

# Novelty Assessment Report

**Paper:** The Counting Power of Transformers

**PDF URL:** <https://openreview.net/pdf?id=IAFwK6NyrP>

**Venue:** ICLR 2026 Conference Submission

**Year:** 2026

**Report Generated:** 2026-01-05

## Abstract

Counting properties (e.g. determining whether certain tokens occur more than other tokens in a given input text) have played a significant role in the study of expressiveness of transformers. In this paper, we provide a formal framework for investigating the counting power of transformers. We argue that all existing results demonstrate transformers' expressivity only for (semi-)linear counting properties, i.e., which are expressible as a boolean combination of linear inequalities. Our main result is that transformers can express counting properties that are highly nonlinear. More precisely, we prove that transformers can capture all semialgebraic counting properties, i.e., expressible as a boolean combination of arbitrary multivariate polynomials (of any degree). Among others, these generalize the counting properties that can be captured by support vector machines via polynomial kernel in the vector space model. To complement this result, we exhibit a natural subclass of (softmax) transformers that completely characterizes semialgebraic counting properties. Through connections with the Hilbert's tenth problem, this expressivity of transformers also yields a new undecidability result for analyzing an extremely simple transformer model --- surprisingly with neither positional encodings (i.e. NoPE-transformers) nor masking. We also experimentally validate trainability of such counting properties.

### Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

## Core Task Landscape

This paper addresses: **Expressiveness of Transformers for Counting Properties**

A total of **50 papers** were analyzed and organized into a taxonomy with **12 categories**.

### Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Theoretical Expressiveness and Formal Characterization**
- **Empirical Analysis and Training Dynamics**
- **Practical Counting Applications**
- **Unrelated or Tangential Work**

### Complete Taxonomy Tree

- Expressiveness of Transformers for Counting Properties Survey Taxonomy
- Theoretical Expressiveness and Formal Characterization
  - Logic-Based Expressiveness Characterization ★ (5 papers)
    - [0] The Counting Power of Transformers (Anon et al., 2026) [View paper](#)
    - [5] Counting like transformers: Compiling temporal counting logic into softmax transformers (Chiang, 2024) [View paper](#)
    - [16] Logical languages accepted by transformer encoders with hard attention (Barcelo, 2023) [View paper](#)
    - [23] Tighter bounds on the expressivity of transformer encoders (Chiang, 2023) [View paper](#)
    - [49] Expressive Power of Graph Transformers via Logic (Ahvonen, 2025) [View paper](#)
    - Complexity-Theoretic Bounds and Limitations (4 papers)
      - [1] When Can Transformers Count to n? (Yehudai, 2024) [View paper](#)
      - [2] Nope: the counting power of transformers with no positional encodings (C KÄ¶cher, 2025) [View paper](#)
      - [32] Attention is turing-complete (J PÄ©rez, 2021) [View paper](#)
      - [47] The Expressivity of Fixed-Precision Transformers without Positional Encoding (N Negishi, 2025) [View paper](#)
    - Positional Encoding and Architectural Component Analysis (3 papers)
      - [25] Counting in Small Transformers: The Delicate Interplay between Attention and Feed-Forward Layers (Behrens, 2024) [View paper](#)
      - [30] Understanding counting in small transformers: The interplay between attention and feed-forward layers (F Behrens, 2024) [View paper](#)
      - [40] Contextual Position Encoding: Learning to Count What's Important (Golovneva, 2024) [View paper](#)
- Empirical Analysis and Training Dynamics
  - Length Generalization and Inductive Biases (2 papers)
    - [8] Language models need inductive biases to count inductively (Chang, 2024) [View paper](#)
    - [36] Autoregressive + Chain of Thought = Recurrent: Recurrence's Role in Language Models' Computability and a Revisit of Recurrent Transformer (Zhang Xiang, 2024) [View paper](#)
  - Mechanistic Interpretability of Counting (1 papers)
    - [13] Contextual Counting: A Mechanistic Study of Transformers on a Quantitative Task (Golkar, 2024) [View paper](#)
  - Training Procedures and Data Design (2 papers)
    - [9] Can Transformers Do Enumerative Geometry? (Hashemi Baran, 2024) [View paper](#)
    - [18] Teaching arithmetic to small transformers (Lee Nayoung, 2023) [View paper](#)

