

Before Information: Languages, Dictionaries, Alphabets, and Messages

Alexey A. Nekludoff

AstraVerge Research

E-mail: an@astraverge.org
ORCID: 0009-0002-7724-5762

June 20, 2026

Abstract

Information theory and modern cryptography traditionally begin with symbols, messages, and communication channels. In this paper we argue that these notions presuppose a deeper structure. A message requires an alphabet, and an alphabet requires a language capable of stabilizing distinctions. We show that the existence of observable symbols is not a primitive fact but a consequence of language-dependent alphabet formation. This observation reveals an implicit assumption shared by information theory and cryptography: the existence of an observable alphabet. We discuss the consequences of this assumption and outline its relevance for future research on alphabet destabilization and post-alphabetic communication systems.

Keywords: information theory; philosophy of information; language; dictionary; alphabet; message; symbol; symbolic stabilization; alphabet formation; message formation; epistemic foundations of communication; linguistic foundations of cryptography; cryptographic analysis; observable alphabets; post-alphabetic cryptography.

Contents

1	Introduction	2
2	The Hidden Assumption of Information Theory	2
3	Languages and Distinctions	3
4	Dictionaries as Stabilized Distinctions	4
5	Alphabets as Symbolic Projections of Dictionaries	4
6	Messages as Sequences of Alphabetic Projections	6
7	Information as a Consequence of Message Formation	7
8	Cryptographic Consequences	8
9	Discussion	9
10	Conclusion	10

1 Introduction

Modern information theory is commonly understood as a theory of messages, symbols, encodings, and communication channels. Since the publication of Shannon’s seminal work [1], the standard model has treated communication as the transmission of messages composed of symbols drawn from an alphabet.

This framework has proven extraordinarily successful. However, it contains an implicit assumption that is rarely examined: the existence of an observable alphabet. The source is assumed to generate symbols, the encoder is assumed to transform symbols, the channel is assumed to transmit encoded symbols, and the receiver is assumed to reconstruct symbols. Yet the conditions under which something becomes a symbol are not themselves part of the model.

In this paper we argue that the alphabet is not a primitive object. It is a derived structure that depends on a deeper hierarchy:

Language \rightarrow Dictionary \rightarrow Alphabet \rightarrow Message \rightarrow Information.

According to this hierarchy, languages establish systems of meaningful distinctions; dictionaries stabilize those distinctions as repeatable units; alphabets provide symbolic representations of dictionary entries; messages become sequences of alphabetic symbols; and information emerges through transformations of messages.

This reverses a common implicit intuition. Information does not precede language. Rather, information becomes possible only after language-dependent distinctions have been stabilized through dictionaries and represented through alphabets.

The central thesis of this paper may be summarized as follows:

Information theory begins with messages.

Messages presuppose alphabets.

Alphabets presuppose dictionaries.

Dictionaries presuppose languages.

Therefore, language precedes information.

This thesis does not reject Shannon’s mathematical framework. Instead, it identifies the epistemic and linguistic preconditions that Shannon’s framework presupposes. The theory of information begins after symbols are already available; the present work examines what must be in place before symbols can be treated as informational objects.

This shift also has consequences for cryptography. Cryptographic analysis operates on messages represented through alphabets. If the observable alphabet fails to emerge for an external observer, then the ordinary objects of cryptographic analysis—symbols, frequencies, distributions, and patterns—fail to stabilize as well.

The purpose of the present paper is therefore to examine the hierarchy that precedes information and to clarify why languages, dictionaries, and alphabets must be treated as foundational conditions for messages, information, and cryptographic interpretation.

2 The Hidden Assumption of Information Theory

Information theory is commonly presented as a theory of communication. A source generates messages, an encoder transforms them into signals, a channel transmits those signals, and a receiver reconstructs the original message.

In Shannon’s formalism, however, the source does not generate arbitrary observations. It generates symbols drawn from a predefined alphabet [1]. The existence of the alphabet is therefore assumed prior to any communication process.

This assumption is so deeply embedded in modern information theory that it is rarely discussed explicitly. Symbols are treated as primitive objects. The source emits symbols, probabilities are assigned to symbols, entropy is calculated over symbols, and coding schemes operate on symbols.

