# Exploring the Zero-Shot Potential of Large Language Models for Detecting Algorithmically Generated Domains

Tomás Pelayo-Benedet

University of Zaragoza, Spain

# Introduction & Motivation

**The Threat: Domain Generation Algorithms (DGAs)**

- Malware uses DGAs to establish resilient Command & Control (C&C) channels.
- Dynamically generates thousands of domains (AGDs) to evade static forbiden lists.
- Traditional detection relies on statistical features or DL (LSTM/CNN) requiring specific training and achiving great classification (ACC 99%) with their test datasets, but **fail classifying families which are no present in training dataset**.

# Introduction & Motivation

**The Threat: Domain Generation Algorithms (DGAs)**

- Malware uses DGAs to establish resilient Command & Control (C&C) channels.
- Dynamically generates thousands of domains (AGDs) to evade static forbiden lists.
- Traditional detection relies on statistical features or DL (LSTM/CNN) requiring specific training and achiving great classification (ACC 99%) with their test datasets, but **fail classifying families which are no present in training dataset**.

**The Opportunity: Large Language Models (LLMs)**

- LLMs have pattern recognition capabilities from pre-training.
- **Gap:** Can general-purpose LLMs detect malicious AGDs *without* domain-specific fine-tuning?
- **Goal:** Establish a baseline for Zero-Shot and Few-Shot AGD detection using commercial LLMs.

# What are Domain Generation Algorithms (DGAs)?

**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**

# What are Domain Generation Algorithms (DGAs)?

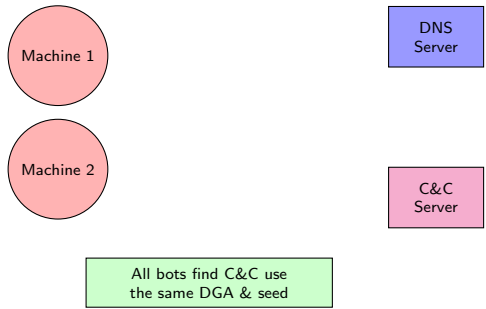
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

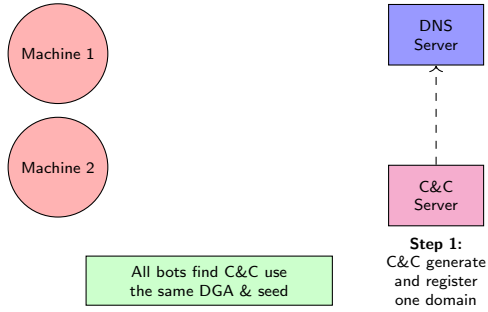
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

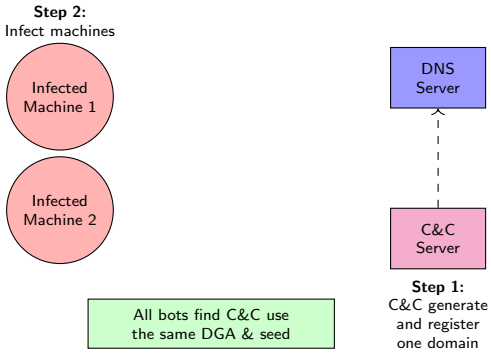
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

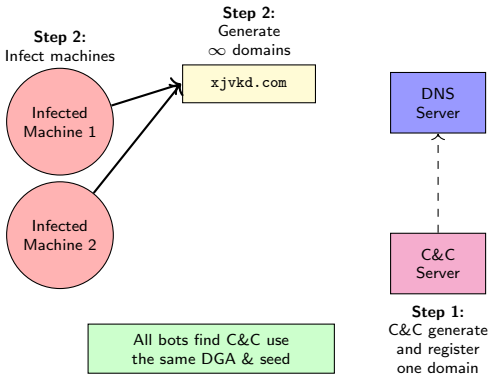
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

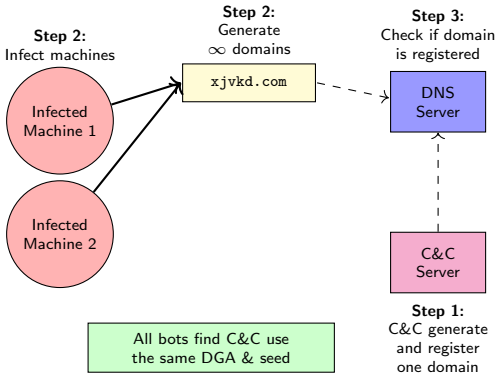
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

