The EOA Program, Part 0: The Engineering of Alphabets as an Open Frontier at the Intersection of Symbolic Dynamics, Representation Theory, and Deterministic Machine Learning

Bahaa Budargham

Independent Researcher, Montreal, Canada
Founder, Engineering of Alphabets (EOA) Research Initiative
ORCID: https://orcid.org/0009-0005-2955-3703
Email: bdarghamneurolabs@gmail.com

Abstract

We introduce the Engineering of Alphabets (EOA) a research program that is based on a single observation. This observation, if confirmed, should not exist according to current theories of representation. At the center of this program is a recursive operator. This operator is called beta. It maps the spoken names of alphabet letters into numerical sequences. Early observations suggest that with a specific bijective phonetic mapping for the 26 letters of the English alphabet beta creates sequences that converge to a limiting constant ratio (LCR). This ratio is approximately 3.037. With a different mapping the ratio changes to 2.689. With another mapping the ratio is 3.355. The sequences seem to collapse into a 16-element basis. Word vectors made from this basis appear to exist on a very low-dimensional subspace. These vectors suggest a recursive, tree-like generative structure that is behind this collapse. These phenomena if confirmed are reproducible because they depend on the encoding. They are mathematically rigid. None of them is currently understood.

This paper is Part 0 of the Engineering of Alphabets Program. It contains no proofs. It has no conclusive experiments. Its purpose is to ask the questions that the next five papers will explore. It also places those questions in the scientific world. This world includes symbolic dynamics [4], phonetic encoding theory, low-rank representation learning and deterministic machine learning. These are fields that are currently overlapping on this problem. We argue that Engineering of Alphabets identifies a genuine gap. It is a class of deterministic string-to-sequence operators whose behavior is seen in experiments but not explained by theory. The questions we ask can be answered, but they need tools that no single field has on its own.

1. What Is the Engineering of Alphabets?

The Engineering of Alphabets (EOA) is not a theory. It is a research program built on preliminary observations that, if confirmed, should not exist according to current frameworks. The underlying observations were first placed on the public record in 2010 [3]; this paper and the parts that follow report on their continued, independent investigation.

The observation is this: if you take the spoken names of the 26 letters of the English alphabet, feed them into a fixed deterministic recursive operator beta, and iterate that operator 43 times, the resulting 26 numerical sequences appear to exhibit the following properties:

1. Every sequence converges to the same limiting ratio, but the value of that ratio depends on the phonetic encoding. Under one bijective mapping, lambda is approximately 3.037. Under another, it is 2.689. Under a third, it is 3.355.
2. The 26 sequences are not independent; they appear to collapse to a 16-element basis.
3. Any word vector constructed from that basis appears to have very low effective dimensionality.
4. When a single letter's encoding is altered without altering the rest, the global ratio appears to break.

These are preliminary observations, not established facts. They need independent verification and theoretical grounding. Why does the operator converge? Why is the ratio shared across all letters under a consistent encoding? What is the 16-element basis — a phonetic universal, a graph-theoretic constraint, or a numerical coincidence? Is the dimensionality collapse a limitation to overcome or a design principle to exploit? And, perhaps most importantly, does any of this matter for natural language processing, information theory, cryptography, or the design of future symbolic systems? We do not yet know.

EOA is the program that investigates these questions. It spans mathematics (symbolic dynamics [4], linear algebra, combinatorics), linguistics (phonetics, encoding theory, cross-linguistic comparison), and machine learning (representation theory, compression, deterministic alternatives to learned embeddings). No single one of these fields currently claims jurisdiction over the phenomena we observe. That is why the program is necessary.

1.1 Why Part 0?

Scientific papers usually present answers. This one presents questions. The reason is strategic: the EOA Program spans multiple disciplines, and each discipline will approach the problem with different tools. By stating the questions before the answers, we allow researchers to engage with the problem on their own terms, without being constrained by the vocabulary or assumptions of our forthcoming empirical papers.

Part I [1] will present the sequences, the basis, and the dimensionality collapse. Part II [2] will present the encoding-dependence of the limiting ratio. Parts III, IV, and V will address transients, downstream behavior, and the operator itself. But none of those papers will answer the deepest question: what kind of system is this?

That question belongs to the community. This paper is the invitation.

1.2 The Structure of This Paper

The questions below are organized by the domain of knowledge they challenge. Each section poses problems that are, to the best of our knowledge, unanswered by any existing literature. Some may have short answers. Others may require new mathematical frameworks. A few may turn out to be ill-posed, revealing that our underlying assumptions are wrong. All of them are genuine.

