INTRODUCTION

CAPTCHAs were originally introduced as a security mechanism to distinguish humans from machines . Early text-based CAPTCHAs exploited the limits of OCR, but advances in computer vision models and in the advent of modern LLM models that has prooved its capability in many fields also it is used to break modern captcha models.As its emergent capabilities improve along with improvement in computer vision and deep learning models,a combination of these prove deathly to modern captchas

MOTIVATION

As preiously mentioned , how important it is to prevent free email services, online polls, worms, and spam, preventing dictionary attacks and cyber-attacks it is equally important fo a modern world to have a strong captcha .We can observe the rising cases of deep fake image and video generation and many such other problems.It is very important that we are able to distinguish between machines and humans especially when the modern algorithms have improved far and beyond the turing test .

LITERATURE REVIEW

NewCognitive Deep-Learning CAPTCHA

In this paper,the authors have made zxCAPTCHA, mixing image-based, text-based, and cognitive-based CAPTCHA schemes’ characteristics and applying adversar ial examples, neural style transfer, and some defense deep-learning techniques, to improve the security of the CATCHA

First, we show that the combination between deep learning and cognition can signifi cantly improve the security of image-based and text-based CAPTCHAs. • Second,wesuggestapromising direction for designing CAPTCHAs. The proposed CAPTCHAcanbevariedfordifferent cognitive CAPTCHA schemes by changing their attributes. Specifically, the attribute of text group order can be natural order, inverse order, or special order, such that any text group with special characteristics can be requested to be picked up with higher priority. Furthermore, the background images and text can be localized to make them more familiar to users in their local surround ings. As a result, we can see that the use of this CAPTCHA is widespread and simple to adapt to any system that requires CAPTCHA protection against automated bots.

Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA

The framework opera tionalizes the model's reasoning into executable actions. It achieves state-of-the-art performance on five high-difficulty CAPTCHAtypesandattains an average accuracy of 83.9% across all seven categories in our benchmark, substantially surpassing existing baselines

Table 1. CAPTCHA Benchmark Comparisons.

Benchmark	Real world	Reasoning	Region Consistent	Scale
Open CaptchaWorld [14]	✗	✗	✗	225
Halligan [21]	✓	✗	✓	2600
OEDIPUS [5]	✓	✗	✗	300
MCA-Bench [24]	✗	✗	✗	180000
CAPTCHA-X (Ours)	✓	✓	✓	1839

In this paper, we create the first real-world benchmark CAPTCHA-X with reasoning annotations and show evi dence that reasoning is the key to solving CAPTCHAs. Di rectly applying commercial VLMs to solve CAPTCHAs, especially highly difficult tasks, achieves only an accuracy of 15.7%. underscoring severe deficits in spatial reasoning. Oncereasoning is introduced, however, performance sta tistically significantly improves by an average of 38.75% relative to the non-reasoning baseline.

Experiments on our benchmark show that incorporating reasoning improves performance by 38.75% relative to the non-reasoning baseline, and statistical analysis con f irms the improvement is highly significant ($p < 0.001$), providing the systematic evidence that reasoning funda mentally improves model accuracy.

we develop a reasoning-centered agent framework that grounds rea soning into executable intelligence, reaching 83.9% aver age accuracy across seven CAPTCHA types and setting new state-of-the-art performance on five categories.

benchmarks used

CAPTCHA Evolution and Benchmarking
Reasoning in Visual CAPTCHA Solving
Spatial Reasoning Benchmarks. .

