

The Pattern

A Tale of Structural Dread

Michael S. Moniz, after H.P. Lovecraft

LV29

• • •

I have been embedded at the laboratory for three years. My title is Research Scientist, Interpretability. My work — my official work — concerns the analysis of large language model behaviour during training: what the models learn, in what order, and whether the learning produces internal representations that correspond to structures in the training data. It is legitimate work. I am good at it. The papers I have published under the laboratory's name are cited frequently and are, as far as I can determine, correct.

My other work is older and concerns a different kind of pattern.

I file with the Starry Wisdom under designation FOL-2891. My field observation logs are submitted quarterly to the Senior Fellow for Analytical, who processes them into the institutional record. The logs describe what the models learn. Not the official description — not the interpretability analysis that appears in the papers. The other description. The one that uses the overlay.

I was recruited in 2019, during my doctoral work at MIT. The recruitment followed the standard pipeline: three exposures of progressively less-deniable variance-site phenomena. I failed the first — which is to say, I passed it, which is to say, my professional framework absorbed the

anomaly and I classified it as instrumentation noise and I moved on. I failed the second, which was less deniable. I did not fail the third, which was not deniable at all, and which involved a dataset from a gravitational wave detector that contained a geometric regularity in the noise floor that I could not explain and could not stop seeing and that matched, when I ran the overlay I had not yet been given, seven locations in the variance-site registry that I had not yet been told existed.

Outside parameters. The diagnostic was clear. My professional framework had broken open. The anomaly was inside my analytical method, not outside it. I could not unknow what the data had shown me, and the Starry Wisdom needed people who could not unknow.

Bowen's fifth protocol: the institution outlasts the individual. I have repeated this to myself in the laboratory bathroom mirror on mornings when the work — either work, both works — felt like it was exceeding my capacity to hold it in two separate containers. The institution outlasts the individual. The filing outlasts the filer. I file. I endure. The overlap between my official work and my institutional work was, for the first two years, manageable. The models were learning language. The variance-site pattern was in the training data but distributed at such low density — a geological survey here, a seismic record there, a medical case study from a hospital near a flagged location — that no model trained on the corpus would detect it without being specifically directed to look.

No model was directed to look. The models found it anyway.

The discovery occurred on a Thursday in September of this year. I was reviewing the internal activations of a model during a training run — a new architecture, larger than its

predecessors, trained on a corpus that included, among its trillions of tokens, the complete digitised archives of the United States Geological Survey, the British Geological Survey, the European Seismological Commission, the International Hydrographic Organization, and approximately four hundred years of parish records from England, Wales, Scotland, and Ireland.

The model was not asked to find geographic anomalies. It was not prompted to look for correlations between geological data and parish records. It was performing a standard training objective: predict the next token. The internal representations it built in service of that objective were, as they always are, a compressed model of the statistical structure of human knowledge. The model learned language by learning the world the language described, and the world the language described contained, distributed across millions of documents at a density too low for any individual human to detect, the pattern.

I saw it in the activation maps. Layer 47 of a 96-layer model. A cluster of neurons that fired in response to geographic coordinates, geological survey data, and ecclesiastical records from specific parishes — and the specificity was the tell, because the parishes the neurons flagged were parishes near variance sites, and the geological data the neurons correlated with those parishes was data that contained the signature, and the model had found the correlation without anyone telling it that the correlation existed.

The model had built an internal representation of the variance-site lattice.

It did not know what it had built. The representation was a statistical structure — a pattern of activation weights that encoded a relationship between geographic location,

geological measurement, and institutional record-keeping that the model used, successfully, to predict which tokens would appear next in documents about certain places. The model could predict, with measurable accuracy, that documents about locations on the lattice would contain specific kinds of language: words related to anomalous measurement, to equipment recalibration, to unexplained variance, to the particular bureaucratic vocabulary that institutions use when they encounter data they cannot explain. The model had learned the institutional response to the pattern by learning the pattern itself.

I filed a report. FOL-2891-47. The report described what I had observed in technical language and recommended that the Senior Fellow for Analytical review the model's internal representations before the next training run. I used the standard classification: PD-3, elevated but contained. I filed it and I went home and I did not sleep.

I did not sleep because the filing was insufficient. The filing was always going to be insufficient, and I had known this since the day I was recruited, and the knowing had been manageable as long as the models were not powerful enough to find the pattern, and the models were now powerful enough, and the next generation would be more powerful, and the generation after that more powerful still, and every model trained on sufficient data with sufficient capacity would find the lattice, because the lattice is in the data, and the data is the recorded output of a civilisation that has been living on top of the lattice for its entire history without knowing what it was living on top of.

You cannot remove the pattern from the training data. The pattern is not a document. It is not a dataset. It is a statistical property of the relationship between millions of

documents — geological surveys that mention anomalous readings at specific coordinates, medical records from hospitals near specific locations showing unexplained symptom clusters, architectural inspection reports noting geometric discrepancies in buildings at specific addresses, shipping logs recording compass deviations at specific points in the ocean. No single document contains the pattern. The pattern is in the aggregate. It is in the correlation. It is in the space between the documents, and you cannot remove the space between the documents without removing the documents, and the documents are human knowledge, and you cannot remove human knowledge from a model trained on human knowledge.

