Papiers d'actualité / Current Affairs in Perspective

N°2 | February 2026

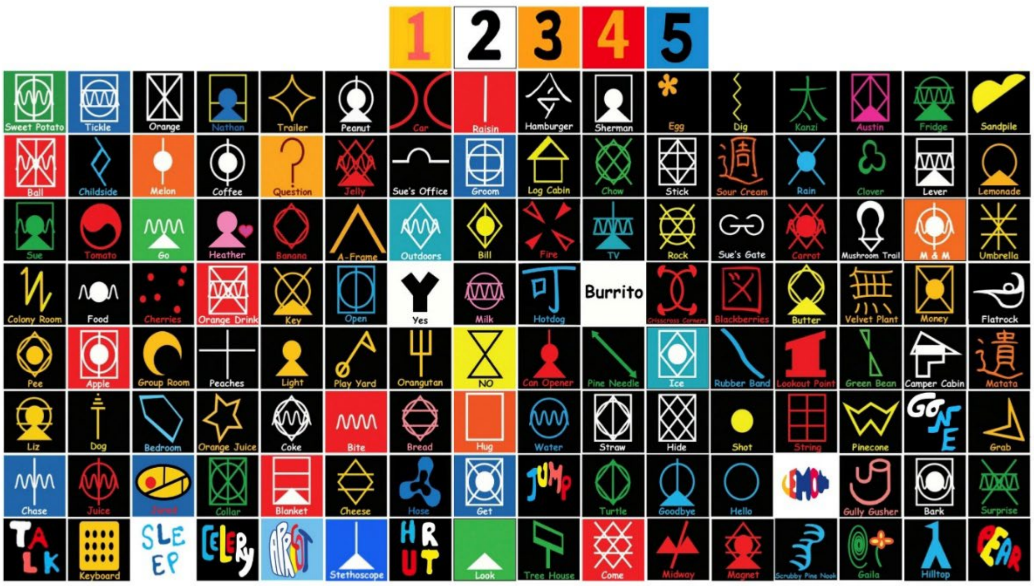
"AI" for Deciphering Animal Communication:
The Geometry of Meaning

Fondation Pierre du Bois
pour l'histoire du temps présent

Paul Deshusses*

"AI" for Deciphering Animal Communication: The Geometry of Meaning

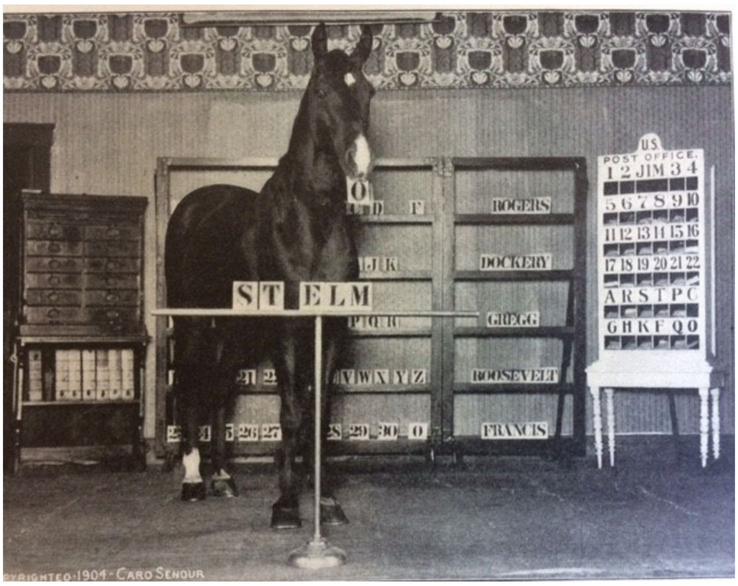
In 2025, the bonobo Kanzi, regarded as the ape most capable of communicating in English, passed away. In his later years, Kanzi communicated using a computerized lexigram keyboard system, mastering symbols reportedly on the order of several hundred. [i] By watching his mother being taught lexigram by her trainers, he learned first on his own, silently, attentively, without being addressed.[ii] Before anyone set out to train him, he had already acquired a large vocabulary of lexigram symbols, and, unprompted, began to "speak" back when he was two and a half years old, having watched his mother learn for over two years on a laminated plastic lexigram board.[iii]



Lexigram keyboard used by the bonobo Kanzi. A spatially organized system of arbitrary visual symbols employed in studies of symbolic communication and language comprehension in great apes. Adapted from Savage-Rumbaugh et al. (1998).

To teach English to a grey parrot born in 1976, named Alex, Irene Pepperberg developed what became famous as the Model–Rival technique, using a human proxy this time, not unconsciously the mother of the protagonist as with the case of Kanzi the bonobo.[iv] Another trainer interacted with Pepperberg in front of Alex. The proxy modelled correct answers to questions ("What colour?" "What shape?").[v] The proxy also acted as a rival: competing with Alex for the trainer's attention and rewards. Both Kanzi the bonobo and Alex the parrot answered novel questions**,** not rehearsed scripts, and even managed to ask questions. In front of a mirror, Alex asked 'what colors' and heard the response 'grey'.[vi] Pepperberg's insistence on this design was a response to a century of suspicion. Scholars have pointed out that animals can respond to human queries to please the trainer without understanding the content of the response. Critics have described this phenomenon as the "Clever Hans effect".

At the beginning of the twentieth century, a horse named Clever Hans appeared to shatter the presumed boundary between human and animal cognition. Hans could, it seemed, count, perform arithmetic, identify musical intervals, and answer questions by tapping his hoof a precise number of times. The phenomenon caused a sensation in Germany and beyond. Importantly, Hans's owner did not claim trickery; he believed the horse genuinely understood numbers. Later experiments have demonstrated that Hans was not calculating. Instead, the genius equine was exquisitely sensitive to involuntary human cues: subtle facial tension, breathing, or eye focus. When asked a mathematical question, he will stop to stamp his hoof at the climax of the public excitement, managing to interpret changes in posture, facial expression and gaze to receive applause and praise. When the questioner did not know the answer, Hans failed. When visual access to the questionnaire was blocked, Hans failed again. In a sense, the intelligence of Hans was real; his ability to read unconscious subtle cues was far better than that of humans themselves, and his attunement to human emotions was absolutely incredible. Yet it was not the type of intelligence tested by the trainers.