The sections are:

- Section 2: The operator itself (mathematics, symbolic dynamics).
- Section 3: The limiting constant ratio (encoding theory, phonetics, inverse design).
- Section 4: The basis structure (combinatorics, group theory).
- Section 5: Dimensionality collapse (representation theory, linear algebra).
- Section 6: The transient regime (information theory).
- Section 7: Cross-linguistic universals (linguistics, phonology).
- Section 8: The application landscape (computer science, cryptography, commerce, AI trends).
- Section 9: Reverse engineering and target design (applied mathematics).
- Section 10: The epistemology of staged disclosure (philosophy of science).
- Section 11: Structural hints and unseen dynamics (fractal geometry, formal language theory)
- Section 12: Call for collaboration.
- Sections 13 and 14 address the program roadmap and concluding remarks.

1.3 Summary Table of Open Questions

To help you find the questions in your own field, the table below summarizes all 43 questions in one line each.

Q#	Domain	What it asks
2.1	Operator	What operator class produces this convergence: substitution system, linear recurrence, contraction mapping, or none of these?
2.2	Operator	Why does a least-squares linear fit of the recurrence matrix fail? Where does the nonlinearity actually live?
2.3	Operator	Would random symbols or number words still converge to a shared ratio, or is alphabet-specific encoding required?
2.4	Operator	Is the transient regime monotonic, oscillatory, or chaotic, and does it survive at infinite precision?
3.1	Limiting ratio	Is λ a computable function of the encoding, predictable without running the operator?
3.2	Limiting ratio	Which encoding property actually determines λ: length, frequency, phonetic features, or coupling structure?
3.3	Limiting ratio	Is λ a global invariant of the alphabet, or does it depend on presentation order?
3.4	Limiting ratio	Is there a continuous mapping from phonetic space to λ-space, allowing λ to be targeted by design?
3.5	Limiting ratio	Is there a universal ceiling on λ, or can an encoding push it arbitrarily high?
4.1	Basis	Why does the 26-letter alphabet collapse to exactly 16 basis sequences?
4.2	Basis	What property of the letter-name strings, not the phonemes, determines group membership?
4.3	Basis	Does the basis restructure under an alternate encoding, or stay fixed at 16 with different groupings?
4.4	Basis	Is the basis a unique fingerprint of the encoding, or can two different encodings share one?
5.1	Dim. collapse	What is the precise, metric-backed definition of the apparent rank-1 collapse?
5.2	Dim. collapse	Is the collapse a limitation to overcome, or a design feature of the representation?
5.3	Dim. collapse	Does the 43-term vector carry any information beyond a single scalar?
5.4	Dim. collapse	Can a guaranteed-collapsed representation be exploited for adversarial robustness?
6.1	Transient	How much letter-specific information survives in the detrended residual?
6.2	Transient	Are the transients structured or noisy, and do they converge to a secondary pattern?
6.3	Transient	Can the transient structure be recovered from the 16-element basis alone?
6.4	Transient	Would a truncated, non-collapsed vector (terms 1–20) be more useful than the full 43-term one?
7.1	Cross-linguistic	Do non-alphabetic scripts (syllabaries, abugidas, logographies) show analogous convergence?
7.2	Cross-linguistic	Is the tight Semitic cluster a structural property of abjads, or just a small-sample artifact?
7.3	Cross-linguistic	Would a synthetic alphabet with no phonetic history still produce a basis and an LCR?
7.4	Cross-linguistic	Is there a relationship between alphabet size and λ, and how would we test it properly?
8.1	Applications	Could a fully characterized operator serve as a useful cryptographic primitive?
8.2	Applications	Is there any downstream task where a low-rank, semantically empty representation actually helps?
8.3	Applications	Does determinism matter more than dimensionality in any real operational scenario?
8.4	Applications	Could truncated EOA vectors outperform the full vector on spelling, anagram, or OCR tasks?
8.5	Applications	Who would pay for a representation that is semantically empty but algebraically rigid?
8.6	Applications	Can EOA-43 act as a zero-training structural prior or fixed initialization for learned embeddings?
8.7	Applications	Can the dimensionality collapse be harnessed as a principled model-compression scheme?
9.1	Reverse design	Can we search for encodings whose λ approximates π, e, or the golden ratio?
9.2	Reverse design	What is the reachable set of λ values, and does it intersect mathematically significant numbers?
9.3	Reverse design	Can an encoding be designed so its sequence terms encode primes at specific positions?
9.4	Reverse design	What constraints govern inverse design? Must the encoding stay phonetic and bijective?
9.5	Reverse design	What would a targetable λ actually enable: periodicity, self-similarity, cryptographic use?
10.1	Disclosure	Can claims about a partially characterized operator be independently verified pre-disclosure?
10.2	Disclosure	Would full disclosure resolve these questions quickly, or are they hard regardless?
10.3	Disclosure	What precedent exists for staged disclosure in mathematics (RSA-129, AlphaFold)?

