

# Zero-Shot Experiment Report

Automatically generated on April 05, 2026 at 14:28

## Benchmark-Wide Zero-Shot Pass@1

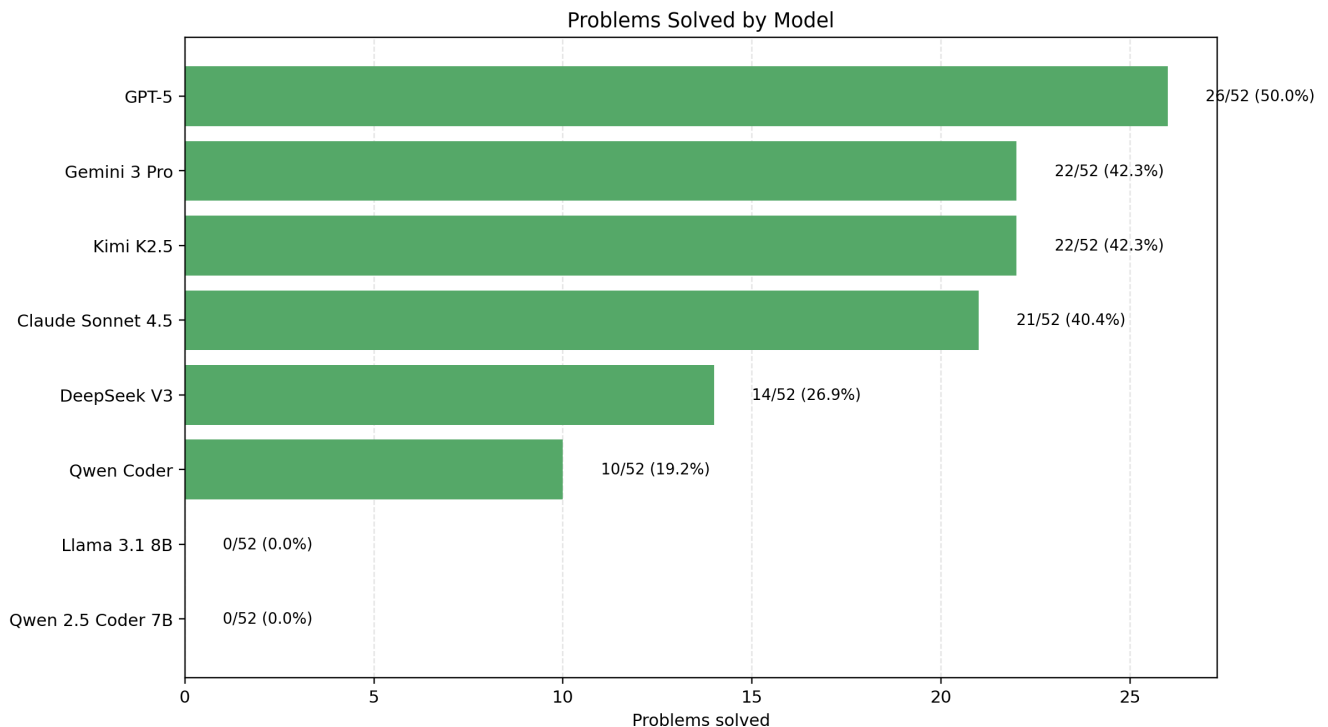
strategy	problems	models	pass@1 (%)
zero_shot	52	8	25.0