- Practical Counting Applications
  - Crowd Counting and Density Estimation
  - Fully-Supervised and Multimodal Crowd Counting (10 papers)
    - [4] A Transformer-based Multimodal Fusion Model for Efficient Crowd Counting Using Visual and Wireless Signals (Cui Zhe, 2025) [View paper](#)
    - [11] DMLViT: Dynamic multi-scale local vision transformer for object counting in congested traffic scenes (Chenxi Lin, 2025) [View paper](#)
    - [14] TransMambaCC: Integrating Transformer and Pyramid Mamba Network for RGB-T Crowd Counting (Yangjian Chen, 2025) [View paper](#)
    - [15] STCC: Scale-Aware Transformer for Crowd Counting (X Jia, 2025) [View paper](#)
    - [17] Countformer: Multi-view crowd counting transformer (Hong Mo, 2024) [View paper](#)
    - [19] Cctrans: Simplifying and improving crowd counting with transformer (Tian Ye, 2021) [View paper](#)
    - [26] Graph Enhancement and Transformer Aggregation Network for RGB-Thermal Crowd Counting (Pan Yi, 2024) [View paper](#)
    - [41] RCCFormer: A Robust Crowd Counting Network Based on Transformer (Liu Peng, 2025) [View paper](#)
    - [43] A Large-Scale Drone based Thermal Infrared Benchmark and Inception Transformer Network for Crowd Counting (X Wang, 2025) [View paper](#)
    - [44] CrowdFormer: An overlap patching vision transformer for top-down crowd counting. (Shaopeng Yang, 2022) [View paper](#)
  - Weakly-Supervised Crowd Counting (5 papers)
    - [10] Transcrowd: weakly-supervised crowd counting with transformers (Dingkang Liang, 2022) [View paper](#)
    - [21] Exploring crowd counting methodology by integrating CNN and transformer: performance optimization under weak supervision (Zengxi Feng, 2025) [View paper](#)
    - [35] CCTwins: A weakly supervised transformer-based crowd counting method with adaptive scene consistency attention (Li Dong, 2023) [View paper](#)
    - [37] Weakly supervised crowd counting with Joint CNN and Transformer network (Fusen Wang, 2025) [View paper](#)
    - [38] Application of improved transformer based on weakly supervised in crowd localization and crowd counting (Hui Gao, 2023) [View paper](#)
  - Object and Instance Counting (7 papers)
    - [7] Cstrans: Correlation-guided self-activation transformer for counting everything (Bin-Bin Gao, 2024) [View paper](#)
    - [12] Countgd: Multi-modal open-world counting (Niki Amini-Naieni, 2024) [View paper](#)
    - [20] CounTR: Transformer-based Generalised Visual Counting (Liu Chang, 2022) [View paper](#)
    - [27] Vision Transformer Off-the-Shelf: A Surprising Baseline for Few-Shot Class-Agnostic Counting (Cao Zhiguo, 2023) [View paper](#)
    - [29] Open-world Text-specified Object Counting (Amini-Naieni, 2023) [View paper](#)
    - [42] PBECCount: Prompt-Before-Extract Paradigm for Class-Agnostic Counting (Chun X, 2025) [View paper](#)
    - [48] Text-promptable Object Counting via Quantity Awareness Enhancement (Shi, 2025) [View paper](#)
  - Domain-Specific Counting Applications (7 papers)
    - [22] Pathway-enhanced Transformer-based model for robust enumeration of cell types from the cell-free transcriptome (S Yan, 2024) [View paper](#)
    - [24] Transformer for Tree Counting in Aerial Images (Guang Chen, 2022) [View paper](#)
    - [31] Vision Transformers for Weakly-Supervised Microorganism Enumeration (Javier Ureña Santiago, 2024) [View paper](#)
    - [33] Enhancing Fish Counting in Sonar Images With Multitask Learning and Local-Global Feature Interaction (Yuhang Wang, 2025) [View paper](#)
    - [34] TreeFormer: A semi-supervised transformer-based framework for tree counting from a single high-resolution image (Amirkolaee, 2023) [View paper](#)
    - [45] Intelligently Counting Agricultural Pests by Integrating SAM with FamNet (Jiajun Qing, 2024) [View paper](#)
    - [50] CTHNet: a network for wheat ear counting with local-global features fusion based on hybrid architecture (Qingqing Hong, 2024) [View paper](#)
  - Auxiliary Vision Tasks and Hybrid Architectures (4 papers)
    - [3] Fusion CNN-transformer model for target counting in complex scenarios (Xingyuan He, 2025) [View paper](#)
    - [28] Visual enumeration is challenging for large-scale generative ai (A Testolin, 2024) [View paper](#)
    - [39] Crowd behavior detection: leveraging video swin transformer for crowd size and violence level analysis (Marwa Qaraqe, 2024) [View paper](#)
    - [46] A Deep Learning-Based Method for Crowd Counting Using Shunting Inhibition Mechanism (Fok Hing Chi Tivive, 2024) [View paper](#)
- Unrelated or Tangential Work (1 papers)
  - [6] Iron: Private inference on transformers (M Hao, 2022) [View paper](#)