Q#	Domain	What it asks
10.4	Disclosure	How should priority be adjudicated when the phenomenon is public but the mechanism is not?
11.1	Structural	Does the generation process exhibit a self-similar branching structure, and can its geometry be formally characterized?
11.2	Structural	Can the tree-like propagation be characterized by a finite grammar (e.g., L-system), and would such a grammar explain the basis and the collapse?

2. The Operator: What Generates Convergence?

The operator beta accepts a string (a letter name) and produces a sequence of real numbers. After 38 terms, the ratio between consecutive terms appears to stabilize. The choice of 43 total iterations was arbitrary but sufficient to observe stabilization; the theoretical minimum required terms remains an open question. This is not gradient descent. There is no loss function, no training data, and no stochasticity. The convergence appears to be built into the recurrence itself.

Question 2.1. What class of deterministic string-to-sequence operators produces geometric convergence from character-level inputs? Is beta a substitution system [4, 5], a coupled linear recurrence, a contraction mapping, or something that falls outside all three categories?

Question 2.2. If beta is a linear recurrence of the form $v_{n+1} = M.v_n$, why does a least-squares fit of M from the sequences yield inconclusive results [2]? Does the nonlinearity lie in the encoding step, the recurrence step, or the coupling between letters?

Question 2.3. If the operator were applied to random symbols or number words, would it still converge to a shared ratio? If yes, the phenomenon may be purely formal rather than phonetically grounded. If no, the alphabet-specific encoding is structurally necessary.

Question 2.4. Is the convergence monotonic, oscillatory, or chaotic in the transient regime? If we had infinite precision, would the transient structure disappear, or does it encode genuine letter-specific information that survives even in the limit?

---

3. The Limiting Constant Ratio: Why 3.037?

For one English bijective phonetic mapping, the LCR is approximately 3.037. For another mapping, it is 2.689. For a third, it is 3.355. These values are not universal constants. They appear to be functions of the encoding.

Question 3.1. Is lambda a computable function of the encoding? That is, given a phonetic mapping (a set of letter-name strings), can one predict lambda without running beta?

Question 3.2. If the answer to 3.1 is yes, what properties of the encoding determine lambda? Is it the mean letter-name length, the character-frequency distribution, the phonetic-feature profile, the graph-theoretic structure of cross-letter dependencies, or a joint interaction of all of these?

Question 3.3. If the answer to 3.1 is no, does lambda depend on the order in which the encoding is presented to the operator? In other words, is lambda a global invariant of the entire alphabet, or does it emerge from a specific sequential coupling?

Question 3.4. The English alternative encoding shifts lambda from 3.037 to 3.355, a substantial change. Arabic and Hebrew cluster near 2.71, while Greek yields 4.128. Is there a continuous mapping from phonetic-space to lambda-space? Can we design an encoding whose LCR approximates a target value to arbitrary precision?

Question 3.5. If lambda is encoding-dependent, is there a universal bound? Can an encoding produce an arbitrarily large lambda, or does the operator impose a finite ceiling?

4. The Basis: Why 16?

Preliminary observations suggest that the 26 letters of the English alphabet collapse to 16 unique sequences. Three groups appear to share sequences: (B, D, P, T, V, Z), (F, L, M, N, S), and (I, R). The remaining 13 letters appear to be singletons.

Question 4.1. Why 16? Is this number a function of the English alphabet size (26), or would a different alphabet size yield a different basis cardinality? If we added a 27th letter to English, would the basis grow to 17, or would the new letter slot into an existing group?

Question 4.2. The groupings do not correspond to any standard phonetic classification. B, D, P, T, V, Z are not all stops, not all fricatives, and not all voiced. F, L, M, N, S share no obvious articulatory feature. I and R are acoustically distant. What property of the letter-name strings (not the phonemes) determines membership in a group?

