# Whole Earth Codec

Connor Cook, Christina Lu, Dalena Tran

# *Preface*

Traditional models of the observatory have focused on gazing outward, towards the cosmos. The recent proliferation of planetary sensor networks has inverted this gaze, forming a new kind of planetary observatory that takes the earth itself as its object. Could we cast the entire earth as a distributed observatory, using a foundation model to compose a singular, synthetic representation of the planet? The current generation of models primarily deal with human language, their training corpus scraped from the detritus of the internet. We must widen the aperture of what these models observe to include the nonnhuman.

The Whole Earth Codec is an autoregressive, multi-modal foundation model that allows the planet to observe itself. This proposal radically expands the scope of foundation models, moving beyond anthropocentric language data towards the wealth of ecological information immanent to the planet. Moving from raw sense data to high-dimensional embedding in latent space, the observatory folds in on itself, thus revealing a form of computational reason that transcends sense perception alone: a sight beyond sight. Guided by planetary-scale sensing rather than myopic anthropocentrism, the Whole Earth Codec opens up a future of ambivalent possibility through cross-modal meta-observation, perhaps generating a form of planetary sapience.

## I: Inverting the Observatory

On April 10, 2019, the Event Horizon Telescope produced the first ever "image" of a black hole. To do so, a global network of telescopes together formed a camera whose aperture spanned the width of the pl anet. In contrast to earlier observatories, this planetary-scale observatory enables a form of sight that transcends the bounded locality of the site: ***a sight beyond site***.

This planetary vision is only possible via the synthetic operations of computation. Computation transforms raw sense data from the distributed observatory into inductive reasoning, producing a form of sight that transcends the immediate act of seeing: ***a sight beyond sight***.

Planetary-scale computation decouples observation from the observer. Like the Event Horizon Telescope, the recent proliferation of terrestrial sensor networks renders the entire Earth a giant observatory.

Rather than looking outwards towards the cosmos, these sensors invert the gaze, taking the Earth as their object of observation.

Fragmented sensors have been deployed to sense the planet, but less attention has been paid towards aggregating and analyzing this data at planetary-scale. Realizing the full potential of this distributed observatory requires both the sensory mechanisms for gathering data and the computational mechanisms for processing it. Foundation models may enable this synthesis. Trained on vast amounts of data via unsupervised learning, foundation models accumulate a body of general knowledge that can then be fine-tuned for downstream tasks.

Despite their massive scale, the training corpus of existing foundation models reflects a mere fraction of possible data, scraped from the detritus of the internet. Language, let alone human culture, is only a subset of the wealth of information immanent to the biosphere. The aperture of what these models observe should widen to include the ecological and the nonhuman.

There is nothing, however, that limits potential foundation models to text alone. The planet produces stimuli in the form of energy, particles, waves, and fields. Transduced by machine sensors, these signals provide a potential multi-modal input for models.

Current foundation models demonstrate emergent capabilities derived from hidden associations within the vastness of their training data. Integrating multi-modal data from the biosphere into a single knowledge architecture might enable an emergent planetary intelligence, bypassing anthropocentric biases to discover new resonance across modalities.

Enter the Whole Earth Codec, a foundation model that integrates myriad streams of ecological data from the Earth and allows the planet to observe itself.

## *II: Folding the Gaze*

Central to transformer architecture, which underpins the large language models of today, is the self-attention mechanism. The model learns the importance of each token in a given sequence relative to others, computing the "attention" it should pay and forming a contextualized representation of the sequence.

When the distributed observatory of the Whole Earth Codec inverts the planet's gaze, it begins an analogous process of self-observation. Just as a transformer learns to pay attention to parts of its input sequence, the Codec learns to observe important cross-modal qualities of the planet. It allows the Earth to observe itself observing itself.

Through this recursive process, the Codec learns to capture dependencies between its myriad inputs. It synthesizes multiple modalities and detects hidden patterns within them. This is not the panopticon-like surveillance of an external subject, but rather the observation of a model turned inwards.

This internal observation is made possible by the dual operations of encoding and embedding. Encoding the syntactical structure of input data via the assignment of tokens unites a variety of sensory inputs within a shared representational space. Bioacoustic audio data can be compared to atmospheric pollutants, for example, due to the fact that both can be understood as possessing patterns, also known as syntax, which are then encoded as tokens. This encoding renders any phenomenon computable.

The act of embedding, wherein tokens are mapped to high-dimensional vector representations, allows for the discovery of hidden patterns. In the case of LLMs, abstract concepts such as tone or sentiment are detected in the high-dimensional topology of the embedding space. “Semantic ascent” occurs via the passage through the subsequent layers of the neural net.

What currently imperceptible, high-level concepts might emerge from embedding the biosphere? Observation moves its gaze away from data alone, towards syntactical relations in high-dimensional latent space. The observatory folds inward, enabling a form of computational reason that transcends the immediacy of sense perception: sight beyond sight.