## Narrative

Core task: expressiveness of transformers for counting properties. The field divides into several main branches that reflect different angles on how well transformers can capture counting and related quantitative reasoning. Theoretical Expressiveness and Formal Characterization investigates the fundamental computational limits of transformer architectures through logic-based frameworks and complexity-theoretic analyses, asking which counting tasks are representable in principle and what architectural features enable or constrain this capacity. Works such as Logical Languages[16] and Tighter Bounds[23] exemplify efforts to map transformer expressiveness onto formal language hierarchies and circuit complexity classes. Empirical Analysis and Training Dynamics explores how transformers learn counting in practice, examining phenomena like length generalization, the role of positional encodings, and inductive biases that emerge during training—illustrated by studies like Counting Like Transformers[5] and Inductive Biases Count[8]. Practical Counting Applications encompasses a dense branch of vision-based crowd counting, object enumeration, and domain-specific tasks (e.g., cell counting, tree counting from aerial imagery), where transformer-based architectures are deployed and refined for real-world performance. A smaller set of works falls outside these core themes, addressing tangential or unrelated topics.

Within the theoretical branch, a particularly active line of inquiry uses logical languages to characterize what transformers can and cannot express, contrasting the power of attention mechanisms with classical automata and first-order logic extensions. Counting Power[0] sits squarely in this logic-based expressiveness cluster, closely aligned with Logical Languages[16] and Graph Logic[49], which similarly formalize counting capabilities through logical frameworks. Compared to Tighter Bounds[23], which emphasizes circuit-theoretic lower bounds, Counting Power[0] focuses more directly on the interplay between transformer depth, attention structure, and the ability to solve counting problems defined in logical terms. Meanwhile, empirical studies like When Count[1] and Teaching Arithmetic[18] probe whether trained models exhibit the counting behaviors predicted by theory, highlighting an ongoing tension

between what transformers can represent in principle and what they reliably learn from data. This landscape reveals open questions about bridging formal guarantees with practical learnability and scaling.

## Related Works in Same Category

---

The following **4 sibling papers** share the same taxonomy leaf node with the original paper:

### 1. Counting like transformers: Compiling temporal counting logic into softmax transformers

**Authors:** Chiang, David, Andy Yang, David Chiang | **Year/Venue:** 2024 | **URL:** [View paper](#)

#### Abstract

Deriving formal bounds on the expressivity of transformers, as well as studying transformers that are constructed to implement known algorithms, are both effective methods for better understanding the computational power of transformers. Towards both ends, we introduce the temporal counting logic  $\text{Kt}\{t\}\{#\}$  alongside the RASP variant  $\text{C-RASP}\{t\}\{#\}$ . We show they are equivalent to each other, and that together they are the best-known lower bound on the formal expressivity of futu...

#### Relationship Analysis

Both papers belong to the Logic-Based Expressiveness Characterization category, establishing formal equivalences between transformer variants and logical systems. The original paper focuses on characterizing transformers' ability to express semialgebraic counting properties (polynomial inequalities) and proves that transformers can capture all such properties, while the candidate paper establishes an equivalence between soft attention transformers and the temporal counting logic  $\text{Kt}\{t\}\{#\}$ , which captures linear counting properties through temporal logic formalism. The key difference is that the original paper demonstrates expressiveness for highly nonlinear (polynomial) counting properties, whereas the candidate paper provides a tighter characterization for linear counting properties using temporal logic, representing different levels of counting complexity within the same logical characterization framework.

---

### 2. Logical languages accepted by transformer encoders with hard attention

**Authors:** Barcelo, Pablo, Pablo Barcelo<sup>3</sup>, Kozachinskiy, Alexander, et al. (14 authors total) | **Year/Venue:** 2023 | **URL:** [View paper](#)

#### Abstract

We contribute to the study of formal languages that can be recognized by transformer encoders. We focus on two self-attention mechanisms: (1) UHAT (Unique Hard Attention Transformers) and (2) AHAT (Average Hard Attention Transformers). UHAT encoders are known to recognize only languages inside the circuit complexity class  $\text{AC}^0$ , i.e., accepted by a family of poly-sized and depth-bounded boolean circuits with unbounded fan-ins. On the other hand, AHAT encoders can recognize languages outsi...