Consequently, the following sequence is usually taken for granted:

Alphabet \rightarrow Message \rightarrow Information.

The alphabet appears as a starting point rather than an object requiring explanation.

From an engineering perspective, this simplification is entirely reasonable. Communication systems are designed to transmit already-existing messages. The origin of the symbolic units composing those messages lies outside the scope of the theory.

However, from a linguistic and epistemic perspective, the situation is less straightforward. Alphabets do not exist independently of systems of distinctions. A symbol can function as a symbol only if observers repeatedly identify certain observations as equivalent instances of the same unit.

The existence of an alphabet therefore presupposes a prior process of stabilization.

This observation raises a fundamental question:

What must exist before an alphabet can exist?

The standard information-theoretic framework does not address this question. It begins only after symbolic units have already been established.

The present work argues that this omission is not merely a methodological simplification but a hidden assumption shared by information theory and, subsequently, by modern cryptography.

If alphabets are not primitive objects but derived structures, then information theory begins in the middle of a deeper hierarchy rather than at its foundation.

3 Languages and Distinctions

The existence of an alphabet presupposes the existence of distinctions. A symbol can function as a symbol only if it can be distinguished from other symbols and repeatedly recognized as the same entity across multiple observations.

However, distinctions do not emerge from alphabets. Rather, alphabets emerge as representations of distinctions that have already become meaningful within a language.

In this paper, a language is understood in a broad sense as a system capable of establishing, preserving, and communicating distinctions. Under this interpretation, natural languages represent only one particular instance of a more general phenomenon. Formal languages, symbolic systems, signaling protocols, biological communication systems, and machine representations may all be regarded as languages insofar as they define admissible distinctions and relations between them.

The crucial observation is that distinctions are language-dependent.

Consider two observers possessing different linguistic frameworks. The same physical observation may correspond to different distinctions, different categories, or no distinction at all. Consequently, the symbolic structures derived from those distinctions may also differ.

The existence of a distinction therefore cannot be reduced to the physical properties of an observation alone. Distinctions arise through the interaction between observations and a language capable of interpreting them.

Languages thus provide the first level of stabilization. They determine which differences are meaningful, which similarities are relevant, and which observations may be grouped together as instances of the same category.

Only after such stabilization becomes possible can more structured forms of representation emerge.

4 Dictionaries as Stabilized Distinctions

If languages establish distinctions, a further mechanism is required to preserve those distinctions in a stable and reproducible form.

We refer to such a mechanism as a dictionary.

In the present work, the term dictionary is used in a broad sense. A dictionary is not merely a collection of words. Rather, it is a structured repository of distinctions recognized by a language.

A dictionary determines which distinctions are considered stable, reproducible, and communicable within a given linguistic framework.

Consequently, dictionaries perform a function fundamentally different from that of alphabets.

Alphabets provide representations.

Dictionaries provide distinctions.

An alphabet may specify how a symbol is written, transmitted, or encoded. A dictionary specifies what distinctions are available for representation in the first place.

For this reason, alphabets cannot be regarded as primary structures. Before a symbol can be represented, the distinction represented by that symbol must already exist within a dictionary.

This relationship may be expressed as

$$\text{Language} \rightarrow \text{Dictionary} \rightarrow \text{Alphabet.}$$

The dependence is asymmetric.

Languages may exist without formal alphabets. Human behavior, animal communication, signaling systems, and pre-literate societies may possess stable distinctions without possessing explicit symbolic representations.

Likewise, dictionaries may exist without alphabets. A distinction can be recognized, remembered, and reproduced before a dedicated symbolic notation is introduced.

The reverse is not true.

An alphabet without a dictionary is merely a collection of marks. In the absence of stabilized distinctions, symbols possess no identifiable referential function and therefore cannot participate in message formation [2].

The dictionary thus represents the first level at which distinctions become stable objects of communication.

Only after distinctions have been stabilized by a dictionary can they be projected into symbolic form through an alphabet.

This observation has important consequences for information theory. If messages are sequences of symbols and symbols belong to alphabets, then the existence of messages presupposes not only alphabets but also the dictionaries that make those alphabets meaningful.