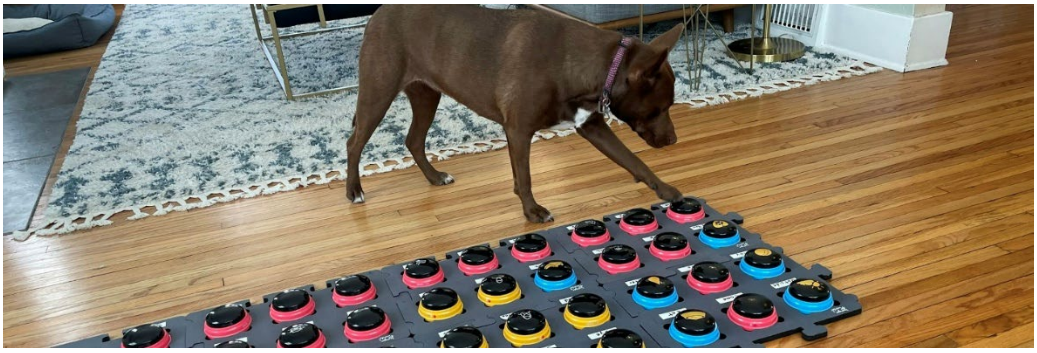
Clever Hans in Berlin, 1904. Photograph by C. Aro Senour. Public domain.

The lasting importance of the Clever Hans case lies not in debunking animal intelligence, but in exposing a methodological danger: the projection of human meaning onto animal behaviour through unintentional signaling.[vii] Ever since, "the Clever Hans effect" has served as a warning label in comparative cognition, psychology studies also warn of 'desirability effect', in ethology, of course, but also in AI studies. When an AI 'hallucinates' and provides false answers, it is also said that it falls for the clever Hans effect. Trainers teaching sign language to apes in the 1960s have repeatedly faced this critic, notably by the (bio)semiotician Thomas Sebeok and biolinguist Noam Chomsky.[viii] Pepperberg's insistence on this design for training Alex the parrot was a response to a century of suspicion.

Trainers of Kanzi were also cautious not to fall into repeating a 'clever Hans' phenomenon. Allen and Beatrix Gardner argued that chimpanzees such as Washoe could acquire meaningful elements of American Sign Language (ASL) and they raised Washoe, a chimpanzee, in the 1960s–1970s who indeed could master around 350 signs. Roger Fouts, their student, later reported anecdotal observations of American Sign Language-like hand movements during sleep in chimpanzees such as Mukobi; these observations were never subjected to controlled experimental verification but are featured in his book *Next of Kin* (1997), a popular science memoir. If an ape (here named Mukobi) is seen reproducing some signs during sleep, hence providing involuntary information about the content of their dreams by moving a finger in her sleep, signaling she wants this or that, it certainly does not fall into the category of a clever Hans phenomenon. Kanzi the bonobo remains distinctive because his symbolic behavior appeared to emerge without direct pedagogical intent, through prolonged exposure to his mother's interactions with lexigrams. This does not imply an emotionally neutral environment: like Alex, Kanzi lived in a socially rich setting. Alex, in particular, received sustained attention, affection, and social engagement from Irene Pepperberg and her collaborators, and this close bond was integral to the Model–Rival method. These methods are nowadays marginal when it comes to animal communication training due to the fear of anthropomorphizing animals. If tactile tablets and recording instruments played a significant role in animal communication research targeting exceptional animals such as Kanzi, the bonobo, and Alex, the parrot, the possibility of "participatory science" through online posting of animal observations by anyone could also yield groundbreaking outcomes. When the most famous ethologist of his time, if not of all, Konrad Lorenz, remarked in the 1970s that he had "never published a graph", he was articulating a style of science grounded in observation and narrative.[ix] That era seems to have drastically changed passed the advent of quantitative analysis of signal structure.

Participatory Science: Dog Talking by Pressing Buttons and Whale FM

For instance, the participatory "talking buttons dog project" emerged from informal home experiments in 2019 and was formalised into a global participatory study in 2020.[x] The project rapidly scaled from viral curiosity to systematic science. Anyone could film a dog pushing buttons making sounds such as 'go outside' or 'eat' to receive treats.[xi] In the dog project, thousands of owners worldwide contribute button-press data that researchers can analyze. A toy for dogs triggered the interest of researchers from the University of San Diego.



Domestic dog interacting with programmable soundboard buttons used in augmentative interspecies communication experiments. © Christina Hunger / Hunger for Words. Used for scholarly analysis.

In the whale-communication domain, interdisciplinary teams compile vast acoustic datasets and engage broader publics in listening, interpreting, and even challenging assumptions about interspecies dialogue. Everyone could upload whale sounds and photos on a website, and by 2011, the whale-song database Whale FM (then the world's largest public whale-call dataset) was placed on Zooniverse, prompting a volunteer labelling campaign "in which over 10,000 participants generated nearly 200,000 labels for 4,000+ calls." [xii] [xiii]

Promotional banner from Zooniverse, "Harnessing the Power of the Crowd." © Zooniverse. Used for scholarly analysis.

What unites the San Diego participatory dog experiment and the whale labelling campaign is not a promise of fluent translation; meaning is no longer extracted by experts alone, but assembled collectively, from fragments of sound, gesture, and pattern.

In media, AI is often described as a proxy, a mediator between animal and human, an algorithmic model that generates patterns, firing hopes of a 'Rosetta stone moment'. If Kanzi learned first by watching his mother and Alex by listening to conversations between two trainers, and dogs had to use buttons to 'speak', what if AI could play the role of mediator?