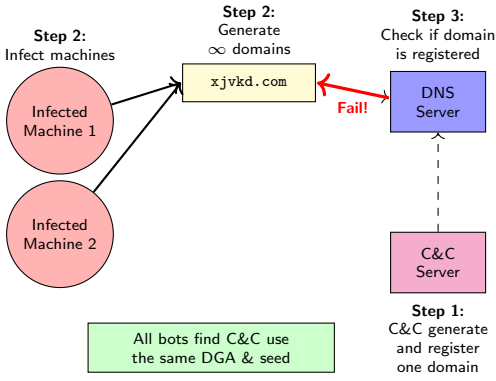
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

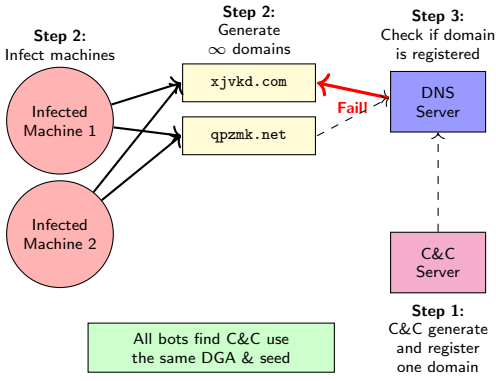
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

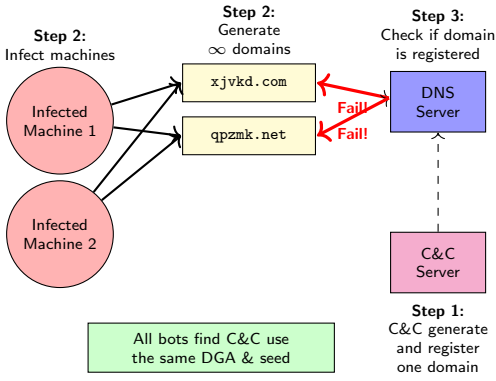
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

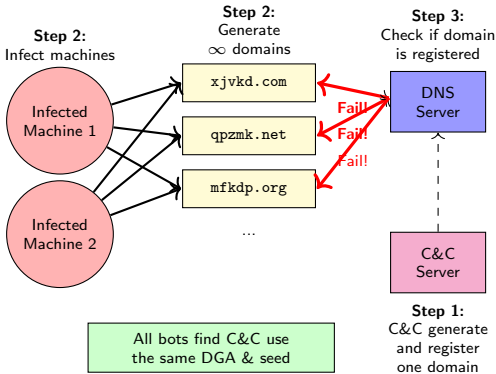
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

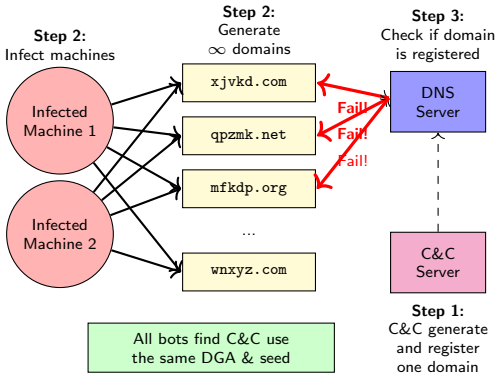
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

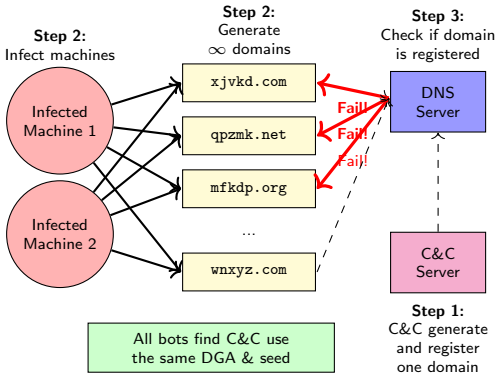
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

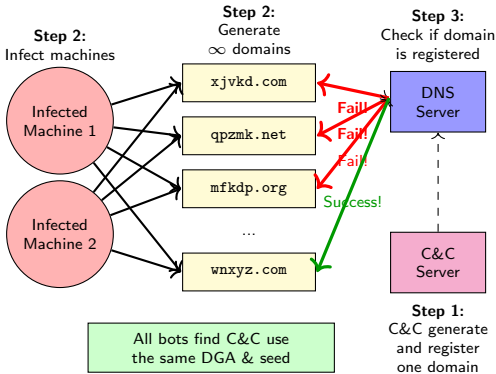
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