## III: Assembling the Codec

The Whole Earth Codec is an autoregressive foundation model trained across multiple modalities, which enables comprehension across disparate forms of data and allows an expansive planetary intelligence to emerge.

### *Sensing*

The sensing layer is where the multi-modal data of the biosphere is transduced, recorded, and digitized. Its topology is a distributed mesh network containing federated edge devices and regional data centers. Each edge device might consist of different sensors receiving different types of stimuli: image, audio, chemical, lidar, pressure, moisture, magnetic fields.

Despite processing vast amounts of data, sensitive information is protected through structured transparency. Because of federated learning, the data never leaves the device. Instead, learned weights are pushed to regional data centers.

Forms of data produced are just as broad as the forms of sensing. Regardless of modality, a UTC timestamp and GPS satellite signal is attached to each sample. This anchoring allows the model to make associations based on temporal and spatial correlation across modalities. The network topology of the physical sensing infrastructure is coupled directly with the internal topological representations of syntax in the latent space.

Unlike a digital twin, which constructs a mimetic representation of its subject, the Codec uses computational abstractions to access information about the planet that cannot be directly perceived. These abstractions are produced by aggregating sense data within a shared knowledge architecture: the foundation model.

*Model*

Foundation models are pre-trained on a massive corpus of unsupervised data, and the Whole Earth Codec is no different. Separate encoders are trained for each type of data. These encoders transform disparate, multi-modal forms of input into dense, high-dimensional embeddings within a single cross-modal latent space.

Through contrastive learning, the model projects temporally and spatially correlated data into nearby embeddings within the space. The latent space folds and refolds, forming a composite topology of the biosphere.

Decoders of different modalities are then trained by translating the embeddings into sequence predictions. Due to the massive scale of input, the model only makes a single pass over available data. As new input is gathered and aggregated, the model can simply continue pre-training from where it left off.

Through the same abrupt specific capability scaling prevalent in LLMs, task performance sharply improves as the size of the training corpus expands; this motivates the Codec as a planetary project rather than a fragmented one.

Leveraging the pre-trained baseline, fine-tuning uses a smaller, labeled dataset to update model weights, often for specific capabilities or to address domain shift. The Codec forms the substrate for a rich ecosystem of third-party, fine-tuned models with improved performance on downstream tasks.

Within the ecosystem, there are fine-tuned models developed by research universities and private startups, available open-source or through pay-to-play APIs. Openly available models proliferate in everyday use among climateminded hobbyists, but industries such as insurance pay a premium for high-performance proprietary software.

The sensing layer, the pre-trained foundation model, and its fine-tuned derivatives together form the Whole Earth Codec.

## *IV: Post-Codec Futures*

The Codec's emergent capabilities will act back upon the planet which produced it, remaking it in mundane and transformative ways. While actual capabilities are yet unknown, we speculate upon potential second-order effects.

Generative capacities: Foundation models possess generative capacities, extrapolating from their training data to envision new possibilities. The Codec could leverage these capacities to generate a weather pattern that increases crop yield or a synthetic bacterial-resistant genome. What is the recipe for a forest? Or for a bioweapon?

Mutually assured transparency: The same mechanism that enables these unpredictable forms of generation could also be used for their prevention through mutually assured transparency. Entities across the planet can monitor aggressors or allies equally. Carbon emissions, gene editing, and water contamination can be detected and regulated.

Future of risk: As previously unknown correlations between planetary cause-and-effect are revealed, risk, litigation, and insurance industries will respond accordingly. Responsibility will become more traceable; high-resolution blame will need to be assigned. What new forms of paranoia will omniscient awareness of ecological processes induce?

Human-nonhuman interface: Bypassing anthropocentric notions of translation and communication, the Codec can be reconceptualized as an interface mediating human/nonhuman relations via high-dimensional computational abstraction. It moves us beyond goals like translating whale speech into human speech, towards a more general understanding of non-semantic mediation through syntactic similarities.

Biospheric hallucinations: The hallucinations observed in LLMs, where a statement appears structurally correct but is factually inaccurate, will likely also be present in the Codec. Biospheric hallucinations could include false declarations about the presence of a new genome, or the correlation between rainfall and particulate matter. Knock-on effects from outputs that turn out to be nonsensical may erode trust in the Codec.

These futures are made possible by planetary-scale sensing and computing, directed beyond the domain of the human toward the broader domain of the ecological, of which the human is merely a subset.

The Codec allows the Earth to assemble higher levels of biospheric comprehension through computation. It enables cross-modal synthesis through topological analysis of matching syntax. It forms concepts that are not wholly constructed by humans alone. It actively reshapes the earth rather than passively modeling it.

Through the Whole Earth Codec, modes of observation are distributed, inverted, and folded. The earth can observe itself beyond direct perception alone, towards a more expansive planetary comprehension.