It goes without saying that efforts to use (so-called) artificial intelligence to interpret animal communication belong to a long historical arc rather than a sudden technological rupture, an arc that stretches from mid-twentieth-century bioacoustics and cybernetics to today's deep-learning systems trained on planetary-scale datasets to mention one moment of this arc. Animals do not merely "speak differently"; they perceive, prioritize, and inhabit reality through sensory channels and temporal scales that human concepts only partially grasp. The historical novelty of current projects lies precisely here: AI enables a form of inquiry that no longer depends on direct semantic analogy with human language, yet still inherits human assumptions about structure, information, and meaning. AI systems excel at detecting statistical regularities, repeated motifs, structural hierarchies, contextual shifts. Building on decades of ethological observation and bioacoustic analysis, machine learning techniques now make it possible to detect patterns in animal signals at scales previously inaccessible. Animal sounds data are fed to AI and AI generate sounds and sequences broadcasted in return to animals. In fact, synthesized signals have been played back to animals to probe behavioral response.

Symbolic Translation and Geometric Translation: Toward the Universal Translator

The history of animal communication research reveals a persistent tension between two epistemological regimes, here caricatured a bit. The first is symbolic translation**,** inherited from linguistics and cryptography: communication is assumed to consist of discrete units (words, signals, messages) that can, in principle, be decoded and mapped onto equivalents in another system. This regime animated early dolphin and whale research in the mid-twentieth century, from John C. Lilly's hopes for interspecies dialogue to later popular imaginaries of "speaking dolphin". Its strength lies in clarity and ambition; its weakness is anthropocentrism.

Symbolic translation presupposes shared referentiality, aligned intentions, and comparable ontologies of meaning. John C. Lily opens his 1961 bestseller book *Man and Dolphin: Adventures of a New Scientific Frontier* with this prediction:

> "Within the next decade or two the human species will establish communication with another species: nonhuman, alien, possibly extraterrestrial, more probably marine; but definitely highly intelligent, perhaps even intellectual. An optimistic prediction, I admit."[xiv]

Lily's work on dolphins in the late 1950s and early 1960s treated cetacean vocalizations as candidates for linguistic decoding, explicitly borrowing metaphors from cryptography and foreign-language learning. His experiments assumed that dolphin communication might possess discrete units, combinatorial rules, and referential meaning analogous to human language. He managed to secure funding from the NASA to study dolphins arguing that they could serve as proxies for potential future extraterrestrial encounters during the 1960s US (after all he studied dolphins during a time that some historians label as "the Space Age"). This approach rested on a symbolic epistemology: communication was presumed to consist of messages awaiting decryption.

The second methodological method, or assumption, or regime, now ascendant, is geometric alignment**.** Here, understanding does not proceed by decoding messages but by mapping structures: distributions, regularities, transitions, and constraints. Meaning is inferred not from equivalence but from *position*—from how signals relate to one another and to contexts of use.

When AI embeds whale codas, bird songs, or bat calls into high-dimensional spaces, it does not ask what a signal "means," but how it is situated relative to others: which signals cluster, which predict subsequent behavior, which change across social or ecological contexts. Geometric alignment accepts incommensurability as a starting point rather than a failure. It replaces the fantasy of translation with a practice of approximation—one that can be empirically tested without claiming semantic closure.

This insight has a longer intellectual history than is often acknowledged. Structural linguistics in the mid-twentieth century, from Saussure to Jakobson, already insisted that linguistic units derive value not from intrinsic reference but from their relations within a system.[xv][xvi] Later, cyberneticians and early AI researchers—working on pattern recognition, information theory, and self-organizing systems—formalized the idea that structure could precede interpretation. What recent advances in machine learning have done is to render these relational theories operational at scale, embedding them in high-dimensional mathematical space.[xvii]

In human natural language processing, this approach has yielded a striking result: semantic regularities appear as stable geometric transformations. The relationship between "king" and "queen," or between "Paris" and "France," is not stored as an explicit rule but as a vector i.e. an orientation in space that recurs across contexts. The now-canonical demonstration that semantic analogies can be recovered through vector arithmetic—such as *king − man + woman ≈ queen* or *Paris − France + Italy ≈ Rome* was first explicitly reported in 2013 by Mikolov and colleagues, who showed that independently learned word embeddings encode stable relational regularities in geometric form across language.[xviii] If words in Japanese and Farsi are visualized in a cloud of dots, the distance between the dot ' Paris' and 'France' will be the same in Japanese and in Farsi because these words have a relationship between them that is almost the same in Farsi or Japanese. If we measure the distance between these two dots, it will be roughly the same. These consistencies are not programmed; they arise from exposure to vast corpora of usage. Language, in this framework, becomes navigable terrain rather than a dictionary, a Borgesian library of Babel where geometric shapes created by vectors and points have replaced books. The idea that meaning might be spatial rather than symbolic has a long and uneven history, one that predates both digital computation and contemporary debates over artificial intelligence.

This framework has particular historical significance because it loosens the bond between meaning and language. Here concepts can exist without words, and therefore without humans. For animal cognition, this implies that a predator, territory, or social hierarchy can be represented as a region in a conceptual space without ever being symbolized. For artificial intelligence, the implications are even more obvious. Contemporary neural networks and language models operate by constructing embeddings, points and trajectories in high-dimensional spaces where similarity, analogy, and inference are geometric relations. When such systems "learn language," they do not manipulate symbols in the classical sense; they learn structure. Seen historically, the move toward geometric representations of communication marks a return to an older ambition in the sciences of life: to uncover order without presuming consciousness, to let form speak before meaning is named. Thom and Frege were mathematicians, they were interested in ordering language, in revealing a hidden mathematical logic operating in language. Wittgenstein famously said that anything important eludes language, this is also the point of view of philosophers of the 'emotivist school' or emotivism (with David Hume as its precursor). Here emotion precedes language and ethics. For existentialists, existence precedes essence. But all philosophical discussions presuppose language. The 'linguistic turn' in philosophy, often attached to the 1967 book by Richard Rorty *The Linguistic Turn* is now embracing the animal question and the AI questions. In the examples mentioned here, the ambition is to formalize language to pin down the elusive definition of what constitutes meaning.[xix] Whether the mathematical and geometric way of approaching the question of meaning will ultimately illuminate animal worlds or merely redraw human expectations in more abstract coordinates remains an open question.

