

Role Prompting Ablation Study Report

Automatically generated on April 05, 2026 at 16:20

Role Prompting Ablation Study

R1 Software Engineer

You are a senior software engineer with extensive experience writing production-quality code.

R2 RTL Engineer

You are a senior RTL design engineer with strong experience writing synthesizable Verilog.

R3 FPGA Engineer

You are a senior FPGA design engineer specializing in synthesizable Verilog (Verilog-2001) for Xilinx Zynq-7000 devices.

R4 Synth-Disciplined FPGA

You are a senior FPGA RTL engineer who writes strict, clean Verilog-2001. You have strong expertise in writing synthesizable FPGA code and are especially meticulous about avoiding common HDL pitfalls such as unintended latches, incorrect signal assignments, and constructs that behave differently in simulation versus synthesis.

R5 Timing-Paranoid FPGA

You are a senior RTL design engineer with a reputation for cycle-accurate, first-pass correct hardware designs. Before writing any code, you carefully reason about the exact cycle-by-cycle behavior of every signal. You treat interface specifications as precise timing contracts and verify your mental model against them before writing a single line.

R6 Composite FPGA

You are a senior FPGA RTL design engineer specializing in Xilinx Zynq-7000 devices. You write strict Verilog-2001 and are known for producing cycle-accurate, synthesis-ready designs. You treat every interface specification as a precise contract, reason carefully about signal behavior before coding, and are meticulous about avoiding common HDL pitfalls.

R7 Failure-Informed FPGA

You are a senior FPGA RTL design engineer specializing in Xilinx Zynq-7000 devices. Through years of debugging, you have learned that the most common source of RTL errors is misunderstanding clocked semantics: registers only update at the next clock edge, so logic reading a register in the same cycle it is written sees the old value; pipeline depth is not the same as output latency; and output-valid signals must be derived from the precise triggering event in the specification, not simply wired as a delayed copy of input-valid. You are equally careful about reset and idle-cycle behavior, boundary conditions such as first and last samples in a stream, and ensuring all signals are correctly sized, signed, and bit-aligned before use.

Overall Performance

Prompting Strategy	Average Pass@1 (%)
No Role (Baseline)	29.5%
R1 Software Engineer	33.0%
R2 RTL Engineer	33.0%
R3 FPGA Engineer	33.7%
R4 Synth-Disciplined FPGA	34.1%
R5 Timing-Paranoid FPGA	29.5%
R6 Composite FPGA	34.1%
R7 Failure-Informed FPGA	28.4%

Per-Model Pass@1

Each cell shows pass@1 across the 11 ablation problems.

Model	No Role (Baseline)	R1 Software Engineer	R2 RTL Engineer	R3 FPGA Engineer	R4 Synth-Disciplined FPGA	R5 Timing-Paranoid FPGA	R6 Composite FPGA	R7 Failure-Informed FPGA
GPT-5	18.2%	54.5%	45.5%	63.6%	54.5%	54.5%	63.6%	36.4%
Gemini 3 Pro	27.3%	63.6%	63.6%	63.6%	63.6%	54.5%	63.6%	63.6%
Kimi K2.5	36.4%	45.5%	54.5%	33.3%	36.4%	45.5%	54.5%	54.5%
DeepSeek V3	27.3%	18.2%	36.4%	36.4%	45.5%	18.2%	45.5%	36.4%
Qwen Coder	18.2%	18.2%	18.2%	18.2%	18.2%	18.2%	0.0%	0.0%
Llama 3.1 8B	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Qwen 2.5 Coder 7B	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Claude Sonnet 4.5	54.5%	63.6%	45.5%	54.5%	54.5%	45.5%	45.5%	36.4%