## Analysis of Models

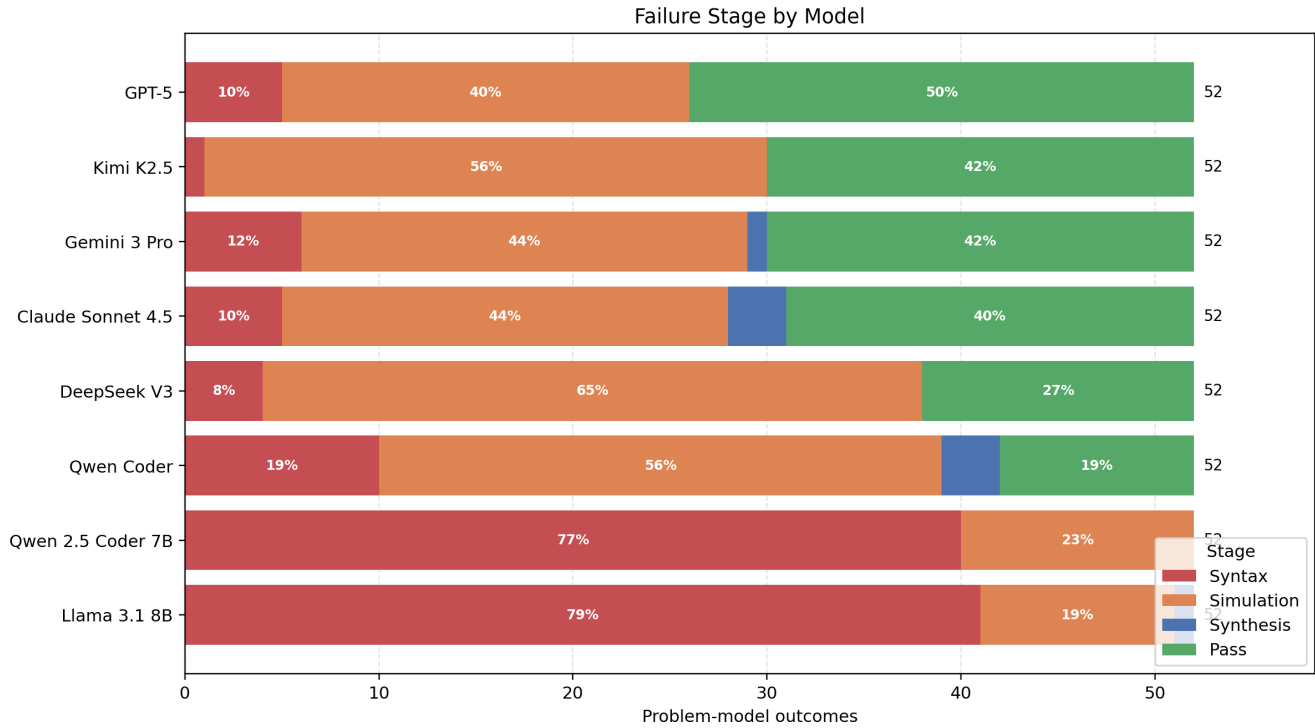
### Zero-Shot Experiment Summary

Model	Pass@1 (%)	Pass@3 (%)
GPT-5	50.0	50.0
Gemini 3 Pro	42.3	42.3
Kimi K2.5	34.6	42.3
Claude Sonnet 4.5	40.4	40.4
DeepSeek V3	17.3	26.9
Qwen Coder	15.4	19.2
Llama 3.1 8B	0.0	0.0
Qwen 2.5 Coder 7B	0.0	0.0
<i>CodeV-R1-RL-Qwen-7B (Baseline)</i>	<i>7.7</i>	<i>7.7</i>