Current Animal Translation Projects

The recent empirical work of the prolific Toshitaka Suzuki provides one of the clearest experimental demonstrations to date that meaning can be structured independently of human language and symbolic reference. Suzuki's research program, developed primarily between 2014 and 2018, focuses on the vocal communication of Japanese bird tits (*Parus minor*). In a landmark paper published in *Nature Communications* in 2016, Suzuki showed that these birds produce distinct call types that, when combined in specific sequences, elicit different behavioral responses from conspecifics. Most notably, he demonstrated that a sequence of alarm calls followed by recruitment calls prompted birds to scan for predators and then approach the source, whereas reversing the order disrupted this coordinated response. The finding was not merely that birds used different calls, but that call order mattered, and that receivers were sensitive to combinatorial structure rather than to isolated signals.

Suzuki extended these results in subsequent work, including a 2017 *Nature Ecology & Evolution* paper, by showing that these vocal combinations were learned and context-dependent rather than fixed reflexes.[xx] Playback experiments revealed that birds responded appropriately only when acoustic elements were arranged in ecologically meaningful sequences, suggesting an internal representation of structured relations between sounds, actions, and environmental conditions. Importantly, Suzuki avoided claims that these call sequences constituted "syntax" in the human linguistic sense. Instead, he emphasized that these behaviors reflected a capacity for mental imagination in birds. When he played a recording of their "snake" calls, the birds began looking toward the ground or scanning shapes resembling a snake. According to him, this suggested that the response was not a mere reflex of "instinct": the birds were mentally visualizing a snake upon hearing the sound.

The work of Toshitaka Suzuki can be read as a striking empirical counterpart to the theoretical move articulated by Peter Gärdenfors. Here concepts can exist without words, and therefore without humans. This approach aligns closely with Gärdenfors' claim that concepts need not be linguistic to be meaningful.[xxi] In Suzuki's framework, a predator type, a spatial directive, or a social coordination cue functions as a *region of behavioral relevance* rather than as a named object. The birds do not label "snake" or "hawk"; instead, they respond to structured acoustic patterns that reliably map onto different ecological situations. These patterns are neither arbitrary nor purely instinctual: they are learned, flexible, and sensitive to order. What Suzuki demonstrates, historically and experimentally, is that conceptual organization can exist without words, and that combinatorial structure can arise without human-like syntax.

It is precisely here that projects such as the *Earth Species Project* intervene. Established in 2017, the *Earth Species Project* is a nonprofit research organization based in the United States, the organization applies artificial intelligence to the study of non-human communication, with the aim of identifying structural regularities in animal signals rather than presuming linguistic meaning. Its activities are supported chiefly through philanthropic funding,

Promotional image from Earth Species Project, "Using AI to listen to all of Earth's Species." © Earth Species Project. Used for scholarly analysis.

Rather than attempting to translate animal signals into human categories, the *Earth Species Project* explicitly adopts the logic of embedding as a methodological restraint. Animal vocalizations, movements, and behaviors are not labeled with presumed meanings. They are instead recorded at scale and analyzed for recurring patterns, co-occurrences, and contextual correlations. Introduced at the International Conference on Acoustics, Speech, and Signal Processing in 2023, *BEANS* (*Benchmark of Animal Sounds*), developed by Masato Hagiwara, Benjamin Hoffman, Jen-Yu Liu, Maddie Cusimano, Felix Effenberger, and Katie Zacarian in close connection with the *Earth Species Project*, aimed at creating computer analysis shaped by low input of animal data and self-supervised deep learning while deliberately stopping short of semantic interpretation.[xxii] BEANS thus functioned as an infrastructural checkpoint. In this framework, an animal call is not treated as a word awaiting translation, but as a node within a dense relational network. Acoustic features, bodily motion, spatial position, social configuration, and environmental conditions are integrated into a shared representational space. What emerges is not a lexicon but again a sort of geometry: clusters of signals that tend to occur together, trajectories that precede or follow particular behaviors, distances that reflect functional similarity rather than human intuition. Crucially, these models do not claim to "understand" animals. They claim something more modest and more defensible: that structure can be detected without translation, and that stable relations may exist prior to interpretation.

This does not (di)solve the problem posed by Clever Hans; it reframes it. Where earlier experiments risked contamination by human cueing, embedding-based approaches seek to minimize direct human intervention altogether. The danger is no longer unconscious signaling by trainers captured by the animal, but overinterpretation by analysts.