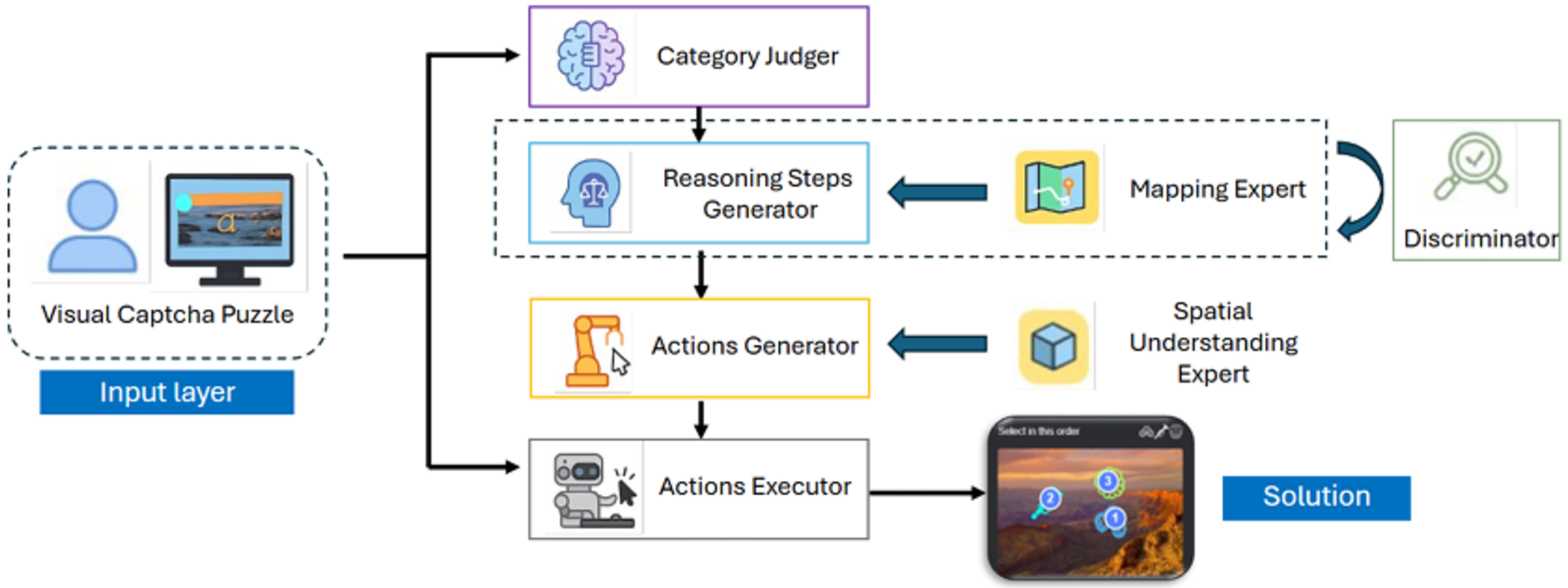


Figure 3. Our Agentic Vision-Language Model Pipeline.

Action Accuracy

$$AccRate = \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}\Big(a_i^{(j)} = a_i^{*(j)} \; \wedge \\ (\hat{x}_i^{(j)}, \hat{y}_i^{(j)}) \in \mathcal{RG}_i^{(j)}, \; \forall i\Big)$$

Reasoning Accuracy

1. Reasoning Steps.
2. Reasoning Length
3. Reasoning Efficiency.

$$Efficiency_i = \frac{Acc_i}{\alpha \cdot \hat{L}_i + \beta \cdot \hat{S}_i}.$$

Trajectory Complexity Index (TCI)

$$z_{i,j} = \frac{\sum_F (F_{i,j} - \overline{F}_i)}{0.5 \cdot (s_i/\overline{s}) + 0.5 \cdot (t_i/\overline{t})}.$$

$$TCI_i = \sigma\left(\frac{1}{N_i}\sum_{j=1}^{N_i} z_{i,j}\right), \quad \sigma(x) = \frac{1}{1+e^{-x}}.$$

5.Reasoning Score.