#### Relationship Analysis

Both papers belong to the Logic-Based Expressiveness Characterization category, establishing formal logical equivalences for transformer variants. They overlap in studying the expressiveness of transformers through logical frameworks, with both examining hard attention mechanisms and their relationship to formal logic systems. The original paper focuses on counting properties and semialgebraic expressiveness using softmax and average hard attention transformers, while the candidate paper characterizes unique hard attention transformers (UHAT) and average hard attention transformers (AHAT) through first-order logic with numerical predicates (FO(Mon)) and temporal logic extensions (LTL), demonstrating that UHATs capture FO(Mon) languages within  $\text{AC}^0$  and AHATs can express counting properties beyond  $\text{AC}^0$ .

---

### 3. Tighter bounds on the expressivity of transformer encoders

**Authors:** Chiang, David, Cholak, Peter, Pillay, et al. (8 authors total) | **Year/Venue:** 2023 | **URL:** [View paper](#)

#### Abstract

Characterizing neural networks in terms of better-understood formal systems has the potential to yield new insights into the power and limitations of these networks. Doing so for transformers remains an active area of research. Bhattamishra and others have shown that transformer encoders are at least as expressive as a certain kind of counter machine, while Merrill and Sabharwal have shown that fixed-precision transformer encoders recognize only languages in uniform  $\text{TC}^0$ . We connect and streng...

#### Relationship Analysis

Both papers belong to the Logic-Based Expressiveness Characterization category, establishing formal logical equivalences for transformer expressiveness. They overlap in using logic-based frameworks to characterize transformers' counting capabilities, with the original paper focusing on semialgebraic counting properties (polynomial inequalities) and the candidate paper establishing equivalences with first-order logic with counting quantifiers (FOC[+; MOD]). The key difference is that the original paper proves transformers can express all semialgebraic counting properties (arbitrary polynomial degree), while the candidate paper provides tighter bounds by characterizing fixed-precision transformers through a specific logical system with linear counting constraints.

---

### 4. Expressive Power of Graph Transformers via Logic

**Authors:** Ahvonen, Veeti, Funk, Maurice, Heiman, et al. (10 authors total) | **Year/Venue:** 2025 • arXiv.org | **URL:** [View paper](#)

#### Abstract

Transformers are the basis of modern large language models, but relatively little is known about their precise expressive power on graphs. We study the expressive power of graph transformers (GTs) by Dwivedi and Bresson (2020) and GPS-networks by Ramp'asek et al. (2022), both under soft-attention and average hard-attention. Our study covers two scenarios: the theoretical setting with real numbers and the more practical case with floats. With reals, we show that in restriction to vertex properti...

#### Relationship Analysis

Both papers belong to the Logic-Based Expressiveness Characterization category, establishing formal logical equivalences for transformer variants. While the original paper characterizes transformers' counting power through semialgebraic properties and connections to polynomial inequalities, the candidate paper characterizes graph transformers through modal logic frameworks (GML with global modality). The key difference is that the original focuses on counting properties in sequential transformers using polynomial expressions, whereas the candidate examines graph transformers' expressive power through graded modal logic and propositional logic with global modalities.

---

## Contributions Analysis

**Overall novelty summary.** The paper establishes that transformers can express semialgebraic counting properties—those defined by boolean combinations of arbitrary multivariate polynomials—and provides a formal framework for analyzing counting power. It resides in the Logic-Based Expressiveness Characterization leaf, which contains five papers total, including the original work. This leaf sits within the broader Theoretical Expressiveness and Formal Characterization branch, indicating a moderately populated research direction focused on formal proofs rather than empirical validation. The sibling papers similarly use logical or formal language frameworks to characterize transformer expressiveness, suggesting this is an active but not overcrowded niche.

The taxonomy reveals that theoretical expressiveness research divides into logic-based characterizations, complexity-theoretic bounds, and positional encoding analyses. The original paper's leaf neighbors the Complexity-Theoretic Bounds leaf (four papers) and the Positional Encoding leaf (three papers), both addressing complementary aspects of transformer power. The taxonomy's scope\_note clarifies that logic-based work establishes equivalences with formal systems, while complexity studies prove computational limits—this paper bridges both by using logical frameworks to demonstrate nonlinear counting capacity. The broader Empirical Analysis branch (seven papers across three leaves) explores whether these theoretical capabilities manifest in practice, highlighting a field-wide gap between formal expressiveness and learned behavior.

Among twenty candidates examined, Contribution A (formal framework) showed no clear refutation across ten candidates, suggesting novelty in how the framework is structured. Contribution B (semialgebraic properties) examined four candidates and found one refutable match, indicating some prior work addresses nonlinear counting expressiveness, though the scale of overlap remains unclear from this limited search. Contribution C (uniform average hard attention characterization) examined six candidates with one refutable match, suggesting partial overlap in architectural characterization approaches. The analysis explicitly covers top-K semantic matches plus citation expansion, not an exhaustive survey of all transformer expressiveness literature.