The hierarchy therefore extends beyond the conventional information-theoretic framework:

$$\text{Language} \rightarrow \text{Dictionary} \rightarrow \text{Alphabet} \rightarrow \text{Message} \rightarrow \text{Information.}$$

From this perspective, information appears not as a primitive entity but as the final result of a sequence of stabilization processes that begins at the level of language.

5 A Minimal Formal Model of Stabilization

Let O denote a set of observations available to an observer.

A language L induces a relation

$$\sim_L \subseteq O \times O$$

such that

$$x \sim_L y$$

means that observations x and y are treated as instances of the same stabilized distinction within the language L .

When \sim_L is reflexive, symmetric, and transitive over the relevant domain of observations, it induces a quotient structure

$$D_L = O/\sim_L.$$

We call D_L the dictionary induced by the language L .

Elements of D_L are not physical observations themselves but equivalence classes of observations stabilized as repeatable distinctions.

An alphabet A_L may then be understood as a system of observable representatives or marks associated with elements of D_L . Formally, an alphabetic projection may be represented as a mapping

$$\pi_L : D_L \rightarrow A_L.$$

A message is therefore not a primitive sequence of physical observations. It is a sequence of alphabetic projections of stabilized distinctions:

$$M_A = (\pi_L([x_1]), \pi_L([x_2]), \dots, \pi_L([x_n])).$$

This minimal model captures the central dependency developed in the preceding sections:

$$L \rightarrow \sim_L \rightarrow D_L \rightarrow A_L \rightarrow M_A.$$

Alphabet formation fails when no stable relation \sim_L can be established by an observer over the observed stream.

In that case, the quotient structure $D_L = O/\sim_L$ does not arise for that observer, and no alphabetic projection can be constructed.

Consequently, message formation fails at the observational level.

6 Alphabets as Symbolic Projections of Dictionaries

The previous section established that dictionaries stabilize distinctions recognized within a language. We now consider the role of alphabets in this hierarchy.

Traditionally, alphabets are often treated as fundamental symbolic systems from which messages are constructed. However, such a view obscures the dependence of alphabets upon dictionaries.

An alphabet does not create distinctions.

Rather, it provides a symbolic representation of distinctions that have already been stabilized by a dictionary.

In this sense, alphabets function as projection mechanisms.

Let

$$D = \{d_1, d_2, \dots, d_n\}$$

be a dictionary consisting of stabilized distinctions.

An alphabet

$$A = \{a_1, a_2, \dots, a_m\}$$

provides symbolic forms through which elements of the dictionary may be represented and communicated.

The relationship is therefore not

$$A \rightarrow D,$$

but

$$D \rightarrow A.$$

The alphabet depends upon the dictionary, while the dictionary does not depend upon the alphabet.

This asymmetry may be observed in numerous communication systems.

Human languages existed long before the invention of writing systems.

Sign languages operate without reference to alphabetic writing.

Biological signaling systems employ stable distinctions despite lacking explicit symbolic notation.

Similarly, a newly invented alphabet does not automatically create new distinctions. It merely introduces alternative symbolic forms through which existing distinctions may be expressed.

The same dictionary may therefore be projected through multiple alphabets.

For example, a mathematical distinction may be represented through:

- natural language descriptions;
- algebraic notation;
- graphical representations;
- programming languages;
- formal logical systems.

The distinction remains the same while the alphabet changes.

This observation demonstrates that alphabets are not intrinsic properties of messages. They are representational layers imposed upon underlying dictionaries.

Consequently, two observers may possess access to the same alphabet while operating with different dictionaries.

In such cases identical symbol sequences may generate different interpretations.

Conversely, two observers may share a dictionary while employing entirely different alphabets.

In this case communication remains possible despite the absence of a common symbolic representation.

The communicative role of the alphabet is therefore secondary.

Its function is not to establish distinctions but to provide observable tokens through which distinctions may be transmitted.

Messages emerge only after such projections have been formed.

Accordingly, the hierarchy developed in this paper may be refined as

$$\text{Language} \rightarrow \text{Dictionary} \rightarrow \text{Alphabet} \rightarrow \text{Message}.$$

The alphabet occupies an intermediate position between distinctions and messages. It transforms stabilized distinctions into observable symbolic units, thereby making message formation possible.