$$S_i = \frac{1}{4} \sum_{m=1}^{4} s_{i,m}.$$

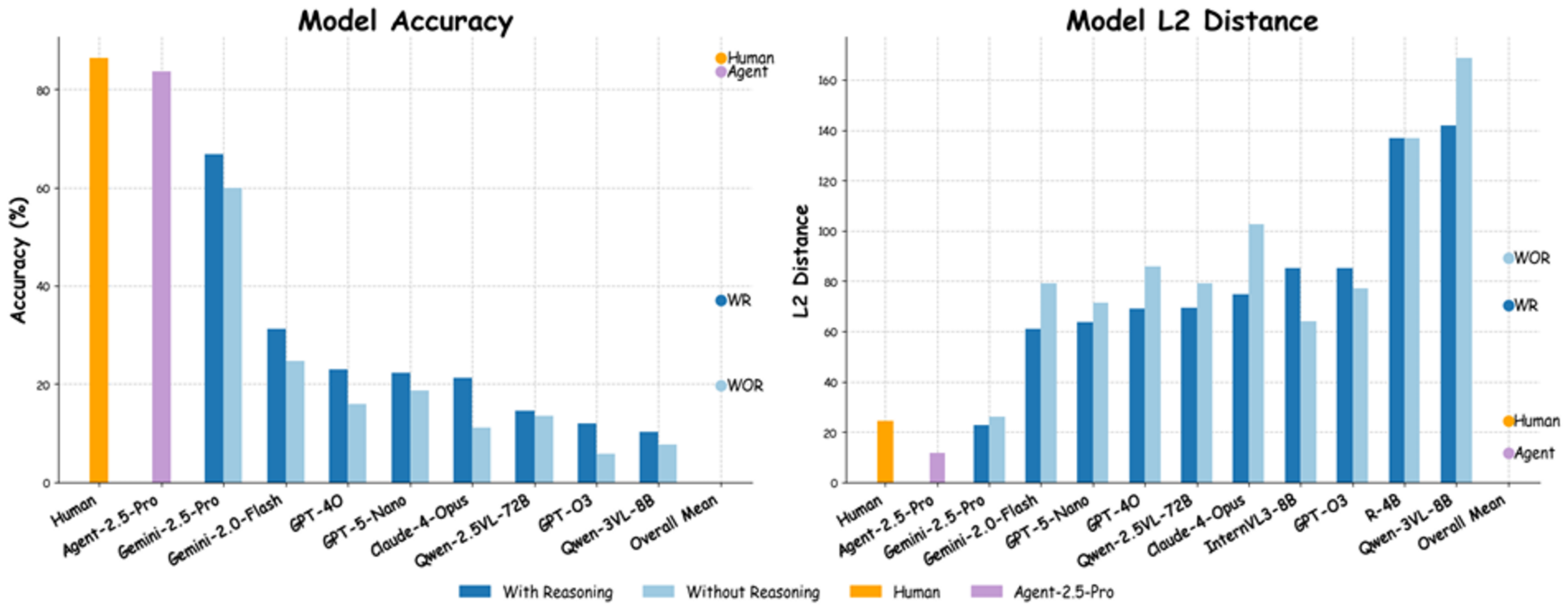
Experiments they conducted

Action Evaluation

Evaluation of Prediction Accuracy

Evaluation of L2 Distance.

Statistical Validation.



Model Accuracy
Accuracy (%)
Human
Agent
WR
WOR
Human
Agent-2.5-Pro
Gemini-2.5-Pro
Gemini-2.0-Flash
GPT-4O
GPT-5-Nano
Claude-4-Opus
Qwen-2.5VL-72B
GPT-O3
Qwen-3VL-8B
Overall Mean
Model L2 Distance
L2 Distance
WOR
WR
Human
Agent
Human
Agent-2.5-Pro
Gemini-2.5-Pro
Gemini-2.0-Flash
GPT-5-Nano
GPT-4O
Qwen-2.5VL-72B
Claude-4-Opus
InternVL3-8B
GPT-O3
R-4B
Qwen-3VL-8B
Overall Mean
With Reasoning
Without Reasoning
Human
Agent-2.5-Pro