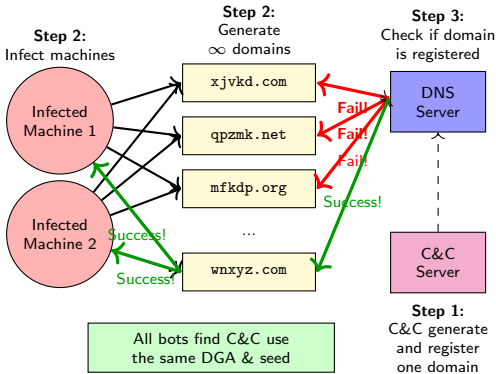
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

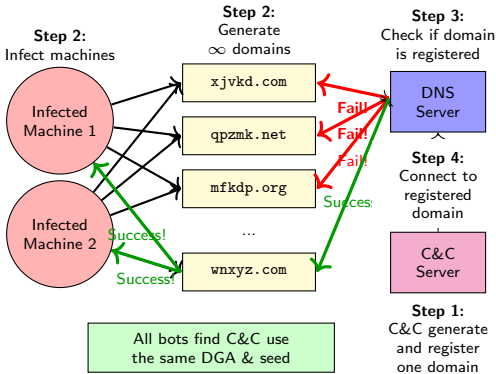
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# What are Domain Generation Algorithms (DGAs)?

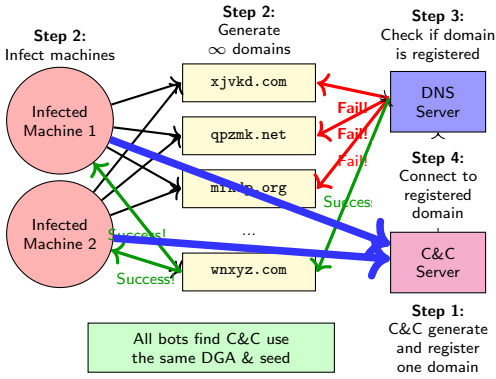
**Definition**

- Malware technique to generate large numbers of pseudo-random domain names
- Used by botnets to establish Command & Control (C&C) communications
- Enables evasion of traditional blocklists and detection efforts

**Key Characteristics**

- Thousands of domains generated daily
- Only few domains actually registered
- Malware and C&C share DGA & seed

**Botnet Communication**



# Research Questions

We formulated three key questions to guide the evaluation:

- **Detection Capabilities**
  - How effective are LLMs in binary classification (Malicious vs. Benign) based solely on the domain string?
  - Does providing linguistic feature context improve accuracy?
- **Family Classification**
  - Can LLMs distinguish between specific malware families (e.g., Conficker vs. Banjori)?
- **Real-World Robustness**
  - How do LLMs perform against real-world, non-malicious DNS traffic that may share structural similarities with AGDs?

# Experimental Methodology

**Models Evaluated (9 Models, 4 Providers)**

- **Paid:** GPT-4o, GPT-4o-mini, Claude 3.5 Sonnet, Claude 3.5 Haiku.
- **Free:** Gemini 1.5 Pro, Flash, Flash-8B; Mistral Large, Small.

**Datasets**

- $D_1$ **(Binary):** 25k Malicious (DGArchive) + 25k Benign (Tranco).
- $D_2$ **(Multiclass):** 50k Malicious from 25 distinct malware families.
- $D_3$ **(Real-World):** 50k benign domains from University of Zaragoza DNS logs.

**Iterative Prompting Strategy ($P_1 - P_4$)**

- $P_1$: Minimal (Zero-shot).
- $P_2$: Lexical features context (randomness, pronounceability).
- $P_3$: Few-shot (10 examples per family).
- $P_4$: Real-world context (focus on TLD/SLD tuples).

# Binary Detection Performance

**Key Findings:**

- **Relative High Accuracy:** Models achieved between **77.3% and 89.3%** accuracy without fine-tuning.
- **Prompt Engineering:** Adding lexical guidance ($P_2$) yielded minimal improvement over simple prompt ($P_1$).

**The False Positive (FP) Problem:**

- All models exhibited high False Positive Rates (FPR), ranging from **13.2% to 31.9%**.
- **Impact:** In a network with 1M queries/day, this would block 130k+ legitimate domains.

