Small Language Reasoning Models

Pablo Leyva, Xiaoning Ding

pl33@njit.edu, xiaoning.ding@njit.edu

November 25, 2025

Abstract

A rise of Small Language Models (SLMs) has been seen in the last few years; however, these models are being used as a general purpose models for a wide range of tasks such as text generation, translation, and summarization. These models became more popular due to usability and accessibility in limited hardware resource use cases. We propose a new class of foundational models that are trained to reason and problem solve one step at a time. Small Language Reasoning Models (SLRMs) are not meant for general chatbot applications but rather for internal reasoning and planning in agentic workflows.

1 Introduction

2 Objectives

3 Significance

3.1 Content

3.2 Innovations

4 Methodology

5 Expected Results

5.1 Outcomes

5.2 Challenges

5.3 Mitigation

6 Timeline

References

[1] Peter Belcak, Greg Heinrich, Shizhe Diao, Yonggan Fu, Xin Dong, Saurav Muralidharan, Yingyan Celine Lin, Pavlo Molchanov.
Small Language Models are the Future of Agentic AI.
arXiv:2506.02153, 2024.

[2] Elliot Meyerson, Giuseppe Paolo, Roberto Dailey, Hormoz Shahrzad, Olivier Francon, Conor F. Hayes, Xin Qiu, Babak Hodjat, Risto Miikkulainen.
Solving a Million-Step LLM Task with Zero Errors.
arXiv:2511.09030, 2025.

[3] Alexia Jolicoeur-Martineau.
Less is More: Recursive Reasoning with Tiny Networks.
arXiv:2510.04871, 2025.

[4] Nexusflow, Venkat Krishna Srinivasan, Zhen Dong, Banghua Zhu, Brian Yu, Damon Mosk-Aoyama, Kurt Keutzer, Jiantao Jiao, Jian Zhang.
NexusRaven: A Commercially-Permissive Language Model for Function Calling.
openreview.net, 2024.

[5] Shishir G. Patil, Tianjun Zhang, Xin Wang, Joseph E. Gonzalez.
Gorilla: Large Language Model Connected with Massive APIs.
arXiv:2305.15334, 2023.