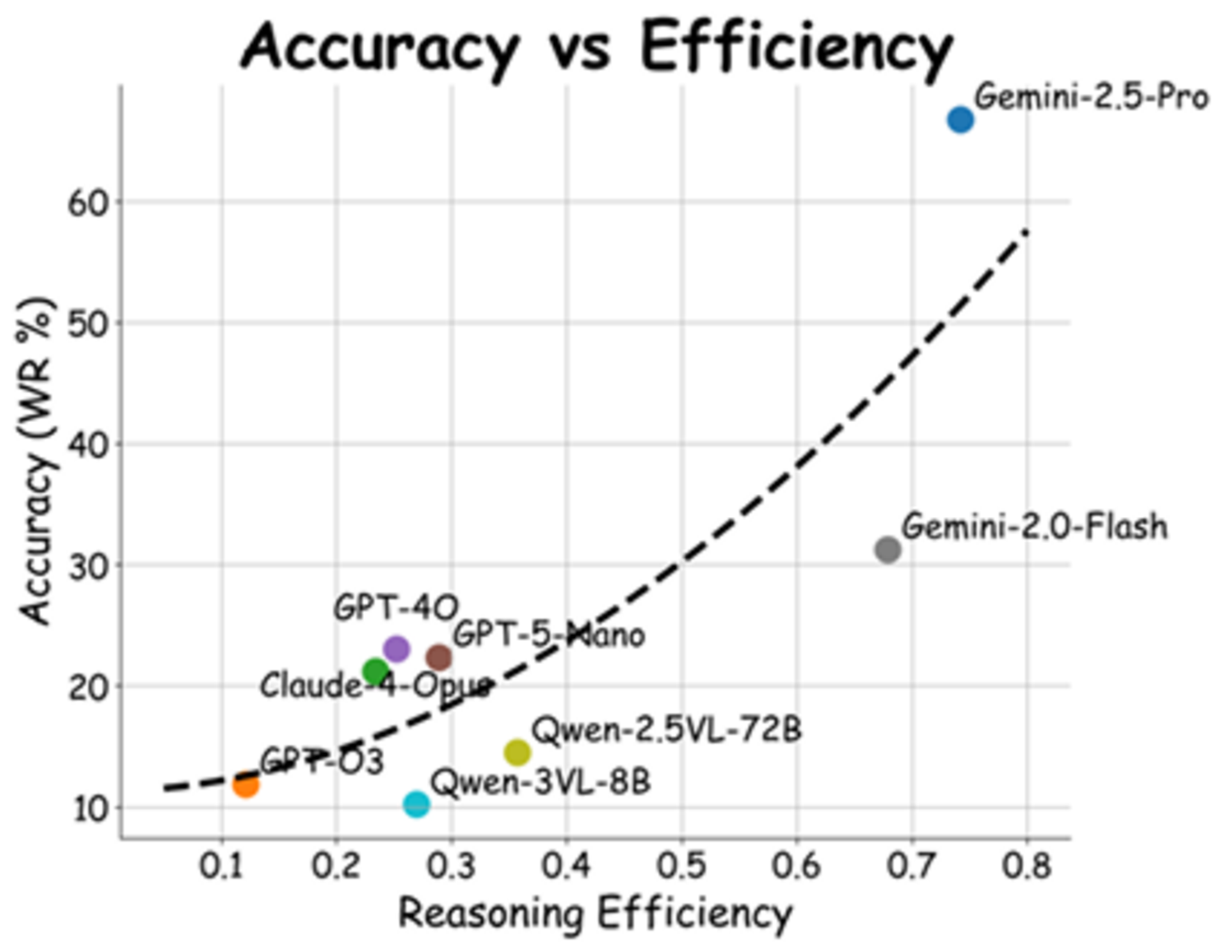


Figure 6. Reasoning Scaling Law.

RESEARCH GAPS

OBJECTIVES

PROPOSED WORK

CONCLUSION

REFERENCES

[1] N. D. Trong, T. H. Huong, and V. T. Hoang, "New Cognitive Deep-Learning CAPTCHA," Sensors, vol. 23, no. 4, Art. no. 2338, Feb. 2023, doi: 10.3390/s23042338.

[2] C. Iliou, T. Kostoulas, T. Tsikrika, V. Katos, S. Vrochidis, and I. Kompatsiaris, "Detection of Advanced Web Bots by Combining Web Logs with Mouse Behavioural Biometrics," Digit. Threats: Res. Pract., vol. 2, no. 3, Art. no. 24, pp. 1–26, Sep. 2021, doi: 10.1145/3447815.

[3] A. A. Nia, "A Hybrid CAPTCHA Combining Generative AI with Keystroke Dynamics for Enhanced Bot Detection," Sep. 2025, arXiv:2510.02374.

[4] Z. Ding, G. Deng, Y. Liu, J. Ding, J. Chen, and Y. Sui, "Illusion CAPTCHA: A CAPTCHA based on Visual Illusion," in Proc. ACM Web Conf. 2025 (WWW '25), Sydney, NSW, Australia, Apr. 2025, doi: 10.1145/3696410.3714726.

[5] J. Wang, L. Li, C. Zhu, and X. He, "COGNITION: From Evaluation to Defense against Multimodal LLM CAPTCHA Solvers," Dec. 2025, arXiv:2512.02318.

[6] P. Song, L. T. Chang, Y. Tsai, P. Li, and J. Yang, "Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA," Nov. 2025, arXiv:2510.06067.