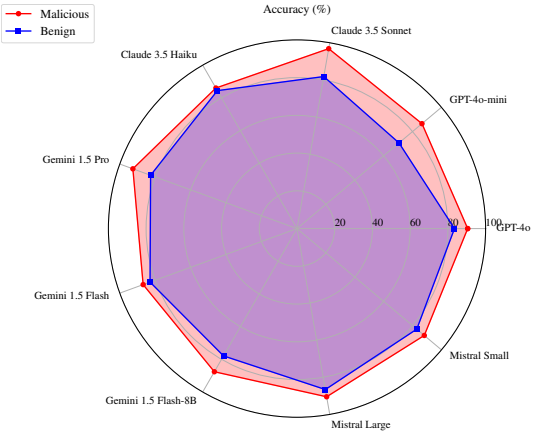



Figure: $P_2$ accuracy

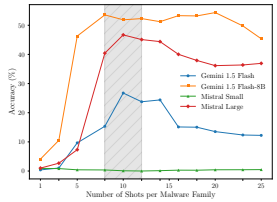
# Malware Family Classification

**Few-Shot Learning Setup ($P_3$)**

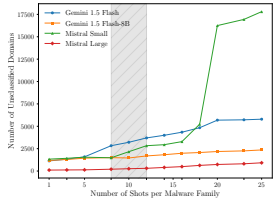
- 10-shot learning proved to be the optimal balance point.
- Performance varied significantly by DGA generation scheme.

**Performance by DGA Type:**

- **Hash-based (e.g., Dyre): Excellent.** Claude 3.5 Sonnet achieved nearly 100% accuracy. Hexadecimal patterns are easy for LLMs.
- **Arithmetic-based: Mixed.** Good for complex patterns, poor for simple ones (e.g., Conficker).
- **Dictionary-based (e.g., Suppobox): Poor.** Models struggle to distinguish "random word combinations" from legitimate domains.




(a) accuracy




(b) API fail (because allucination)