Given the limited search scope of twenty candidates, the paper appears to advance the logic-based expressiveness agenda by moving beyond semilinear properties to semialgebraic ones, though at least two contributions show some prior work overlap. The taxonomy context suggests this work extends an established research direction rather than opening an entirely new area. A broader literature search might reveal additional overlapping results, particularly in the complexity-theoretic or empirical branches not fully covered here.

---

This paper presents **3 main contributions**, each analyzed against relevant prior work:

### **Contribution 1: Formal framework for counting power of transformers**

**Description:** The authors introduce a formal framework to study the expressiveness of transformers for counting properties, which are permutation-closed languages where acceptance depends only on the number of occurrences of tokens rather than their order.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

#### **1. When Can Transformers Count to n?**

URL: [View paper](#)

##### **Brief Assessment**

When Count[1] focuses on specific counting tasks (query counting and most frequent element) with architectural constraints, while the original paper introduces a broader formal framework studying expressiveness for general counting properties as permutation-closed languages. The candidate does not challenge the novelty of the original's formal framework approach.

---

#### **2. Tighter bounds on the expressivity of transformer encoders**

URL: [View paper](#)

##### **Brief Assessment**

Tighter Bounds[23] focuses on characterizing transformer expressivity through first-order logic with counting quantifiers (FOC[+; mod]), studying what formal languages transformers can recognize. The original paper's framework specifically studies counting properties (permutation-closed languages) and their semialgebraic characterization, which is a different formal approach than the logic-based characterization in the candidate.

---

#### **3. Homomorphism Counts as Structural Encodings for Graph Learning**

URL: [View paper](#)

##### **Brief Assessment**

Homomorphism Encodings[55] focuses on graph learning with structural encodings based on homomorphism counts for graph transformers, not on formal frameworks for studying transformer expressiveness on counting tasks in sequential/textual data.

---

#### **4. Nope: the counting power of transformers with no positional encodings**

URL: [View paper](#)

##### **Brief Assessment**

Nope Counting[2] focuses on transformers without positional encodings (NoPE) and their ability to express semi-algebraic counting properties. The original paper's framework is broader, studying general counting properties with various transformer architectures including those with positional encodings.

---

#### **5. Counting like transformers: Compiling temporal counting logic into softmax transformers**

URL: [View paper](#)

##### **Brief Assessment**

Counting Like Transformers[5] focuses on temporal counting logic (KT[#]) and compiling it into transformers for recognizing formal languages with counting properties. The ORIGINAL paper introduces a framework specifically for studying semialgebraic counting properties (polynomial inequalities), which is a different mathematical formalism than temporal logic.

---

#### **6. Dynamic unary convolution in transformers**

URL: [View paper](#)

##### **Brief Assessment**

Dynamic Unary[54] focuses on unary convolution operations in transformers for computer vision tasks, not on formal frameworks for studying expressiveness of transformers on counting properties or formal language theory.

---

#### **7. Trees in transformers: a theoretical analysis of the Transformer's ability to represent trees**

URL: [View paper](#)

##### **Brief Assessment**

Trees Analysis[56] focuses on transformers' ability to represent tree structures and tree backbones, not on counting properties or permutation-closed languages. The formal frameworks address entirely different aspects of transformer expressiveness.

---

#### **8. Language models need inductive biases to count inductively**

URL: [View paper](#)

##### **Brief Assessment**

Inductive Biases Count[8] focuses on empirical training experiments for length generalization in counting tasks, not on formal frameworks for studying transformer expressiveness. The candidate investigates practical learning dynamics and positional embeddings rather than theoretical expressivity characterizations.

---

## 9. Counting ability of large language models and impact of tokenization

URL: [View paper](#)

### Brief Assessment

Tokenization Impact[57] focuses on how tokenization choices affect counting performance in LLMs, not on establishing a formal framework for studying transformer expressiveness on counting properties. The candidate examines practical counting failures due to BPE tokenization rather than theoretical expressiveness frameworks.

---

## 10. Contextual Counting: A Mechanistic Study of Transformers on a Quantitative Task

URL: [View paper](#)

### Brief Assessment

Contextual Counting[13] focuses on a specific counting task (counting ones within delimited regions) and mechanistic interpretability, not on establishing a formal framework for studying the expressiveness of transformers for general counting properties or permutation-closed languages.

---

## Contribution 2: Transformers capture semialgebraic counting properties

**Description:** The main theoretical result demonstrates that transformers can express counting properties defined by boolean combinations of polynomial inequalities of arbitrary degree, generalizing beyond the linear counting properties captured by existing models like C-RASP.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. BOLD: Boolean Logic Deep Learning

URL: [View paper](#)

#### Brief Assessment

BOLD Logic[59] focuses on boolean logic deep learning for neural network training with binary weights and activations, not on transformer expressiveness or counting properties with polynomial inequalities.