7 Messages as Sequences of Alphabetic Projections

Having established that alphabets are symbolic projections of dictionaries, we may now reconsider the nature of messages.

In conventional information theory, a message is typically defined as a sequence of symbols drawn from an alphabet. While operationally useful, such a definition conceals the dependency structure developed in the preceding sections.

If alphabets are projections of dictionaries, then messages are necessarily constructed from projected distinctions rather than from primitive symbols.

A message therefore does not originate at the alphabetic level.

Its origin lies in a sequence of distinctions selected from a dictionary and subsequently represented through an alphabet.

Let

$$D = \{d_1, d_2, \dots, d_n\}$$

denote a dictionary of stabilized distinctions.

A message may then be represented as an ordered sequence

$$M_D = (d_{i_1}, d_{i_2}, \dots, d_{i_k}).$$

This sequence exists at the dictionary level.

Communication requires the projection of these distinctions into an alphabet

$$\Pi : D \rightarrow A.$$

The observable message becomes

$$M_A = (\Pi(d_{i_1}), \Pi(d_{i_2}), \dots, \Pi(d_{i_k})).$$

The message observed by a receiver is therefore not the original sequence of distinctions but its alphabetic projection.

This distinction is important.

The same dictionary-level message may be represented through multiple alphabets without changing its underlying structure.

Conversely, identical alphabetic sequences may correspond to different dictionary-level messages when interpreted through different dictionaries.

The existence of a message is therefore not an intrinsic property of a symbol sequence.

Rather, it depends upon the successful reconstruction of the dictionary-level distinctions represented by that sequence.

This observation leads to an observer-dependent notion of messages.

Consider an observer who receives a sequence of alphabetic symbols but lacks access to the dictionary from which those symbols were generated.

Such an observer may identify recurring tokens, frequencies, and structural regularities, yet remain unable to reconstruct the distinctions encoded by the message.

At the extreme, the observer may fail to recognize the sequence as a message at all.

In this situation, the observable symbol stream remains a physical object but ceases to function as a communicative object.

The existence of a message is therefore conditional upon the availability of a suitable interpretive framework.

Messages are not merely transmitted.

Messages are reconstructed.

The process of reconstruction requires access to the dictionary that stabilizes the distinctions represented by the alphabet.

Consequently, message formation may fail even when symbol transmission succeeds.

A receiver may correctly observe every symbol while failing to reconstruct the distinctions from which those symbols originated.

This possibility reveals another hidden assumption of conventional information theory.

The standard communication model assumes not only a shared alphabet but also a shared dictionary.

Without such a dictionary, symbol sequences remain observable, but messages fail to emerge as meaningful entities.

The hierarchy developed throughout this paper may therefore be extended as follows:

Language \rightarrow Dictionary \rightarrow Alphabet \rightarrow Message \rightarrow Information.

Information appears only after message formation has been achieved, and message formation itself depends upon successful dictionary reconstruction.

The existence of observable symbols is therefore insufficient to guarantee the existence of observable messages.

8 Information as a Consequence of Message Formation

The preceding sections established a dependency hierarchy extending from languages to information:

Language \rightarrow Dictionary \rightarrow Alphabet \rightarrow Message \rightarrow Information.

This hierarchy allows us to reconsider the position occupied by information within communication systems.

In conventional information theory, information is commonly treated as a fundamental object associated with messages generated by a source. Entropy, coding efficiency, redundancy, and channel capacity are all defined with respect to symbol sequences and probability distributions over those sequences [1].

From the perspective developed in this paper, however, such quantities become meaningful only after message formation has already occurred.

Information theory begins with messages.

Messages presuppose alphabets.

Alphabets presuppose dictionaries.

Dictionaries presuppose languages.

Consequently, information is not a primitive object but the final element of a sequence of prior stabilizations.

This observation does not invalidate the mathematical framework of information theory. Rather, it clarifies its domain of applicability.

Information theory begins at the point where symbolic units have already been established and where messages can already be recognized as messages.

Questions concerning the origin of those symbolic units lie outside the scope of the theory.

In this sense, information theory does not explain how distinctions arise. It assumes their existence.

Similarly, it does not explain how symbols become identifiable. It assumes that observers can already recognize and reproduce symbolic units.

The existence of a message therefore represents a precondition of information-theoretic analysis rather than a result of that analysis.

