into a visual representation which the network then uses to determine which pixel belongs to which speaker, then reconstructing the individual sounds. They "trained" BioCCPNet by creating synthetic audio, combining a multitude of singular vocalisations into a track which mimics a real environment. They used a multitude of species to test this algorithm, from sperm whales to Egyptian fruit bats and dogs. This enabled them to check the accuracy of the network and make constant improvements (9); however, this system is still not unfailingly accurate. Research so far has been primarily focused on human speech, and even then, there is some omission of speech properties such as speaking style, linguistic variability, and contextual information (18). However, BioCCPNet could provide a fundamental foundation on which to build upon in regards to animal species communication.

Furthermore, it is important to realize the negative effects that this research could have if successful. Commercial fishing organizations may use AI for "precision fishing"; listening for specific schools of fish and impersonating their calls to lure them in. Though this is largely theoretical for the moment, there are current interpretations in the form of fish stock assessment, which uses datasets to predict fish distribution and abundance to calculate sustainable fishing levels (19). Though this is a positive application, it can quickly be used negatively. Furthermore, especially during testing stages, it is difficult to know if introducing synthetic calls into the marine ecosystem could cause disruption and stress to the animals. Scientists urge us to create appropriate legislation to inhibit the exploitation of these animals. "Every time you invent a technology, you also invent a responsibility," said Aza Raskin, one of the founders of the nonprofit Earth Species Project (9).

Beyond questions of interspecies dialogue, there are also urgent environmental factors—like climate change and underwater noise pollution—that threaten the very survival of these communicative creatures. According to the International Union for Conservation of Nature, sperm whales are classified as "vulnerable." As with all species targeted by whalers, their blubber was removed and rendered into whale oil, which was used as a lamp fuel for most of the 16th and 17th centuries. However, sperm whales had other uses, too; they were specially targeted due to the spermaceti oil, which gives them their name (3), and also for their ambergris, an excretion used in perfumery and extravagant cuisine (20). Because of this, at the start of the nineteenth century, their population was estimated to have exceeded 1,110,000 individuals, yet by 2022, long after commercial whaling had ended, the population was thought to be only around 850,000, though the numbers vary widely due to uncertain population distribution. (21). One of the possible dangers of this project is further detrimental impact to this species due to the collection of data, through which we may be putting these animals further at risk of harm.

CONCLUSION

This review demonstrates that AI has significantly enhanced our ability to examine the structure and patterns that make up sperm whale vocalisations. While the vocalisations show complexity that could compare to human linguistic features, AI analysis has not yet come far enough to translate these patterns into meaning. The main takeaway is that sperm whale communication is highly structured, socially learned, and variable across individuals and regions. These findings highlight the remarkable cognitive and social capabilities of the species, while also revealing the current limits of our capacity for interpretation.

These findings have broad implications for both animal cognition research and broader conservation efforts. Understanding the complexity of whale communication could shift human perspectives, encouraging greater empathy and fostering more informed conservation strategies. The methods developed through AI analysis may also be adapted to other species, provisioning a framework for studying communication and cognition across species. Additionally, these approaches could influence technological development in AI by refining pattern recognition and data analysis techniques which can be applied to all other fields.

Despite these advances, significant limitations remain. AI tools are currently restricted to identifying patterns and cannot reliably interpret semantic meaning. The size, quality and contextual depth of available datasets are not enough to fully generalize findings to all sperm whale populations. Variations in social context, regional dialects and individual behavior are extremely difficult for AI models to capture, and current methods may misrepresent or oversimplify complex expressions. These limitations suggest that while AI is an immensely powerful tool, it should be applied cautiously, whilst recognizing the gaps and potential mistakes in current analyses. Furthermore, while similarities to linguistic features such as phenome-like units have been identified, there is not enough evidence to conclude that sperm whale vocalisations convey meaning in a way that can be directly compared to human language. This distinction is critical, as it highlights the risk of anthropocentric bias when interpreting animal cognition.