---

### 2. Transformers in Uniform TC

URL: [View paper](#)

#### Brief Assessment

Uniform TC[58] focuses on circuit complexity characterization of transformers (showing they are in dlogtime-uniform TC0), not on their ability to express counting properties defined by polynomial inequalities. The papers address fundamentally different aspects of transformer expressiveness.

---

### 3. Nope: the counting power of transformers with no positional encodings

URL: [View paper](#)

#### Prior Art Analysis

Nope Counting[2] demonstrates that transformers without positional encodings (NoPE-AHAT) can express semi-algebraic counting properties, which are defined as boolean combinations of polynomial inequalities. This directly overlaps with the original paper's claim about transformers capturing semialgebraic counting properties. Both papers prove that transformers can express counting properties defined by polynomial inequalities of arbitrary degree, with Nope Counting[2] providing the characterization theorem 1 that establishes this equivalence.

#### Evidence

Evidence 1 - **Rationale:** Both papers define semi-algebraic counting properties identically as boolean combinations of polynomial inequalities, establishing the same mathematical framework. - **Original:** our main result is that transformers can express counting properties that are highly nonlinear. more precisely, we prove that transformers can capture all semialgebraic counting properties, i.e., expressible as a boolean combination of arbitrary multivariate polynomials (of any degree). - **Candidate:** a subset  $s \subseteq \mathbb{N}^m$  is semi-algebraic if it is a boolean combination of sets of the form  $\{x \in \mathbb{N}^m \mid p(x) > 0\}$  for some polynomial  $p \in \mathbb{Z}[x_1, \dots, x_m]$ . a language  $l \subseteq \Sigma^*$  is semi-algebraic if there is a semi-algebraic set  $s \subseteq \mathbb{N}^m$  and  $\sigma = \{a_1, \dots, a_m\}$  such that  $l = \{w \in \{a_1, \dots, a_m\}^* \mid \psi(w) \in s\}$ ...

Evidence 2 - **Rationale:** Both papers prove that transformers can capture all semi-algebraic counting properties. Nope Counting[2]'s theorem 1 establishes this equivalence for NoPE-AHAT, demonstrating the same capability claimed in the original paper. - **Original:** theorem 1.1. transformers can capture all semialgebraic counting properties, i.e., those expressible as a boolean combination of inequalities between multivariate polynomials, where each variable counts the number of occurrences of a specific token in the text. - **Candidate:** theorem 1. nope-ahat = nope-ahat[u] = semialg.

Evidence 3 - **Rationale:** Both papers prove the identical proposition that transformers can express languages defined by polynomial inequalities, using nearly identical notation and formulation. - **Original:** proposition 3.1. for every polynomial  $p \in \mathbb{Z}[x_1, \dots, x_m]$ , the language  $lp > 0 = \{w \in \{a_1, \dots, a_m\}^* \mid p(\psi(w)) > 0\}$  belongs to nope-ahat[u]. thus,  $lp > 0$  is in smat. - **Candidate:** proposition 9. for every polynomial  $p \in \mathbb{Z}[x_1, \dots, x_m]$ , the language  $lp > 0 = \{w \in \{a_1, \dots, a_m\}^* \mid p(\psi(w)) > 0\}$  belongs to nope-ahat[u].

---

### 4. Solving Mathematical Problems with Transformers

URL: [View paper](#)

#### Brief Assessment

Mathematical Problems[60] focuses on solving mathematical problems with transformers and analyzing their complexity classes, not on expressing counting properties defined by polynomial inequalities. The candidate does not address semialgebraic counting properties or boolean combinations of polynomial inequalities.

---

## Contribution 3: Characterization via uniform average hard attention transformers

**Description:** The authors prove that NoPE-AHAT and NoPE-AHAT[U] (transformers without positional encodings using average hard attention with uniform layers) precisely capture semialgebraic counting properties, providing a complete characterization of this expressiveness class.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

## 1. Expressive Power of Graph Transformers via Logic

URL: [View paper](#)

### Brief Assessment

Graph Logic[49] focuses on graph transformers' expressive power via modal logic characterizations, not on characterizing counting properties through uniform average hard attention transformers for sequential text data.

---

## 2. Transformers as transducers

URL: [View paper](#)

### Brief Assessment

Transformers Transducers[51] focuses on sequence-to-sequence transductions and finite transducers, not on characterizing counting property expressiveness through uniform average hard attention transformers.