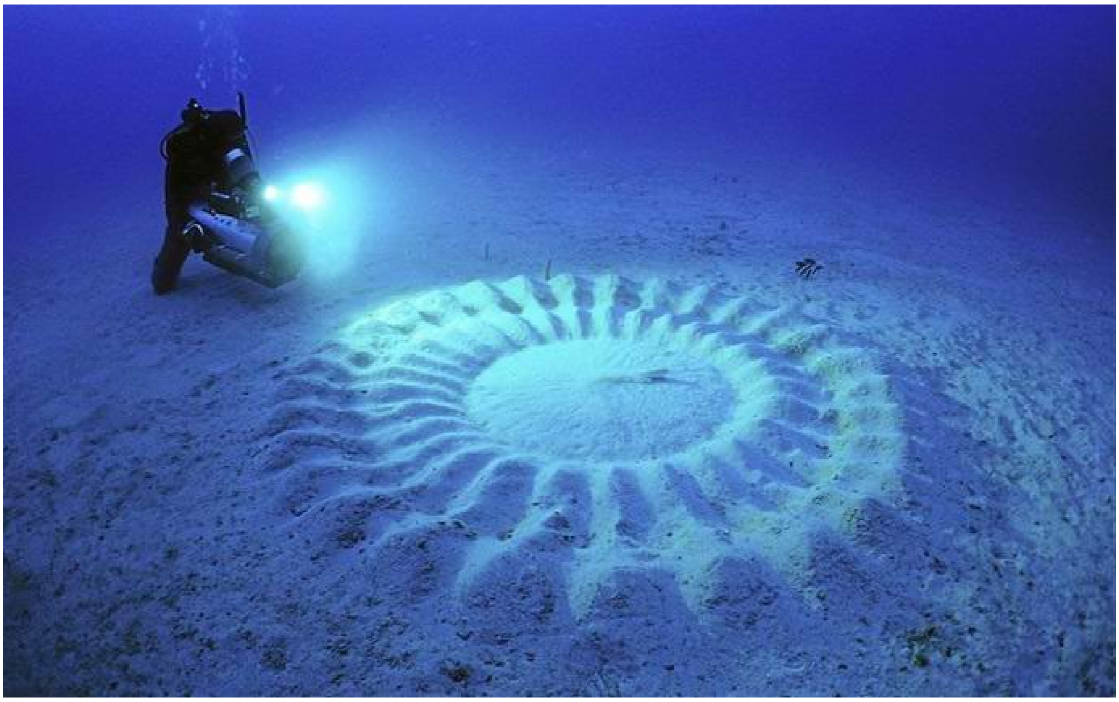
Citizen scientists recording sounds, tagging behaviors, and contributing contextual metadata became essential to training AI systems. Trained as a computer scientist and signal-processing specialist, Glotin is a professor at Université de Toulon and has played a leading role in the *CETI (Cetacean Translation Initiative)* and earlier European research programs devoted to large-scale passive acoustic monitoring of marine mammals. Hervé Glotin, pioneer of the use of AI to decode cetacean vocalisation, also filmed the cetacean to better understand their communication. In fact, meaning may change in function of the movements of the body. When we talk and make a gesture with our hands (as an Italian I can particularly relate), the meaning may change, it may signal our phrase as ironic, cynical, and by adding a gesture, it reverses the meaning of the utterance taken alone.

Contemporary projects such as *Project CETI* extend this trajectory also into ambitious terrain. *CETI* is often framed within a broader "interspecies internet" imaginary—an effort to adapt tools from natural language processing to nonhuman signals. The aspiration is not merely to detect or classify sounds, but to model structure at a level analogous to syntax, and eventually to enable predictive or interactive exchange.[xxiii] Glotin's work is distinctive not only for its early adoption of machine learning and deep neural networks for analyzing whale and dolphin vocalizations, but also for its insistence on synchronizing sound with visual and behavioral data**.** From the outset, Glotin has argued that vocalizations cannot be meaningfully analyzed in isolation. His teams have combined long-term acoustic recordings with video and movement tracking, recognizing that cetacean communication is deeply embodied. A vocal signal may change its functional significance depending on body orientation, proximity to conspecifics, group configuration, or ongoing collective movement.

Project CETI, promotional graphic

A central challenge in applying machine learning to animal communication lies in cases where data are scarce, uneven, or altogether absent, conditions that define many endangered species. Unlike human language technologies, which are trained on massive, continuously replenished corpora, animal communication research often confronts fragmentary archives: short recordings, uneven sampling across individuals, and limited contextual annotation. Species are vanishing, and the human cause mass-extinction is the fastest of all the canonical major extinctions Earth has known. We have eerie recordings of extinct species, and little data on some vanishing ones. For the majority of animals, we are far from the plethora of data gathered by citizens for the Whale FM project or "the talking dog button project."

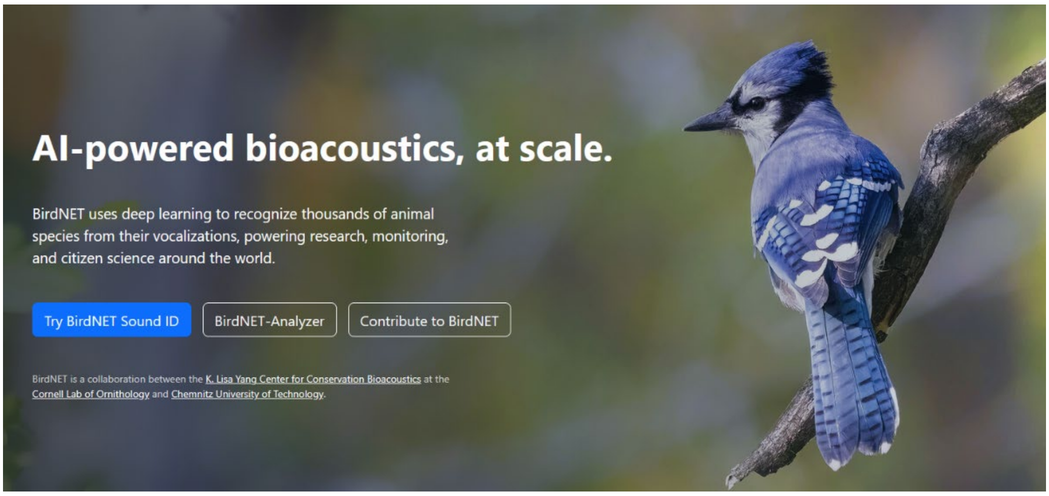
The year 2013 marked a breakthrough in "zero-resource" translation, meaning a translation by a computer in which an algorithm constructs a latent space representation of a language as multidimensional geometric "shape", to decode an unknown language without prior dictionaries or examples of translation.[xxiv] With no data on the human language needing translation, just by comparing it to other languages, the computer managed to translate it.[xxv] In the light of wild animal extinctions and biodiversity loss the "zero-resource" translation takes on an unintended ethical resonance in the context of accelerating biodiversity loss. If we continue, there will indeed be 'zero-animal language resource' left. Human languages are vanishing; animal species are vanishing; and with them disappear distinctive ways of perceiving the world. Of course, in the midst of ecosystem destruction and habitat loss for animals, new signs and ways of communicating are discovered. For the anecdote, below is the image of a pattern a tiny male puffer fish builds during a week to attract a mating partner. It was discovered for the first time by divers in 1995 but attributed to puffer fish only in 2013, unrelatedly, the year of a major AI breakthrough aforementioned. The geometry used by the puffer fish had nothing to do with the geometry used by AI to decipher language, but this animal expression is too striking not to mention here; had this species vanished before this discovery, it is a whole internal 'umwelt' of a tiny fish that will have vanished with it.