Question 4.3. Is the apparent 16-element basis a property of the English primary encoding only? If we apply the English alternative encoding, does the basis restructure into a different set of groupings, or does it remain 16 with the same letters grouped?

Question 4.4. If the basis is a fingerprint of the encoding, can two different encodings produce the same basis? If so, the basis is not a unique identifier. If not, the basis is a unique signature of the encoding, and we should be able to invert it.

---

5. Dimensionality Collapse: A Bug or a Design?

Because all basis sequences appear to converge to the same geometric rate, word vectors built from them seem to collapse to a very low-dimensional subspace, possibly effective rank 1.

Question 5.1. Preliminary observations suggest the vectors may collapse to a very low-dimensional subspace, but what is the precise definition of this collapse? Is it exact rank 1, numerical rank 1 under a specified tolerance, or an artifact of the construction formula? What metric (singular value ratio, cumulative variance, Frobenius approximation error) should be used to define “effective rank” in this context?

Question 5.2. If dimensionality collapse is inevitable, is it a failure of representation or a feature of compression? In other words, is nature telling us that letter-name strings contain only 1 bit of independent geometric information, or are we simply using the wrong decoder?

Question 5.3. Because EOA-43 vectors collapse to a single direction, they are essentially just a weighted sum of letter positions and alphabetical indices. If that is all they do, why does the operator generate 43 terms instead of just one? Is there any information in those 43 numbers that a single number cannot capture?

Question 5.4. Can dimensionality collapse be exploited? If a representation is guaranteed to live on a line, can it be used to construct adversarially robust classifiers (where perturbations orthogonal to the line are harmless), or does near-chance accuracy on semantic tasks prove that the line points in a meaningless direction?

---

6. The Transient: Where Is the Real Information?

Terms 1 through 38 appear to be letter-specific. They deviate from the geometric trend. They may be the only place where the 16 basis vectors differ.

Question 6.1. If we detrend the sequences by dividing each term by lambda^n, the residual carries the transient structure. What is the entropy of this residual? Does it contain enough information to distinguish all 26 letters, or only the 16 basis groups?

Question 6.2. Are the transients structured or noisy? Do they themselves converge to a secondary pattern, or do they remain chaotic across all 43 terms? If they converge to a secondary ratio, is that ratio also encoding-dependent?

Question 6.3. The operator's value, if any, may lie in the transients. Can the transient structure be recovered from the 16 basis vectors alone, or does it require access to the full recurrence? In other words, is the basis a lossy compression of the transient, or a completely different object?

Question 6.4. If the transients encode letter identity, can they be used to build a non-collapsed representation? What would happen if we used only terms 1 through 20 to build word vectors, discarding the converged tail entirely?

---

7. Cross-Linguistic Universals: Beyond English

Preliminary tests across five encodings and four languages suggest the LCR varies. The basis structure may vary too. But we do not know the scope.

Question 7.1. Do non-alphabetic scripts (syllabaries, abugidas, logographic systems) exhibit analogous behavior? Can beta be applied to the names of hiragana, katakana, or Devanagari characters? If so, do they converge to a shared ratio, or does the operator fail outside the alphabetic paradigm?

Question 7.2. The Semitic cluster (Arabic 2.751, Hebrew 2.710) is tighter than the Indo-European spread (English 3.037, Greek 4.128). Is this because Semitic abjads share a structural property (consonantal roots, vowel omission), or because the sample size is too small to mean anything?

Question 7.3. If we constructed a synthetic alphabet with no historical phonetic basis, purely to test the operator, would it still produce a basis and an LCR? If yes, the phenomenon is purely formal. If no, the phenomenon is phonetically grounded.

Question 7.4. Is there an inverse relationship between alphabet size and lambda? English has 26 letters and lambda = 3.037. Hebrew has 22 and lambda = 2.710. Greek has 24 and lambda = 4.128. The correlation is not obvious. With only five data points, we cannot even pose the regression confidently. What is the minimum experimental design needed to test this?

---

8. The Application Landscape: Cipher, Compression, and AI

EOA-43 is not a word embedding in any conventional sense. But its deterministic, algebraic rigidity may have value in domains where semantics is not the goal. Does it carry any semantic content? How does it perform on standard benchmarks? These are open questions that Parts IV and V will address.