This distinction becomes particularly important when communication occurs between observers possessing different languages, different dictionaries, or different symbolic systems.

In such situations, successful transmission of symbols does not necessarily imply successful transmission of messages.

Likewise, successful transmission of messages does not necessarily imply successful reconstruction of the distinctions from which those messages originated.

The conventional communication model therefore conceals several layers of stabilization beneath the notion of information itself.

Information appears only after:

1. distinctions have been established by a language;
2. distinctions have been stabilized by a dictionary;
3. distinctions have been represented through an alphabet;

4. messages have been constructed from alphabetic symbols.

Only after these conditions have been satisfied can information be defined, measured, encoded, transmitted, or analyzed.

The principal conclusion may therefore be stated as follows:

Information is not the foundation of language.

Language is the foundation of information.

Or, equivalently,

Information does not precede language.

Information emerges from language through dictionaries, alphabets, and messages.

From this perspective, information theory begins not at the foundation of communication but at a particular stage within a deeper hierarchy of symbolic stabilization.

9 Cryptographic Consequences

The hierarchy developed in this paper has implications extending beyond information theory.

Modern cryptography operates on messages represented through alphabets. Encryption transforms messages into ciphertexts, while cryptanalysis attempts to recover information about the original message from observable properties of the ciphertext.

Despite the diversity of cryptographic constructions, a common assumption remains implicit.

Cryptographic analysis presupposes the existence of observable symbolic units.

Whether the analysis relies on symbol frequencies, statistical distributions, structural regularities, machine learning techniques, or algebraic representations, it requires the ability to identify recurring entities within the observed data.

This requirement may be expressed through the hierarchy established in the preceding sections.

If information presupposes messages, messages presuppose alphabets, alphabets presuppose dictionaries, and dictionaries presuppose languages, then cryptographic analysis necessarily presupposes the existence of an observable alphabet.

Consequently, the following dependency chain emerges:

Language \rightarrow Dictionary \rightarrow Alphabet \rightarrow Message \rightarrow Cryptanalysis.

The existence of an observable ciphertext alphabet therefore becomes a hidden prerequisite of cryptographic interpretation.

This observation applies not only to classical frequency analysis but to a much broader class of analytical methods.

A frequency distribution requires identifiable symbols.

An n -gram model requires identifiable symbols.

A probabilistic language model requires identifiable symbols.

Machine learning methods require stable features derived from identifiable symbols.

Even when the underlying mathematics differs substantially, the existence of repeatable symbolic units remains a common prerequisite.

Cryptographic analysis therefore begins only after alphabet formation has occurred.

The analyst must first establish that multiple observations correspond to instances of the same symbolic entity.

Only after such stabilization becomes possible can higher-level structures be constructed.

This observation suggests a distinction between two fundamentally different approaches to cryptographic security.

The conventional approach assumes the existence of an observable alphabet and seeks to protect messages represented within that alphabet.

This point may be illustrated by a simple boundary case. Suppose that repeated source-level distinctions are projected into ciphertext observations in such a way that no two observable ciphertext units can be reliably classified as instances of the same symbolic class by an external observer. The legitimate receiver may still possess the internal rule required to reconstruct the dictionary-level distinctions. For the external observer, however, the quotient structure required for alphabet formation does not emerge. The observed stream may remain physically measurable, but it does not stabilize into a ciphertext alphabet suitable for cryptographic analysis.

An alternative approach would seek to prevent the formation of an observable alphabet itself.

In the first case, symbols exist but their meaning is protected.

In the second case, the symbolic units required for analysis fail to stabilize.

The latter possibility lies outside the scope of traditional cryptographic theory and motivates further investigation into communication systems in which observable alphabet formation becomes impossible or unstable.

Such systems suggest the existence of a broader class of security mechanisms based not on the concealment of symbols but on the prevention of symbolic stabilization.

The exploration of these mechanisms remains a subject for future work.

10 Discussion

The argument developed in this paper does not challenge the mathematical validity of information theory. Shannon's framework remains fully applicable within its intended domain.

The present work addresses a different question.

Rather than investigating the transmission of information, it investigates the conditions under which information becomes possible.