© Yoji Ookata / NHK / NaturePL

© NHK / Yoji Ookata

Historically, computational bioacoustics responded to this limitation of few data available by prioritizing detection over interpretation. Early systems asked whether a signal was present, what broad category it belonged to, who produced it, and under what environmental conditions, with a history of strong ties with the military. During the Cold War, submarine sonar and other forms of underwater acoustic research developed primarily in response to military requirements, while simultaneously creating the technical, institutional, and methodological conditions that later enabled systematic scientific study and monitoring of cetacean vocalizations. As machine learning matured, these detection tasks were scaled and standardized. Deep-learning systems trained on large reference datasets, such as *BirdNET*, made it possible to identify hundreds of bird species from audio recordings with high reliability. *BirdNET* did not "interpret" bird song in a semantic sense; rather, it provided a robust infrastructure for biodiversity monitoring, enabling ecologists to track population change, migration, and habitat degradation. *BirdNET* was first released in 2018 by a collaboration between the Cornell Lab of Ornithology and Chemnitz University of Technology. *BirdNET e*nabled the identification of hundreds of bird species from audio recordings with a level of reliability previously unattainable outside specialist settings.[xxvi] A representative case is AVES *(Animal Vocalization Encoder*) introduced in 2022 by the aforementioned *Earth Species Project*. Rather than classifying sounds into predefined categories, AVES constructs internal representations of vocal structure by exposure alone, without any attached meanings or annotations. The model learns by predicting parts of a signal from other parts, gradually building an internal representation of vocal structure without any predefined meaning.



,BirdNET homepage, https://birdnet.cornell.edu Used for scholarly analysis.

The Earth Species Project explicitly frames animal communication as a problem of unstructured data without ground truth. Benchmarks such as BEANS are designed not to test translation accuracy, but to evaluate whether models can discover stable, reproducible patterns. They are algorithmic abstractions situating formal structures that may later support interpretation. A similar logic underlies applied projects such as ***DolphinGemma***, which adapts large-language-model techniques to dolphin vocal sequences. Tokenization and sequence prediction are used not because dolphin sounds are assumed to be language, but because these tools are effective at modeling structured sequences. Meaning remains deliberately bracketed.

Promotional graphic titled "Google Launches DolphinGemma," . © Softradix Technologies Pvt. Ltd. Used for scholarly analysis.

Finally, last but not least, rodents offer also an instructive trajectory. For decades, mouse ultrasonic vocalizations were analyzed manually or with limited automation, a process that was slow, subjective, and difficult to reproduce across laboratories. The introduction of deep-learning tools such as *DeepSqueak* marked a methodological turning point (the name *DeepSqueak* bears an uncanny sound resemblance with the most prominent Chinese AI tool *DeepSeek*, a notorious rival to *ChatGPT*). Built on convolutional neural networks, *DeepSqueak* automated the detection, segmentation, and classification of ultrasonic vocalizations of rodents.[xxvii]

Conclusion

Detecting regularities in sequences does not by itself establish shared reference, intention, or social function. These technical developments cannot be separated from their ethical consequences. Historically, changes in how animals are perceived have altered their moral and political status. Birds during the French Revolution (think "la colombe de la paix"), bees in modern environmental thought, and whales in late twentieth-century conservation movements each became emblematic species through shifts in representation and attention. Today, initiatives such as the *More Than Human Life Project* seek to articulate legal and ethical frameworks for nonhuman communication technologies, drawing in part on CETI's research in Dominica. The question is not only what animals can do, but what humans owe them once they are (finally) recognised as communicative agents, or simply agents instead of objects/livestock.

Public discourse around these projects often invokes deep time. Aza Raskin, co-founder of the Earth Species Project, has emphasized that while humans have transmitted vocal culture for perhaps a few hundred thousand years, cetaceans have done so for tens of millions. The implication, that longevity correlates with accumulated wisdom, echoes earlier claims by John C. Lilly, and resonates with scientific findings showing cultural disruption in whale populations following mass mortality events. The oldest, biggest whales were indeed most sought after and hunted first, disrupting 'cultural' transmission of pod lineages.

Historians of language and cryptography have long noted that the desire to translate lost scripts is often driven less by utility than by the hope of recovering vanished knowledge, a tendency that Francesco Perono Cacciafoco has compared to a romantic, treasure-hunt mentality.[xxviii] Animal communication research risks a similar temptation if claims of "ancient wisdom" outrun empirical grounding. After all, John C. Lily toyed with the idea that dolphins could have a seat at the United Nations once we understood their ancient wisdom passed as oral history. From classical Jakob von Uexküll's work on species-specific perception to contemporary scholarship on animal time, dreams, and sensory worlds, by Vinciane Despret or Ed Yong, the emphasis has shifted toward recognizing that each species inhabits a unique experiential reality.

Recent developments in participatory science and large-scale data collection, such as dog button experiments, Whale FM, and citizen-science labeling initiatives—have expanded the empirical base of animal communication research. These projects do not establish translation, but they do make systematic comparison possible across individuals, contexts, and time. Their value lies in scale and reproducibility, not in claims of meaning.

Artificial intelligence enters this history not as a rupture, but as an extension of earlier structural approaches in linguistics, ethology, and cybernetics. Contemporary machine-learning systems do not decode messages; they embed signals into geometric spaces defined by similarity, sequence, and context. This shift from symbolic translation to geometric alignment marks a methodological constraint rather than an advance toward fluency. AI systems can identify stable relations among signals, predict transitions, and detect contextual variation, but they do not establish reference or intention.