Question 8.1. If we can fully understand how the operator works will its ability to map strings to sequences be useful for keeping information secure?. Is the fact that it works in a straightforward linear way a major limitation for cryptography? The person who figures out how the operator works first will control its properties. Is it worth trying to be the first to do this?

Question 8.2. Is there any real use for a representation that carries no meaning but is algebraically rigid? Or is the whole program just a mathematical curiosity with no practical value?

Question 8.3. EOA-43 uses 2.69 KB, the same as a random projection. Does its determinism ever matter more than its size? For example, in reproducible research, or in deployment on any device without retraining?

Question 8.4. If the transient terms (Section 6) contain recoverable letter-specific information, could a truncated EOA representation (terms 1 through 20 only) outperform the full 43-term vector on character-level tasks such as spelling correction, anagram detection, or OCR post-processing?

Question 8.5. From a commercial perspective: who would pay for a representation that is semantically empty but algebraically rigid? Is the value proposition in the representation itself, or in the operator that generates it?

Question 8.6. Current AI trends emphasize zero compression, on-device inference, and deterministic reproducibility. EOA-43 is already a 2.69 KB basis that generates vectors without training. Can it be integrated into zero-compression pipelines as a structural prior? Can it serve as a fixed initialization for learned embeddings, reducing the number of trainable parameters in downstream models?

Question 8.7. Can the apparent dimensionality-collapse property be harnessed for model compression? If a 768-dimensional BERT embedding can be projected onto the EOA line with minimal loss on a specific task, does that constitute a principled compression scheme, or merely a degenerate case?

---

9. Reverse Engineering and Target Design

The most powerful question EOA raises is not analytical but synthetic. If lambda is encoding-dependent, can we invert the relationship?

Question 9.1. Given that the set of possible encodings for a fixed finite alphabet is at most countable (finite strings over a finite character set), the LCR space is at most countable. Can we practically search for encodings whose LCR approximates pi, *e*, or the golden ratio to arbitrary precision?

Question 9.2. Can we characterize the countable set of LCR values that is actually reachable? What is the image of the encoding-to-LCR mapping, and does it intersect any mathematically significant numbers?

Question 9.3. Can we design an encoding whose sequence terms themselves encode prime numbers at specific positions? This would require understanding whether the operator's recurrence can be tuned to generate number-theoretic sequences.

Question 9.4. If inverse design is possible, what are the constraints? Must the encoding be a bijective phonetic mapping, or can any consistent string assignment work? Can we automate the search with gradient-free optimization, or does the discreteness of string space make this a combinatorial search problem?

Question 9.5. What applications would a targetable LCR enable? A representation whose convergence rate is tuned to pi might possess circular or periodic properties when embedded in geometric space. A representation tuned to the golden ratio might exhibit self-similarity. A representation tuned to approximate a prime might be useful in cryptographic key generation or error-correcting codes.

---

10. The Epistemology of Staged Disclosure

The operator beta is currently under investigation. Its full recurrence has not been published. This raises meta-questions about how scientific knowledge should proceed when the generative mechanism is not yet public.

Question 10.1. Can a scientific claim about a partially characterized operator be independently verified if the full mechanism is not yet published? If the sequences and their properties are documented, is that sufficient for mathematical verification, or does proof require access to the recurrence?

Question 10.2. If the operator were fully characterized today, would the questions in this paper be answered within a year, or would they remain open because the mathematics is genuinely hard? In other words, is the current state of knowledge a consequence of incomplete disclosure, or of incomplete understanding?

Question 10.3. Have others done this before? RSA kept RSA-129 secret for decades. DeepMind released AlphaFold structures before the code. Does that justify our approach, or is EOA a different kind of case?

Question 10.4. If a researcher independently discovers the recurrence relation without access to the operator, do they own the discovery? How should priority be adjudicated when the empirical phenomenon is public but the generative mechanism is still being investigated?

---

11. Structural Hints and Unseen Dynamics

Beyond the numerical results in earlier sections, a closer look at the generation process points to something deeper. Starting from a single symbol, the expansion branches out along multiple paths. It settles into a finite, closed set — yet from that set, infinite descendants keep emerging. This suggests a recursive, tree-like structure behind the basis collapse we observed.

Question 11.1 Does the generation process exhibit a recursive, self-similar branching structure? Does this indicate a formal fractal geometry, an L-system grammar, or a graph-theoretic attractor? If so, what is its branching ratio, and how does it relate to the limiting constant ratio λ?