---

## 3. ESSLLI 2024 Course Notes Expressivity of Transformers: Logic, Circuits, and Formal Languages

URL: [View paper](#)

### Brief Assessment

ESSLLI Expressivity[53] focuses on characterizing transformer expressivity through formal language theory and circuit complexity, not on counting properties or semialgebraic characterizations. The candidate discusses average-hard attention in the context of simulating Turing machines and general expressivity bounds, which is a different research direction from the original paper's semialgebraic counting property characterization.

---

## 4. Logical languages accepted by transformer encoders with hard attention

URL: [View paper](#)

### Brief Assessment

Logical Languages[16] focuses on hard attention mechanisms (UHAT and AHAT) for recognizing formal languages, not on characterizing semialgebraic counting properties. The candidate's AHAT results concern logical language acceptance (e.g., LTL(C,+) formulas), which is a different expressiveness framework than the original paper's semialgebraic counting property characterization.

---

## 5. SmartPod: An Automated Framework for High-Precision Soybean Pod Counting in Field Phenotyping

URL: [View paper](#)

### Brief Assessment

SmartPod Soybean[52] focuses on soybean pod counting using vision transformers for agricultural phenotyping, not on theoretical characterization of transformer expressiveness for counting properties or semialgebraic sets.

---

## 6. Nope: the counting power of transformers with no positional encodings

URL: [View paper](#)

### Prior Art Analysis

Nope Counting[2] provides the exact characterization claimed in the original paper, proving that NoPE-AHAT and NoPE-AHAT[U] (transformers without positional encodings using average hard attention with uniform layers) precisely capture semialgebraic counting properties. The candidate paper's theorem 1 establishes this equivalence, demonstrating that this characterization was already known.

### Evidence

Evidence 1 - **Rationale:** Both papers prove the identical characterization theorem that NoPE-AHAT[U] precisely captures semi-algebraic counting properties, establishing the same equivalence. - **Original:** theorem 1.2. nope-ahat and nope-ahat[u] capture precisely semialgebraic counting properties. in particular, as far as expressing counting properties, nope-ahat is a subset of smat. - **Candidate:** theorem 1. nope-ahat = nope-ahat[u] = semialg.

Evidence 2 - **Rationale:** Both papers define uniform attention layers identically, where key and query maps are constant (or zero), establishing the same technical framework for the characterization. - **Original:** we say that  $\lambda$  is a softmax (resp. aha) layer if  $w_t = \text{softmax}$  (resp. aha). we say that it is a uniform-aha layer if it is an aha layer such that  $kx = qx = 0$  for all  $x$ , i.e.,  $(kx, qy) = 0$  for all  $x$  and  $y$ . - **Candidate:** we say  $\lambda$  is uniform iff the key and query maps  $k$  and  $q$  are constant

Evidence 3 - **Rationale:** Both papers prove the same proposition establishing one direction of the characterization, showing that NoPE-AHAT languages are contained in semi-algebraic sets. - **Original:** proposition 4.1. nope-ahat  $\subseteq$  semialg. - **Candidate:** proposition 10. nope-ahat  $\subseteq$  semialg.

---

## Appendix: Text Similarity Detection

Textual similarity detection checked 18 papers and found 6 similarity segment(s) across 2 paper(s).

The following **2 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

### 1. Logical languages accepted by transformer encoders with hard attention

**Detected in:** Core Task (sibling), Contribution: contribution\_3

⚠ **Note:** This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

---

### 2. Nope: the counting power of transformers with no positional encodings

**Detected in:** Contribution: contribution\_1, Contribution: contribution\_2, Contribution: contribution\_3

⚠ **Note:** This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

---

## References

- [0] The Counting Power of Transformers [View paper](#)
- [1] When Can Transformers Count to  $n$ ? [View paper](#)
- [2] Nope: the counting power of transformers with no positional encodings [View paper](#)
- [3] Fusion CNN-transformer model for target counting in complex scenarios [View paper](#)
- [4] A Transformer-based Multimodal Fusion Model for Efficient Crowd Counting Using Visual and Wireless Signals [View paper](#)
- [5] Counting like transformers: Compiling temporal counting logic into softmax transformers [View paper](#)
- [6] Iron: Private inference on transformers [View paper](#)