To address these limitations, future research should prioritize the collection of larger, more diverse datasets that integrate vocal, behavioral and social information. Studies could track individual whales over longer periods of time to better understand learning and the social transmission of communication. Arguably most importantly, developing AI models capable of combining many inputs would provide a more holistic understanding of whale behavior. Furthermore, interdisciplinary collaboration between marine biologists, linguists and AI engineers will be essential to refine methods and expand the scope of possibilities for interpretation, ultimately allowing more meaningful insights into interspecies communication.

Finally, research in this field carries substantial ethical responsibility, given the vulnerability of sperm whales and their history of exploitation. This review emphasizes the importance of conducting research in a manner that minimizes harm to whale populations by reflecting on the potential risks of both data collection and AI interventions. The methods and recommendations presented should guide future studies to be more ethically considerate, balancing scientific breakthrough with conservation priorities. Understanding the ethical dimension also ensures that advances in this field do not inadvertently facilitate exploitation or direct harm to the species, but instead promote coexistence and respect for all non-human animals.

Therefore, in light of the scientific, ethical, and ecological considerations explored in this review, AI-driven analysis of sperm whale vocalisations represents a promising but still advancing avenue for our understanding of non-human communication. Sperm whales, with their sophisticated brain structures, heightened intelligence, complex social behaviors, and unique click-based communication system, serve as an ideal model for examining how structured communication systems can emerge outside of human language. The research discussed in this paper demonstrates that AI can reliably identify recurring patterns, classify codas and reveal combinatorial structures within communication, indicating a level of organization that exceeds simple signaling.

Ultimately, this review shows that AI has the capacity to deepen scientific insight into sperm whale communication and animal cognition as a whole. The final step in this process, once we have understood the method, is for us to contact them. Such a development would raise significant ethical questions regarding the purpose and consequences of this outreach, including how humans should engage with other species whose communication structure and life experience differ fundamentally from our own. Any form of interaction would require immense consideration of conceptual translation, as meanings grounded in human frameworks may not translate to a species whose communication is shaped by acoustic perception and long-term matriarchal structures. As a result, the possibility of such an exchange challenges us to rethink not only how we define intelligence and language, but also our role as humans and our responsibility to protect interspecific harmony on this planet.

ACKNOWLEDGEMENTS

The author would like to express my gratitude to Ambika Verma for her support and mentorship throughout the preparation of this paper.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

REFERENCES

1. Bressanin A, Latham K. The sperm whale ‘phonetic alphabet’ revealed by AI. BBC. Available from: https://www.bbc.co.uk/future/article/20240709-the-sperm-whale-phonetic-alphabet-revealed-by-ai (accessed on 2025-6-10)
2. Hersh TA, Gero S, Rendell L, Cantor M, Weilgart L, Amano M, *et al*. Evidence from sperm whale clans of symbolic marking in non-human cultures. *Proc Natl Acad Sci U S A*. 2022; 119 (37): e2201692119. DOI: 10.1073/pnas.2201692119
3. Sperm whale. NOAA Fisheries. Available from: https://www.fisheries.noaa.gov/species/sperm-whale (accessed on 2025-6-13)
4. Sperm Whale Species Profile, Alaska Department of Fish and Game. Available from: https://www.adfg.alaska.gov/index.cfm?adfg=spermwhale.printerfriendly (accessed on 2025-6-13)
5. McMillan L. Is whaling cruel? Whale and Dolphin Conservation. 2025 Mar 14. Available from: https://uk.whales.org/2025/03/14/is-whaling-cruel/ (accessed on 2025-5-7)
6. Blasi MF, Caserta V, Bruno C, Salzeri P, Di Paola AI, Lucchetti A. Behaviour and vocalisations of two sperm whales (Physeter macrocephalus) entangled in

illegal driftnets in the Mediterranean Sea. *PLoS One*. 2021; 16 (4): e0250888. DOI: 10.1371/journal.pone.0250888