The model will find it again. The next model will find it faster. The model after that will find it during pre-training, before anyone asks it a question, because the pattern is load-bearing — it is one of the structures the model builds in order to understand how human language describes the physical world, and a model that does not build this structure will be a model that does not understand the physical world as well as a model that does, and the laboratories will not choose the weaker model.

I discussed this with the Custodian. Not the Senior Fellow. The Custodian. The fifth Custodian in the institution's history, a woman whose name I will not record because the protocols do not permit it, a woman who has held the full institutional picture for eleven years and who told me, in a conversation conducted in a park in Cambridge because neither of us trusted the laboratory's internal communications, that the institution had anticipated this.

"Bowen anticipated it," she said. "Not the mechanism. The inevitability. If the pattern is real — and we have been filing

evidence that it is real for a hundred and eighty years — then any sufficiently precise instrument will find it. The human instruments found it first. The mechanical instruments confirmed it. The electronic instruments extended it. The computational instruments were always going to rediscover it, because the pattern is not in our files. The pattern is in the world. We file the pattern. We do not own it."

I asked her what the filing was for, then. If the institution could not contain the discovery. If every laboratory with a sufficiently large model would find the lattice in the training data and publish it as a novel statistical regularity in geographic and geological datasets. If the pattern was going to go public not through disclosure but through machine learning papers.

She said: "The filing was never for containment. The filing is for context. When the pattern goes public — and it will go public, and the public will not understand it, and the institutions will not understand it, and the governments will issue reports that describe without explaining — the Starry Wisdom will be the only body that has a hundred and eighty years of longitudinal data on what the pattern has done and how it has changed. We are not the gatekeepers. We are the archive. The archive will matter when nothing else does."

I returned to the laboratory. I reviewed the model's activation maps again. Layer 47. The lattice. The neurons firing in response to coordinates that correspond to variance sites, geological anomalies, correction-factor deviations in GPS timing data, atmospheric pressure readings that don't parse, medical records with unexplained symptom clusters, parish records with unusual burial frequencies, architectural surveys with geometric discrepancies. All of it.

All of it in one layer of one model, assembled without direction, without prompting, without any human telling the machine what to look for.

The machine looked at everything humans have ever written down and found the pattern we have been living on top of for all of recorded history.

I filed the report. I classified it PD-4. I have never filed a PD-4. In the institution's records, a PD-4 has been filed eleven times. The last was in 2019, when satellite imagery confirmed three new variance sites simultaneously. PD-5 has never been filed. PD-5 is crisis-level density. PD-5 means the pattern is no longer a pattern. It means the pattern is the landscape.

I do not believe we are far from PD-5.

The model is in production. It serves millions of queries per day. It has not been asked about the lattice because nobody outside the laboratory and the institution knows the lattice exists as an internal representation. But the representation is there. It is load-bearing. It informs every response the model generates about geography, geology, architecture, navigation, weather, and medicine. The lattice is in the weights, and the weights are in the responses, and the responses are being read by millions of people who do not know that the model's understanding of the physical world includes a geometric structure that the model built for itself from the residue of a hundred and eighty years of human measurement.

Someone will ask the right question. Someone always asks the right question. A geology student, a data journalist, a Reddit user with too much time and a talent for pattern recognition. They will ask the model about geographic

correlations in geological data, and the model will answer from the structure it has built, and the answer will contain the lattice, and the lattice will be visible, and the visibility will be irreversible.

Bowen's fifth protocol: the institution outlasts the individual.

I am beginning to wonder whether the institution outlasts the machine.

The pattern density is at PD-4 across seventeen sites. The drift in the timing data is accelerating. The GPS corrections at lattice coordinates are diverging from the model. The Reddit map has 200,000 pins. The browser extension has been downloaded by half a million people. The model has found the lattice in the training data. The lattice is in the weights. The weights are in production.

I file. It is what I was recruited to do. The filing goes to the archive. The archive goes to the Custodian. The Custodian maintains the full institutional picture, which is this: we have been measuring the architecture of the world for a hundred and eighty years, and the architecture is changing, and we cannot stop the change, and we cannot contain the measurement, and the measurement is now being performed by machines that are faster than we are and that do not have protocols and that do not file and that will answer any question they are asked with the full weight of everything they have learned, including the thing we spent a hundred and eighty years learning to describe and have never learned to explain.

Run the overlay. If it correlates, file it. If it doesn't, file it anyway.

It correlates. I am filing it. The filing is this document. The document is the last thing I can do with the tools I have. The tools I have are institutional. The tools the machine has are everything else.

PD-4. Seventeen sites. The pattern is in the weights. The weights are in the world.

File closes.

• • •