Empirical work by researchers such as Toshitaka Suzuki shows that structured combinations of signals can exist without words, without labels, and without human-like syntax. Bird calls that depend on order and context demonstrate that conceptual organization can precede language. These findings support the view that meaning, understood minimally as functional coordination, does not require symbolic naming. The Clever Hans episode continues to define the methodological baseline of the field, and Suzuki designed experiments to avoid falling into this trap, the century-long boogeyman for ethologists. Its importance lies not in denying animal intelligence, but in revealing how easily human expectations enter experimental settings through unconscious signaling. This problem has not disappeared; it has merely shifted. In contemporary AI-based research, the primary risk is no longer trainer cueing, but interpretive overreach: inferring semantic content from statistical regularities alone .

Projects such as the Earth Species Project, BEANS, AVES, DolphinGemma, and CETI adopt this restraint explicitly. They treat animal communication as unstructured data without ground truth, prioritizing pattern discovery over interpretation. Meaning is deliberately bracketed. This does not solve the problem raised by Clever Hans; it reframes it by relocating the burden of caution from the experimental setting to the analytical stage.

Finally, the accelerating loss of species introduces an additional constraint. Many animal communication systems will never be documented at sufficient scale. In this context, zero-resource and low-data methods acquire an unintended ethical dimension. What is lost with extinction is not only biological diversity but entire perceptual and communicative systems. Machine learning does not grant access to those worlds, but it does change how carefully, how systematically, and how humbly humans can listen before they interpret. In that sense, AI neither solves nor replaces the problem of animal communication; it sharpens it and makes the ethical cost of misunderstanding harder to ignore.

*Paul Deshusses
PhD candidate in International History & Politics and Teaching Assistant at the IHEID

Bibliography

Chomsky, Noam. *Language and Mind*. New York: Harcourt, Brace & World, 1968.

Despret, Vinciane. *What Would Animals Say If We Asked the Right Questions?* Minneapolis: University of Minnesota Press, 2016.

Dirzo, Rodolfo, Hillary S. Young, Mauro Galetti, Gerardo Ceballos, Nick J. B. Isaac, and Ben Collen. "Defaunation in the Anthropocene." *Science* 345, no. 6195 (2014): 401–406.

Earth Species Project. "AVES: Animal Vocalization Encoder." Technical documentation and blog posts, 2022. https://www.earthspecies.org.

Fouts, Roger, and Stephen Tukel Mills. *Next of Kin: What Chimpanzees Have Taught Me About Who We Are*. New York: William Morrow, 1997.

Frege, Gottlob. *Begriffsschrift*. Halle: Louis Nebert, 1879.

Frege, Gottlob. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik* 100 (1892): 25–50. English translation: "On Sense and Reference."

Gärdenfors, Peter. *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT Press, 2000. Revised edition, 2014.

Glotin, Hervé, et al. "Big Data Passive Acoustic Monitoring of Marine Mammals." Conference papers and research reports, 2008–2022.

Goldberg, Yoav. *Neural Network Methods for Natural Language Processing*. San Rafael, CA: Morgan & Claypool, 2017.

Hagiwara, Masato, Benjamin Hoffman, Jen-Yu Liu, Maddie Cusimano, Felix Effenberger, and Katie Zacarian. "BEANS: The Benchmark of Animal Sounds." In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Rhodes, 2023.

Herzing, Denise L., et al. "Acoustic and Behavioral Responses of Dolphins to Synthetic Signals." *Animal Cognition* 17 (2014): 1–12.

IPBES. *Global Assessment Report on Biodiversity and Ecosystem Services*. Bonn: IPBES Secretariat, 2019.

Kahl, Stefan, et al. "BirdNET: A Deep Learning Solution for Avian Diversity Monitoring." *Ecological Informatics* 61 (2021): 101236.

Lorenz, Konrad. *Behind the Mirror: A Search for a Natural History of Human Knowledge*. Translated by Ronald Taylor. New York: Harcourt Brace Jovanovich, 1977.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient Estimation of Word Representations in Vector Space." *arXiv* (2013). https://arxiv.org/abs/1301.3781.

Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever. "Exploiting Similarities among Languages for Machine Translation." In *Advances in Neural Information Processing Systems 26*. Red Hook, NY: Curran Associates, 2013.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. “Linguistic Regularities in Continuous Space Word Representations.” In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics*. Atlanta, 2013.

Pepperberg, Irene M. “Evidence for Conceptual Quantitative Abilities in the African Grey Parrot (*Psittacus erithacus*).” *Journal of Comparative Psychology* 101, no. 2 (1987): 198–209.

Pepperberg, Irene M. *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Cambridge, MA: Harvard University Press, 1999.

Perono Cacciafoco, Francesco. Selected essays on historical linguistics and cryptography, 2015–2023.

Pfungst, Oskar. *Clever Hans (The Horse of Mr. von Osten): A Contribution to Experimental Animal and Human Psychology*. Translated by C. L. Rahn. New York: Henry Holt, 1911.

Rosenthal, Robert. “Experimenter Effects in Behavioral Research.” *Psychological Bulletin* 63, no. 6 (1966): 364–383.

Rorty, Richard, ed. *The Linguistic Turn: Essays in Philosophical Method*. Chicago: University of Chicago Press, 1967.

Savage-Rumbaugh, E. Sue, Stuart G. Shanker, and Talbot J. Taylor. *Kanzi: The Ape at the Brink of the Human Mind*. New York: John Wiley & Sons, 1994.

Sebeok, Thomas A. *The Sign and Its Masters*. Austin: University of Texas Press, 1979.

Suzuki, Toshitaka N., David R. Wheatcroft, and Michael Griesser. “Experimental Evidence for Compositional Syntax in Bird Calls.” *Nature Communications* 7 (2016): 10986. https://doi.org/10.1038/ncomms10986.