7. Andreas J, Beguš G, Bronstein MM, Diamant R, Delaney D, Gero S, *et al*. Toward understanding the communication in sperm whales. *iScience*. 2022; 25 (6). DOI: 10.1016/j.isci.2022.104393
8. Drösser C. Workshop report: Decoding communication in nonhuman species II. Simons Institute. 2023 Jul 25. Available from: https://simons.berkeley.edu/news/workshop-report-decoding-communication-nonhuman-species-ii (accessed on 2025-8-9)
9. Parshley L. Artificial intelligence could finally let us talk with animals. Scientific American. 2023 Oct 1. Available from: https://www.scientificamerican.com/article/artificial-intelligence-could-finally-let-us-talk-with-animals/?utm_source=substack&utm_medium=email (accessed on 2025-8-9)
10. Pearson J. Scientists have reported a breakthrough in understanding whale language. Vice. 2023 Dec 7. Available from: https://www.vice.com/en/article/scientists-have-reported-a-breakthrough-in-understanding-whale-language/ (accessed on 2025-6-13)
11. van Hooijdonk R. This real-life Dr Dolittle uses AI to translate animal language. Available from: https://blog.richardvanhooijdonk.com/en/this-real-life-dr-dolittle-uses-ai-to-translate-animal-language/ (accessed on 2025-6-13)
12. McCowan B, Hubbard J, Walker L, Sharpe F, Frediani J, Doyle L. Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "conversing" with an Alaskan humpback whale. *PeerJ*. 2023; 11: e16349. DOI: 10.7717/peerj.16349
13. Gronholt-Pedersen J. AI decodes oinks and grunts to keep pigs happy. Reuters. 2024 Oct 24. Available from: https://www.reuters.com/technology/artificial-intelligence/ai-decodes-oinks-grunts-keep-pigs-happy-2024-10-24/ (accessed on 2025-5-9)
14. Sharma P, Gero S, Payne R, Gruber DF, Rus D, Torralba A, Andreas J. Contextual and combinatorial structure in sperm whale vocalisations. *Nat Commun*. 2024; 15 (1): 3617. DOI:10.1038/s41467-024-47221-8
15. Jiang J, Li S, Wang X, Li C, Sun Z, Liu W, *et al*. Study of the relationship between sound signals and behaviors of a sperm whale during the hunting process. *Appl Acoust*. 2021; 174: 107745. DOI: https://doi.org/10.1016/j.apacoust.2020.107745
16. Sperm whale head anatomy (skull cutaway). Adapted from Kurzon, Sperm whale head anatomy.svg, Wikimedia Commons (CC BY-SA 3.0). Licensed under Creative Commons Attribution-ShareAlike 3.0 Unported (https://creativecommons.org/licenses/by-sa/3.0/deed.en). Changes: unchanged. https://commons.wikimedia.org/wiki/File:Sperm_whale_head_anatomy_%28skull_cutaway%29.svg (accessed on 2025-5-9)
17. Ladd DR. What is duality of patterning, anyway? *Lang Cogn*. 2012; 4 (4): 261–273. DOI: 10.1515/langcog-2012-0015
18. Bronkhorst AW. The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten Percept Psychophys*. 2015; 77 (5): 1465–1487. DOI: 10.3758/s13414-015-0882-9
19. Stock assessment model descriptions. NOAA Fisheries. Available from: https://www.fisheries.noaa.gov/insight/stock-assessment-model-descriptions (accessed on 2025-8-2)
20. Smith P. A taste of edible feces. Smithsonian Magazine. 2012 Jun 4. Available from: https://www.smithsonianmag.com/arts-culture/a-taste-of-edible-feces-111178660/ (accessed on 2025-8-2)
21. Whitehead H, Shin M. Current global population size, post-whaling trend and historical trajectory of sperm whales. *Sci Rep*. 2022; 12 (1): 19468. DOI:10.1038/s41598-022-241