- [7] Cstrans: Correlation-guided self-activation transformer for counting everything [View paper](#)
- [8] Language models need inductive biases to count inductively [View paper](#)
- [9] Can Transformers Do Enumerative Geometry? [View paper](#)
- [10] Transcrowd: weakly-supervised crowd counting with transformers [View paper](#)
- [11] DMLViT: Dynamic multi-scale local vision transformer for object counting in congested traffic scenes [View paper](#)
- [12] Countgd: Multi-modal open-world counting [View paper](#)
- [13] Contextual Counting: A Mechanistic Study of Transformers on a Quantitative Task [View paper](#)
- [14] TransMambaCC: Integrating Transformer and Pyramid Mamba Network for RGB-T Crowd Counting [View paper](#)
- [15] STCC: Scale-Aware Transformer for Crowd Counting [View paper](#)
- [16] Logical languages accepted by transformer encoders with hard attention [View paper](#)
- [17] Countformer: Multi-view crowd counting transformer [View paper](#)
- [18] Teaching arithmetic to small transformers [View paper](#)
- [19] Cctrans: Simplifying and improving crowd counting with transformer [View paper](#)
- [20] CounTR: Transformer-based Generalised Visual Counting [View paper](#)
- [21] Exploring crowd counting methodology by integrating CNN and transformer: performance optimization under weak supervision [View paper](#)
- [22] Pathway-enhanced Transformer-based model for robust enumeration of cell types from the cell-free transcriptome [View paper](#)
- [23] Tighter bounds on the expressivity of transformer encoders [View paper](#)
- [24] Transformer for Tree Counting in Aerial Images [View paper](#)
- [25] Counting in Small Transformers: The Delicate Interplay between Attention and Feed-Forward Layers [View paper](#)
- [26] Graph Enhancement and Transformer Aggregation Network for RGB-Thermal Crowd Counting [View paper](#)
- [27] Vision Transformer Off-the-Shelf: A Surprising Baseline for Few-Shot Class-Agnostic Counting [View paper](#)
- [28] Visual enumeration is challenging for large-scale generative ai [View paper](#)
- [29] Open-world Text-specified Object Counting [View paper](#)
- [30] Understanding counting in small transformers: The interplay between attention and feed-forward layers [View paper](#)
- [31] Vision Transformers for Weakly-Supervised Microorganism Enumeration [View paper](#)
- [32] Attention is turing-complete [View paper](#)
- [33] Enhancing Fish Counting in Sonar Images With Multitask Learning and Local-Global Feature Interaction [View paper](#)
- [34] TreeFormer: A semi-supervised transformer-based framework for tree counting from a single high-resolution image [View paper](#)
- [35] CCTwins: A weakly supervised transformer-based crowd counting method with adaptive scene consistency attention [View paper](#)
- [36] Autoregressive + Chain of Thought = Recurrent: Recurrence's Role in Language Models' Computability and a Revisit of Recurrent Transformer [View paper](#)
- [37] Weakly supervised crowd counting with Joint CNN and Transformer network [View paper](#)
- [38] Application of improved transformer based on weakly supervised in crowd localization and crowd counting [View paper](#)
- [39] Crowd behavior detection: leveraging video swin transformer for crowd size and violence level analysis [View paper](#)
- [40] Contextual Position Encoding: Learning to Count What's Important [View paper](#)
- [41] RCCFormer: A Robust Crowd Counting Network Based on Transformer [View paper](#)
- [42] PBECCount: Prompt-Before-Extract Paradigm for Class-Agnostic Counting [View paper](#)
- [43] A Large-Scale Drone based Thermal Infrared Benchmark and Inception Transformer Network for Crowd Counting [View paper](#)
- [44] CrowdFormer: An overlap patching vision transformer for top-down crowd counting. [View paper](#)
- [45] Intelligently Counting Agricultural Pests by Integrating SAM with FamNet [View paper](#)
- [46] A Deep Learning-Based Method for Crowd Counting Using Shunting Inhibition Mechanism [View paper](#)
- [47] The Expressivity of Fixed-Precision Transformers without Positional Encoding [View paper](#)
- [48] Text-promptable Object Counting via Quantity Awareness Enhancement [View paper](#)
- [49] Expressive Power of Graph Transformers via Logic [View paper](#)
- [50] CTHNet: a network for wheat ear counting with local-global features fusion based on hybrid architecture [View paper](#)
- [51] Transformers as transducers [View paper](#)
- [52] SmartPod: An Automated Framework for High-Precision Soybean Pod Counting in Field Phenotyping [View paper](#)
- [53] ESSLI 2024 Course Notes Expressivity of Transformers: Logic, Circuits, and Formal Languages [View paper](#)
- [54] Dynamic unary convolution in transformers [View paper](#)
- [55] Homomorphism Counts as Structural Encodings for Graph Learning [View paper](#)
- [56] Trees in transformers: a theoretical analysis of the Transformer's ability to represent trees [View paper](#)
- [57] Counting ability of large language models and impact of tokenization [View paper](#)
- [58] Transformers in Uniform TC [View paper](#)
- [59] BOLD: Boolean Logic Deep Learning [View paper](#)
- [60] Solving Mathematical Problems with Transformers [View paper](#)