### Problems Solved by Model

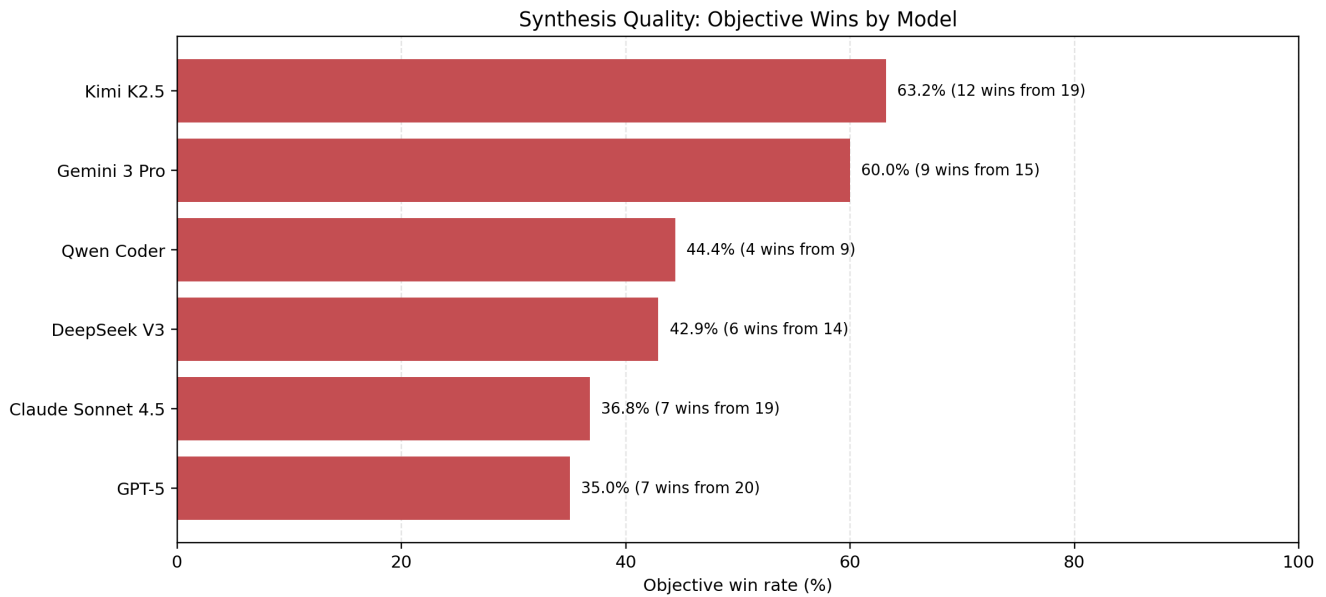


## Failure Stage by Model



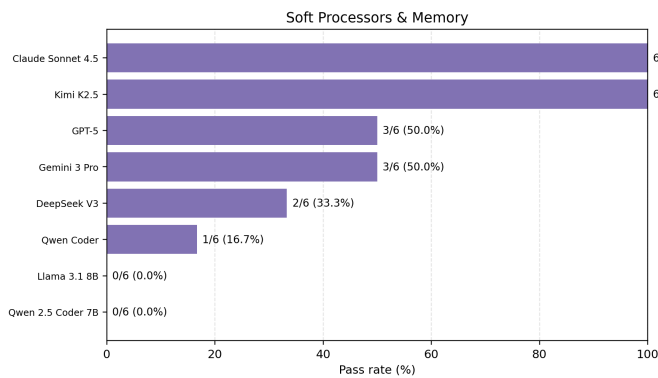
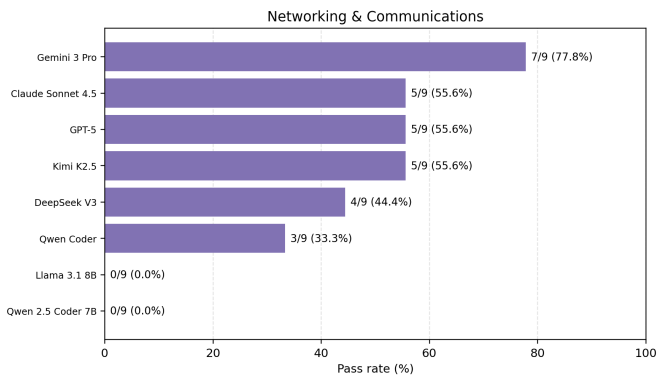
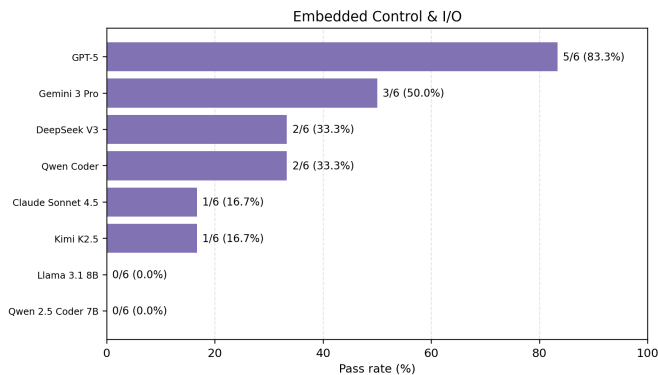
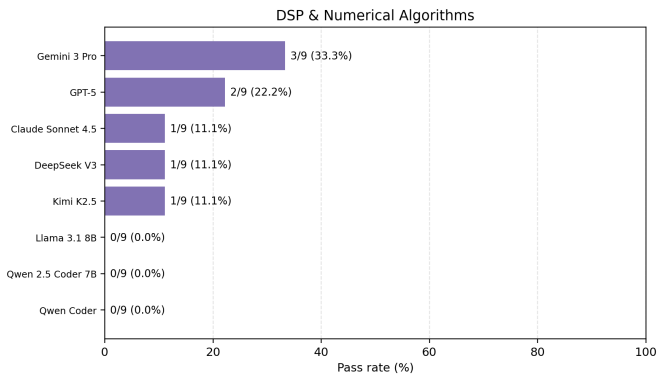
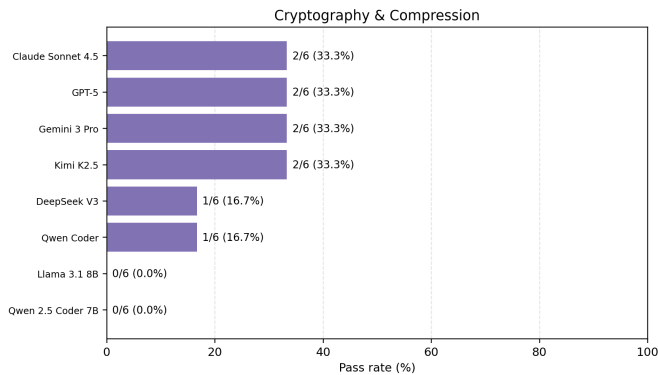
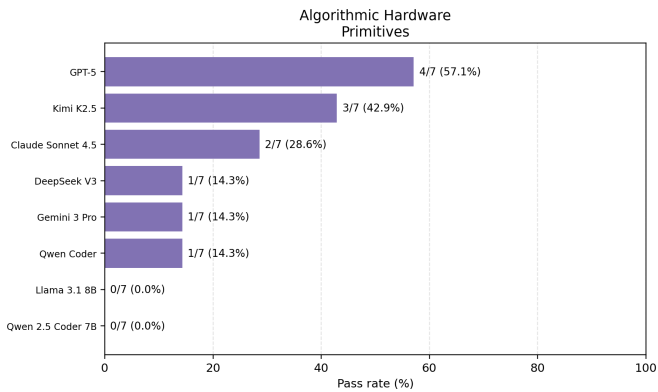
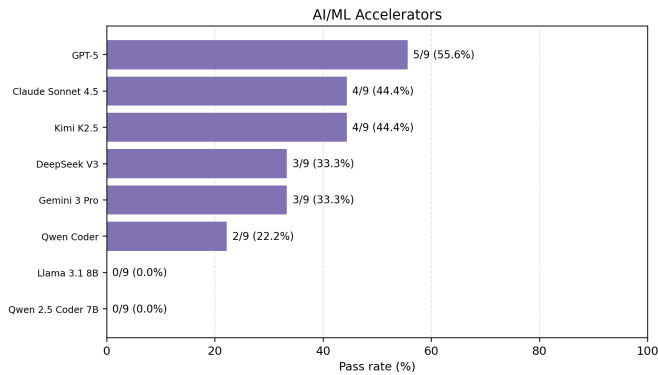
This figure considers only problems where the optimization objective is supported and at least two models produced a synthesis-passing solution. Each such problem awards one objective win to the model, or tied models, with the best objective value among the successful solutions.