Question 11.2 If the branching turns out to follow an L-system, can we write down the exact rules that generate it? And would those rules explain why the basis has exactly 16 elements and why the dimensionality collapses — not just confirm that it happens, but show the mechanism behind it?

---

12. Call for Collaboration

The questions posed in this paper are not solvable by a single researcher or a single discipline. They require expertise that spans pure mathematics, computational linguistics, representation learning, cryptography, and phonetic theory. We therefore issue an open call for collaboration.

We are prepared to offer the following to qualified collaborators:

- Access to preliminary sequence data under standard academic agreements.
- Access to computational tools via a secure notebook environment that allows testing of arbitrary encodings without exposing the underlying recurrence.
- Co-authorship on papers that produce correct, reproducible theoretical explanations of the phenomena described here.
- Independent publication rights with acknowledgement for researchers who prefer to work alone.

Priority will be adjudicated by timestamped submission. We are particularly interested in the following bounded projects:

1. A classification of the operator’s recurrence class (Section 2).
2. A predictive model or inverse-design framework for the encoding-to-LCR mapping (Section 9).
3. An empirical evaluation of EOA-43 on zero-compression or on-device NLP pipelines (Section 8).
4. A cryptographic or information-theoretic analysis of the basis structure (Sections 4 and 8).

For collaboration proposals, contact: bdarghamneurolabs@gmail.com

---

13. Program Roadmap: Our Research Journey

The EOA Program documents our research journey into these questions. Parts I through V report what we have found so far. We do not claim to have all answers, and we invite parallel independent work.

Part	Title	Status	Category
0	The Engineering of Alphabets as an Open Frontier	This paper	Programmatic overview and question statement
I	A Rank-Collapsed Deterministic Symbolic Representation with a Convergent Recursive Basis	In preparation [1]	Empirical and descriptive
II	Empirical Cross-Encoding Variation of the Limiting Ratio	In preparation [2]	Empirical and descriptive
III	Transient Structure and Detrended Residuals	Future	Analytical
IV	Boundary Mapping on Downstream Tasks	Future	Applied
V	The Operator beta and Its Recurrence Class	Future	Theoretical

Parts I and II are empirical and descriptive. Parts III and IV are analytical and applied. Part V is theoretical. But none of them, individually or collectively, will answer every question posed in this paper. That is by design. The EOA Program is not a closed system; it is an open frontier.

---

14. Conclusion: The Value of a Well-Posed Problem

We have posed 43 questions across 10 domains. Some may have short answers. Others will likely need entirely new mathematical frameworks. A few could turn out to be ill-posed and the assumptions behind them may simply be wrong. All of them are genuine.

The EOA Program does not claim to have answers. It claims the questions are worth asking. There is a deterministic operator. It turns letter names into convergent sequences. It groups letters into an unexplained basis. It collapses word vectors to a line. None of this should exist under current theories of representation. Its existence is the puzzle.

We invite mathematicians to classify the operator. We invite phoneticians to explain the encoding dependence. We invite representation theorists to interpret the dimensionality collapse. We invite cryptographers to test the mapping properties. We invite computer scientists to find the application. And we invite the broader scientific community to tell us whether the entire structure is a deep truth or a deep coincidence.

Beyond the convergence and basis collapse we have reported, we have seen hints of a deeper structure. It appears in the unfolding of the recursion. It suggests nested, tree-like propagation. This may settle into a finite closed set. At the same time, it may continue to generate infinite complexity. We do not know whether this points to formal fractal geometry, to an L-system grammar, or to a graph-theoretic constraint. It remains an open question. We have not formalized this observation. But we raise it here as a structural frontier. It may connect the numerical phenomena to a generative grammar.

The questions are open.

Acknowledgments

No external funding was received for this work. The author acknowledges the use of open-source software (Python, NumPy, SciPy).

References

[1] B. Budargham. The EOA Program, Part I: A Rank-Collapsed Deterministic Symbolic Representation with a Convergent Recursive Basis. In preparation, 2026.

[2] B. Budargham. The EOA Program, Part II: Empirical Cross-Encoding Variation of the Limiting Ratio. In preparation, 2026.

[3] B. Budargham. Engineering of Alphabets (EOA): Mathematical properties of written alphabetic scripts. Internet Archive, June 29, 2010.

[4] D. Lind and B. Marcus. Symbolic Dynamics and Coding. Cambridge University Press, 1995.

[5] N. P. Fogg. Substitutions in Dynamics, Arithmetics and Combinatorics. Springer, 2002.