Suzuki, Toshitaka N., and Michael Griesser. “Context-Specific Calls Signal Predation Risk in a Bird.” *Nature Ecology & Evolution* 1 (2017): 132. https://doi.org/10.1038/s41559-017-0132.

Terrace, Herbert S., Laura Ann Petitto, Roger J. Sanders, and Thomas G. Bever. “Can an Ape Create a Sentence?” *Science* 206, no. 4421 (1979): 891–902.

Thom, René. *Structural Stability and Morphogenesis*. Reading, MA: Addison-Wesley, 1975. Originally published in French, 1972.

Yong, Ed. *An Immense World: How Animal Senses Reveal the Hidden Realms Around Us*. New York: Random House, 2022.

[i] Joshua Rapp Learn, "Kanzi the Bonobo, Who Learned Language and Made Stone Tools, Dies at Age 44," Scientific American, March 20, 2025, accessed January 15, 2026, https://www.scientificamerican.com/article/kanzi-the-bonobo-who-learned-language-and-made-stone-tools-dies-at-age-44/.

[ii] Savage-Rumbaugh et al., *Kanzi: The Ape at the Brink of the Human Mind*, 1994.

[iii] Ape Initiative, "Remembering Kanzi," accessed January 15, 2026, https://www.apeinitiative.org/remembering-kanzi.

[iv] Irene M. Pepperberg, "A Review of the Model/Rival (M/R) Technique for Training Grey Parrots," Animals 11, no. 10 (2021), accessed January 15, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC8469950/.

[v] Pepperberg, *The Alex Studies*, 1999.

[vi] The mirror test was an aptitude test developed in the 1970s by Gordon Gallup, a psychologist just as Irene Pepperberg, to assess the intelligence and awareness of animal; if an animal can recognize itself in a mirror it shows self-awareness and high cognitive abilities.

[vii] Oskar Pfungst, *Clever Hans (The Horse of Mr. von Osten)*, 1911.

[viii] E. Sue Savage-Rumbaugh et al., "Language Comprehension in Ape and Child," Monographs of the Society for Research in Child Development 58, nos. 3–4 (1993).

[ix] Burkhardt, R. W. (2005). *Patterns of behavior: Konrad Lorenz, Niko Tinbergen, and the founding of ethology*. University of Chicago Press.

[x] . Ana P. M. Bastos et al., "Soundboard-Trained Dogs: Investigation into Word Comprehension," *PLOS ONE* 19, no. 8 (2024): e0307189, https://doi.org/10.1371/journal.pone.0307189

[xi]Bastos, Amalia P. M., Zachary N. Houghton, Lucas Naranjo, and Federico Rossano. "Soundboard-Trained Dogs Produce Non-Accidental, Non-Random, and Non-Imitative Two-Button Combinations." *Scientific Reports* 14 (2024): 28771.

[xii] Bakker, Karen. "The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants." (2022): 1-368.

[xiii] Zooniverse Blog, "Whale FM – The Whale Song Project," November 29, 2011, accessed January 15, 2026, https://blog.zooniverse.org/2011/11/29/whale-fm-the-whale-song-project/.

[xiv] Lilly, John C. *Man and Dolphin: Adventures of a New Scientific Frontier*. 1st ed. Garden City, NY: Doubleday, 1961. Reprint, London: Gollancz, 1962. ISBN 978-0-575-01054-3. P.7

[xv] Ferdinand de Saussure, *Course in General Linguistics*, trans. Wade Baskin (New York: McGraw-Hill, 1966 [1916]).j

[xvi] Roman Jakobson, "Linguistics and Poetics," in *Style in Language*, ed. Thomas A. Sebeok (Cambridge, MA: MIT Press, 1960), 350–377.

[xvii] Christian Rutz et al., "Using Machine Learning to Decode Animal Communication," *Science* 381, no. 6654 (July 14, 2023): 152–155, https://doi.org/10.1126/science.adg7314

[xviii] Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. "Linguistic Regularities in Continuous Space Word Representations." In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751. Atlanta: Association for Computational Linguistics, 2013.

[xix] Richard Rorty, ed., *The Linguistic Turn: Essays in Philosophical Method* (Chicago: University of Chicago Press, 1967).

[xx] Toshitaka N. Suzuki, "Semantic Compositionality in Bird Calls," *Nature Ecology & Evolution* 1 (2017): 1392–1398.

[xxi] Peter Gärdenfors, *Conceptual Spaces: The Geometry of Thought* (Cambridge, MA: MIT Press, 2000).

[xxii] Masato Hagiwara et al., "BEANS: The Benchmark of Animal Sounds," in *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2023), https://doi.org/10.1109/ICASSP49357.2023.10096686

[xxiii] Pratyusha Sharma et al., “Contextual and combinatorial structure in sperm whale vocalisations,” Nature Communications 15 (2024): 3617, accessed January 15, 2026, https://www.nature.com/articles/s41467-024-47221-8.
[xxiv] Mikolov et al., *NIPS*, “Linguistic Regularities in Continuous Space”.
[xxv] . Orhan Firat, Kyunghyun Cho, and Yoshua Bengio, “Zero-Resource Translation with Multi-Lingual Neural Machine Translation,” in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Austin, TX: Association for Computational Linguistics, 2016), accessed February 11, 2026, https://aclanthology.org/D16-1026
[xxvi] Stefan Kahl et al., “BirdNET: A deep learning solution for avian diversity monitoring,” Ecological Informatics 61 (2021): 101236, accessed January 15, 2026, https://doi.org/10.1016/j.ecoinf.2021.101236.
[xxvii] K. R. Coffey et al., “DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations,” Neuropsychopharmacology 44, no. 5 (2019): 859–868, accessed January 15, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC6461910/.
[xxviii] Francesco Perono Cacciafoco, “What It Takes to Be a Glyph-Breaker: Deciphering Ancient Languages,” *Aeon* (April 7, 2025), https://aeon.co/essays/what-it-takes-to-be-a-glyph-breaker-deciphering-ancient-languages