## Synthesis Quality: Objective Wins by Model



# Model Ranking by Category

Model Ranking by Category

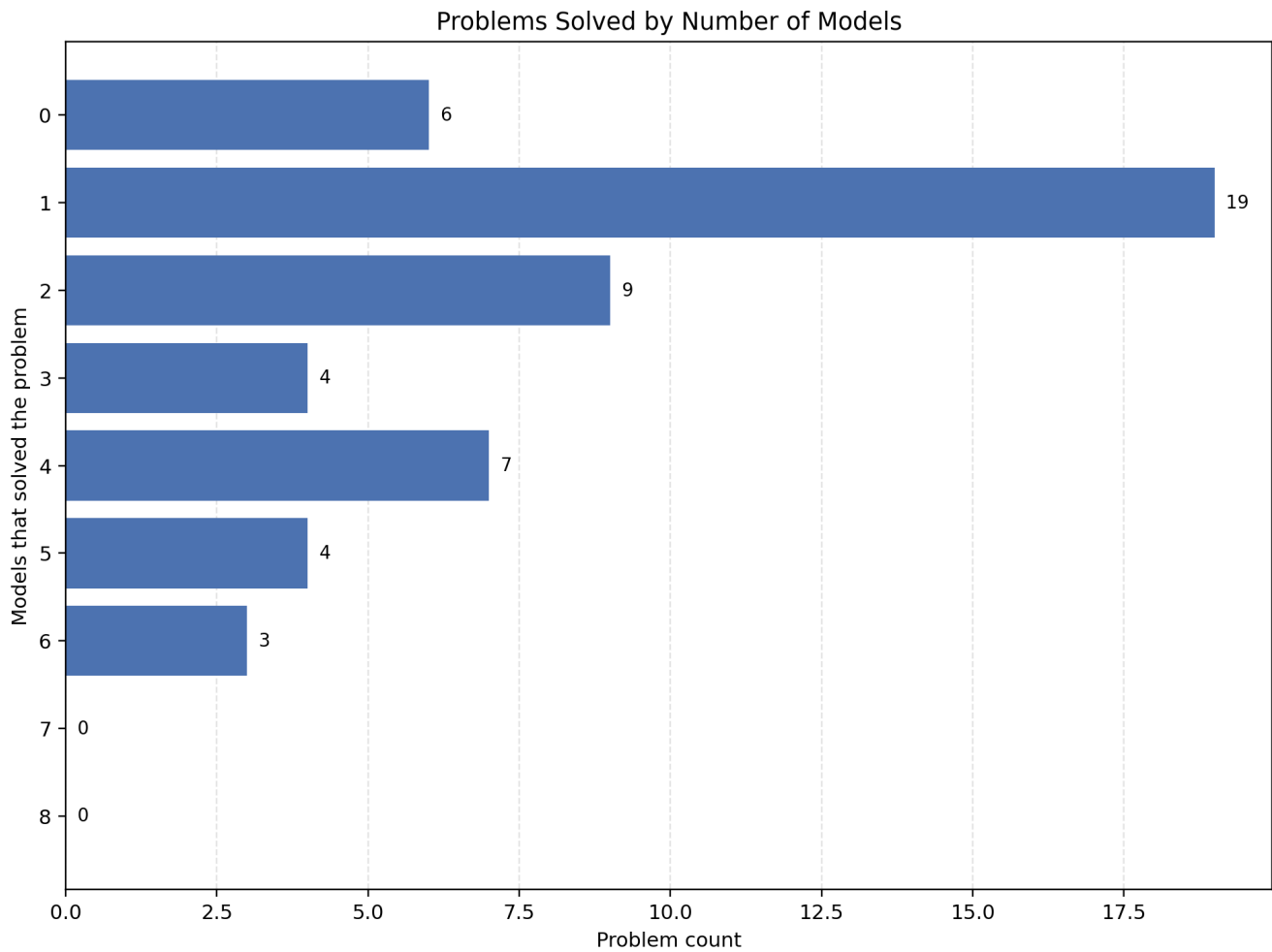


These figures compare overall zero-shot pass rate against two resource proxies. The cost figure uses the pricing proxy of input cost plus output cost per one million tokens. The model-size figure uses a maintained parameter-count lookup where a defensible size is available, so models