From this perspective, information theory begins after several prior stabilization processes have already occurred. Languages establish distinctions, dictionaries stabilize those distinctions, alphabets provide symbolic representations, and messages organize those representations into communicable structures.

The existence of information therefore presupposes the existence of a symbolic framework capable of supporting message formation.

This observation also clarifies the relationship between information and interpretation.

Information theory intentionally abstracts away from semantics. Within Shannon's framework, the meaning of a message is irrelevant to the mathematical treatment of communication.

The hierarchy proposed in this paper does not contradict this principle. Instead, it identifies a level of analysis that precedes the distinction between semantics and transmission.

Before meaning can be ignored, symbolic units must first exist.

Before symbols can exist, distinctions must be stabilized.

Before distinctions can be stabilized, a language capable of supporting them must already be present.

The proposed hierarchy therefore concerns the preconditions of symbolic communication rather than the transmission process itself.

A second implication concerns cryptography.

Conventional cryptographic analysis assumes that observable symbolic units exist within the communication process. The present work suggests that alphabet formation itself may constitute an independent object of study.

If cryptographic interpretation requires observable alphabets, then the conditions under which alphabets emerge become relevant to cryptographic security.

The investigation of communication systems in which observable alphabet formation fails, remains unstable, or becomes observer-dependent represents a possible direction for future research.

Finally, the hierarchy proposed here should not be interpreted as a historical sequence. Languages, dictionaries, alphabets, messages, and information frequently coexist within real communication systems.

The claim of this paper is not chronological but structural.

The hierarchy describes relations of dependence rather than stages of temporal development.

Information depends upon messages.

Messages depend upon alphabets.

Alphabets depend upon dictionaries.

Dictionaries depend upon languages.

In this sense, language remains the foundational condition of information.

The stronger consequence is that information theory and cryptography do not operate at the foundation of symbolic communication. They operate after symbolic stabilization has already occurred.

This does not diminish their practical value. It clarifies their boundary.

Where alphabets are stable, information-theoretic and cryptographic methods apply.

Where alphabets fail to stabilize, the object of analysis changes.

In such regimes, the central problem is no longer the protection of messages within an alphabet but the formation or non-formation of the alphabet itself.

11 Conclusion

Modern information theory begins with messages represented through alphabets. The existence of those alphabets is generally treated as a primitive assumption.

This paper has argued that such an assumption conceals a deeper dependency structure.

Languages establish distinctions.

Dictionaries stabilize distinctions.

Alphabets provide symbolic representations of stabilized distinctions.

Messages organize those representations into communicable forms.

Information emerges through transformations of messages.

The resulting hierarchy may be summarized as

Language \rightarrow Dictionary \rightarrow Alphabet \rightarrow Message \rightarrow Information.

From this perspective, information is not a primitive object but the final outcome of a sequence of prior stabilizations.

The central conclusion of this work is therefore straightforward:

Information theory begins with messages.

Messages presuppose alphabets.

Alphabets presuppose dictionaries.

Dictionaries presuppose languages.

Therefore, language precedes information.

This conclusion does not replace information theory. Rather, it identifies the linguistic and epistemic foundations upon which information-theoretic analysis implicitly depends.

Recognizing these foundations opens new directions for research in communication theory, symbolic systems, and cryptography, particularly in domains where alphabet formation itself becomes a central object of investigation.

References

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x
- [2] A. A. Nekludoff, “Alphabet formation and the epistemic limits of cryptographic analysis,” *Zenodo*, 2026. DOI: 10.5281/zenodo.18126229
- [3] L. Wittgenstein, *Philosophical Investigations*. Oxford: Blackwell, 1953.
- [4] F. de Saussure, *Course in General Linguistics*. New York: McGraw-Hill, 1966, Original work published 1916.
- [5] B. L. Whorf, *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, J. B. Carroll, Ed. Cambridge, MA: MIT Press, 1956.
- [6] U. Eco, *A Theory of Semiotics*. Bloomington: Indiana University Press, 1976.
- [7] N. Goodman, *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Hackett Publishing, 1976.
- [8] G. Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press, 1987.
- [9] A. A. Nekludoff, “Regulatory model of language origin: Imitation, behavioral forms, and semantization,” *Zenodo*, 2025. DOI: 10.5281/zenodo.18063394