| Family | Claude Sonnet 3.5 | | | Gemini 1.5 Pro | | | Gemini 1.5 Flash-8B | | | Mistral Large | | | Type |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prec | Rec | F1 | Prec | Rec | F1 | Prec | Rec | F1 | Prec | Rec | F1 | |
| banjori | 99.9 | 99.6 | 99.7 | 96.9 | 98.9 | 97.9 | 83.0 | 63.9 | 72.2 | 96.2 | 63.0 | 76.1 | |
| conficker | 96.9 | 97.6 | 97.2 | 85.9 | 69.1 | 76.6 | 54.1 | 31.1 | 39.5 | 76.5 | 14.1 | 23.8 | |
| emotet | 98.2 | 100.0 | 99.1 | 97.2 | 98.9 | 98.0 | 98.2 | 99.5 | 98.8 | 95.1 | 42.4 | 58.6 | |
| flubot | 98.9 | 97.1 | 98.0 | 89.5 | 76.1 | 82.3 | 65.4 | 64.9 | 65.1 | 69.8 | 39.5 | 50.5 | |
| gameover | 99.6 | 97.0 | 98.3 | 94.6 | 92.2 | 93.4 | 82.0 | 66.7 | 73.6 | 92.0 | 44.4 | 59.9 | |
| metastealer | 100.0 | 100.0 | 100.0 | 99.7 | 100.0 | 99.8 | 95.4 | 99.7 | 97.5 | 95.0 | 99.6 | 97.2 | |
| necurs | 97.0 | 89.9 | 93.3 | 80.9 | 60.6 | 69.3 | 78.3 | 59.7 | 67.8 | 52.0 | 43.7 | 47.5 | Arithmetic |
| nymaim | 97.6 | 98.7 | 98.1 | 73.6 | 93.6 | 82.4 | 77.4 | 85.2 | 81.1 | 73.0 | 88.4 | 80.0 | |
| pitou | 99.7 | 99.5 | 99.6 | 96.2 | 93.4 | 94.8 | 82.2 | 84.4 | 83.3 | 89.0 | 77.4 | 82.8 | |
| pushdo | 99.2 | 99.9 | 99.6 | 98.6 | 98.6 | 98.6 | 95.4 | 96.4 | 95.9 | 94.8 | 97.7 | 96.2 | |
| qakbot | 96.0 | 98.6 | 97.3 | 85.5 | 84.4 | 84.9 | 60.6 | 49.4 | 54.5 | 38.1 | 57.4 | 45.8 | |
| rovnix | 100.0 | 100.0 | 100.0 | 95.5 | 99.8 | 97.6 | 57.0 | 35.8 | 44.0 | 87.4 | 96.1 | 91.5 | |
| virut | 99.0 | 100.0 | 99.5 | 94.8 | 100.0 | 97.3 | 95.1 | 99.8 | 97.4 | 79.8 | 99.5 | 88.5 | |
| zloader | 100.0 | 100.0 | 100.0 | 97.1 | 97.3 | 97.2 | 70.9 | 93.4 | 80.6 | 74.7 | 44.6 | 55.8 | |
| gozi | 97.2 | 99.9 | 98.5 | 95.8 | 97.2 | 96.5 | 55.9 | 90.3 | 69.1 | 90.1 | 87.5 | 88.7 | |
| matsnu | 94.4 | 99.6 | 96.9 | 80.5 | 92.6 | 86.1 | 75.2 | 95.1 | 84.0 | 7.9 | 63.0 | 14.0 | Dictionary |
| nymaim2 | 96.1 | 99.6 | 97.8 | 83.8 | 97.3 | 90.0 | 46.5 | 42.6 | 44.4 | 9.1 | 22.9 | 13.0 | |
| suppobox | 99.6 | 88.8 | 93.9 | 98.6 | 96.5 | 97.6 | 98.5 | 39.4 | 56.3 | 86.8 | 96.2 | 91.3 | |
| darkwatchman | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 89.3 | 99.9 | 94.3 | 83.7 | 99.8 | 91.1 | |
| dyre | 99.0 | 100.0 | 99.9 | 99.8 | 100.0 | 99.9 | 97.8 | 99.8 | 98.8 | 99.0 | 88.9 | 93.7 | |
| grandoreiro | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.9 | 100.0 | 99.9 | 100.0 | 99.9 | 99.9 | |
| monerominer | 100.0 | 100.0 | 100.0 | 99.8 | 100.0 | 99.9 | 99.2 | 99.9 | 99.6 | 83.2 | 99.6 | 90.7 | Hash |
| pandabanker | 100.0 | 100.0 | 100.0 | 100.0 | 99.8 | 99.9 | 98.8 | 97.6 | 98.2 | 97.8 | 60.7 | 74.9 | |
| tinynuke | 100.0 | 100.0 | 100.0 | 99.8 | 99.3 | 99.6 | 96.0 | 67.7 | 79.4 | 97.1 | 83.9 | 90.0 | |
| wd | 100.0 | 100.0 | 100.0 | 99.3 | 99.9 | 99.6 | 97.9 | 99.6 | 98.7 | 95.7 | 99.0 | 97.3 | |
| **Total** | 96.2 | 91.4 | 93.7 | 88.3 | 72.5 | 79.6 | 62.7 | 51.9 | 56.8 | 46.7 | 37.2 | 41.4 | — |

# Real-World Deployment Challenges

**Performance Degradation**

- When tested on real university DNS logs ($D_3$), accuracy dropped by **11% to 24%** compared to the baseline.
- LLMs struggle with legitimate domains that *look* algorithmic (e.g. cloud storage subdomains).

**Operational Limitations**

- **Throughput:** Too slow for real-time DNS filtering.
  - Fastest (Gemini Flash-8B): 12.6 domains/sec.
  - Slowest (Gemini Pro): 2.5 domains/sec.
  - Requirement: DNS firewalls need thousands/sec.
- **Cost:** High accuracy (Claude Sonnet) is expensive (≈$14 per 50k domains).

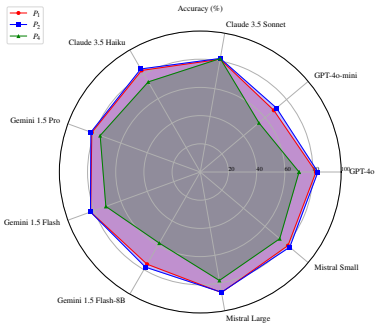



Figure: P1, P2, and P4 benign accuracy

# Concluding remarks

- LLMs *can* detect AGDs with relative high accuracy using only pre-trained knowledge.
- Significant limitations in distinguishing benign AGD-like domains (high False Positives).
- **Conclusion:** Not ready for real-time inline blocking, but promising for **offline forensics**.

# Thank You

**Questions?**

tomaspelayobenedet.github.io