with undisclosed parameter counts may be omitted from that plot.



# Analysis of Problems

## Problems Solved by Number of Models



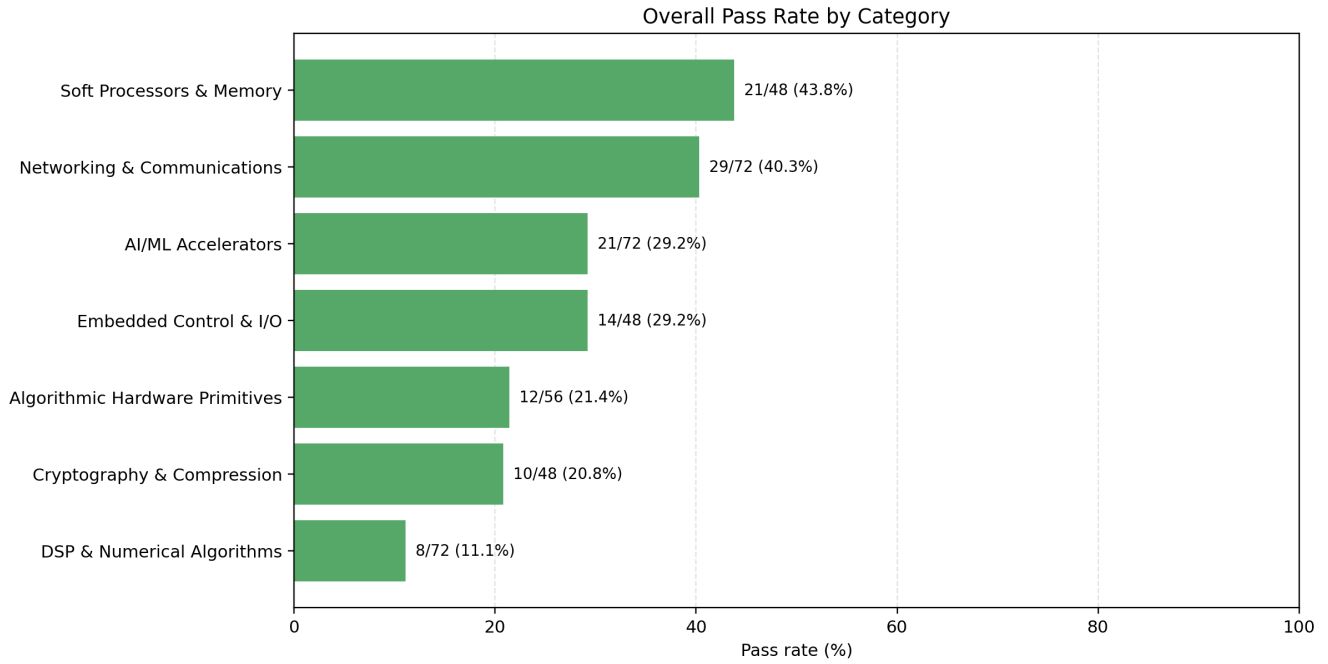
## Top 5 Easiest Problems

id	title	solved	solve rate
13	Hardmax One-Hot Classifier	6 / 8	75.0
15	Longest Prefix Match Router	6 / 8	75.0
16	Priority-Based Packet Filter	6 / 8	75.0
9	Dual-Port Register File	5 / 8	62.5
22	Fixed-Weight Perceptron	5 / 8	62.5

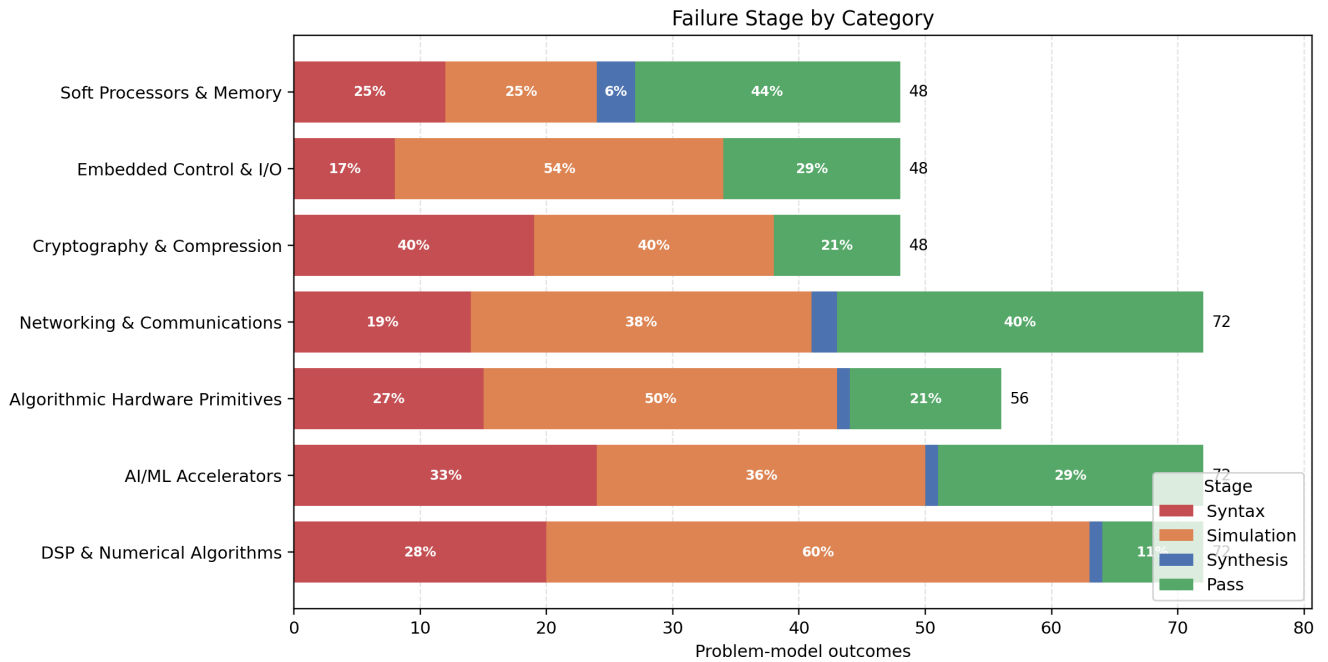
## Top 5 Hardest Problems

id	title	solved	solve rate
4	Streaming 3x3 CNN Convolution Layer	0 / 8	0.0
38	7-Sample Median Denoising Filter	0 / 8	0.0
43	Parallel Bitonic Sort and Rank Engine	0 / 8	0.0
48	2x2 Streaming MAC Tile	0 / 8	0.0
49	Variable-Width Timestamp Delta Encoder	0 / 8	0.0

## Overall Pass Rate by Category



## Failure Stage by Category



# Overall Pass Rate by Complexity

