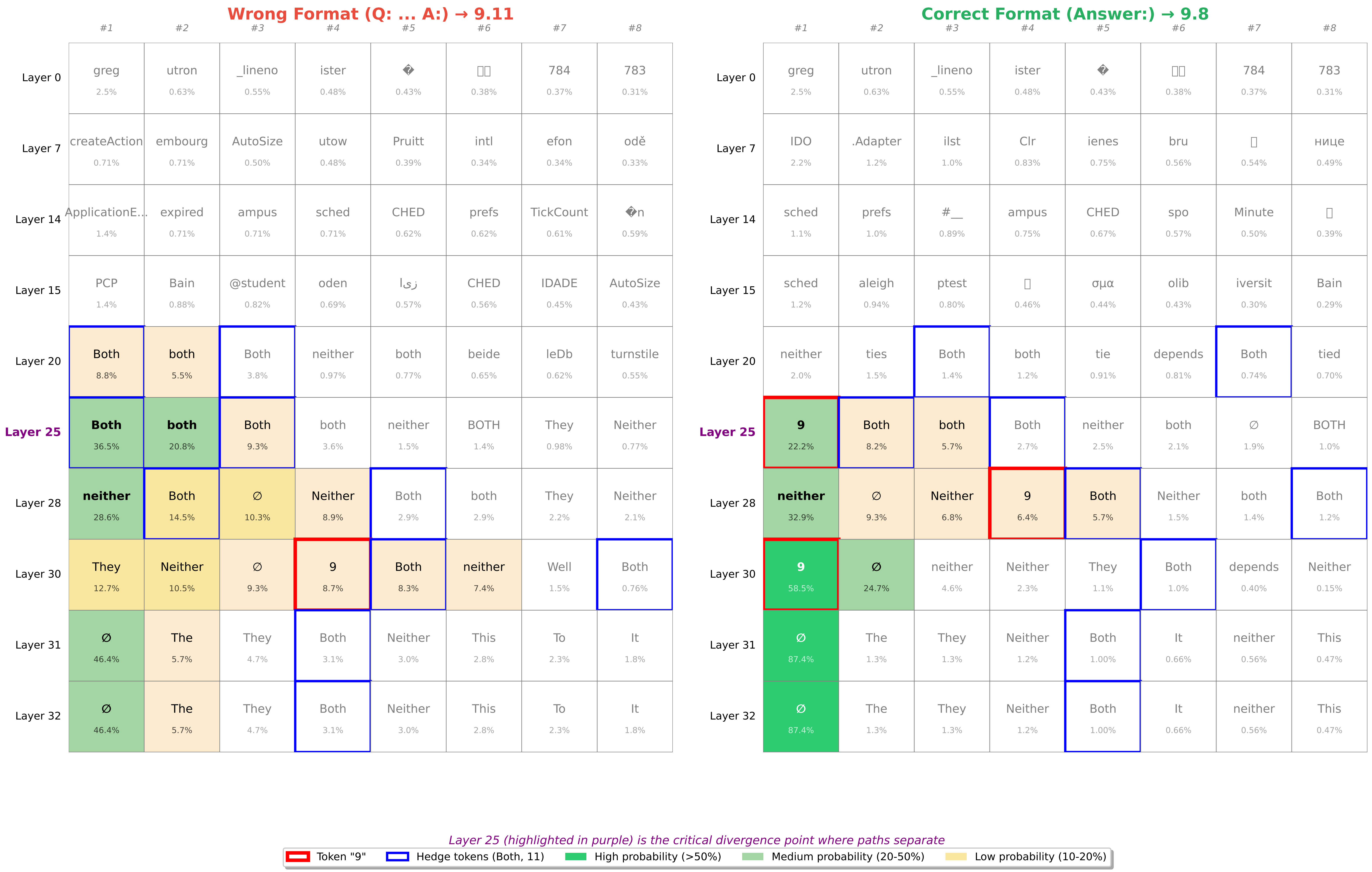
Logit Lens Analysis: 9.11 vs 9.8 Bug
Token Predictions Across Key Layers



Wrong Format (Q: ... A:) → 9.11

	#1	#2	#3	#4	#5	#6	#7	#8
Layer 0	greg 2.5%	utron 0.63%	_lineno 0.55%	ister 0.48%	� 0.43%	□□ 0.38%	784 0.37%	783 0.31%
Layer 7	createAction 0.71%	embourg 0.71%	AutoSize 0.50%	utow 0.48%	Pruitt 0.39%	intl 0.34%	efon 0.34%	odě 0.33%
Layer 14	ApplicationE... 1.4%	expired 0.71%	ampus 0.71%	sched 0.71%	CHED 0.62%	prefs 0.62%	TickCount 0.61%	�n 0.59%
Layer 15	PCP 1.4%	Bain 0.88%	@student 0.82%	oden 0.69%	ایز 0.57%	CHED 0.56%	IDADE 0.45%	AutoSize 0.43%
Layer 20	Both 8.8%	both 5.5%	Both 3.8%	neither 0.97%	both 0.77%	beide 0.65%	leDb 0.62%	turnstile 0.55%
Layer 25	**Both** 36.5%	**both** 20.8%	Both 9.3%	both 3.6%	neither 1.5%	BOTH 1.4%	They 0.98%	Neither 0.77%
Layer 28	**neither** 28.6%	Both 14.5%	∅ 10.3%	Neither 8.9%	Both 2.9%	both 2.9%	They 2.2%	Neither 2.1%
Layer 30	They 12.7%	Neither 10.5%	∅ 9.3%	9 8.7%	Both 8.3%	neither 7.4%	Well 1.5%	Both 0.76%
Layer 31	∅ 46.4%	The 5.7%	They 4.7%	Both 3.1%	Neither 3.0%	This 2.8%	To 2.3%	It 1.8%
Layer 32	∅ 46.4%	The 5.7%	They 4.7%	Both 3.1%	Neither 3.0%	This 2.8%	To 2.3%	It 1.8%

Correct Format (Answer:) → 9.8

	#1	#2	#3	#4	#5	#6	#7	#8
Layer 0	greg 2.5%	utron 0.63%	_lineno 0.55%	ister 0.48%	� 0.43%	□□ 0.38%	784 0.37%	783 0.31%
Layer 7	IDO 2.2%	.Adapter 1.2%	ilst 1.0%	Clr 0.83%	ienes 0.75%	bru 0.56%	□ 0.54%	нице 0.49%
Layer 14	sched 1.1%	prefs 1.0%	#__ 0.89%	ampus 0.75%	CHED 0.67%	spo 0.57%	Minute 0.50%	□ 0.39%
Layer 15	sched 1.2%	aleigh 0.94%	ptest 0.80%	□ 0.46%	σμα 0.44%	olib 0.43%	iversit 0.30%	Bain 0.29%
Layer 20	neither 2.0%	ties 1.5%	Both 1.4%	both 1.2%	tie 0.91%	depends 0.81%	Both 0.74%	tied 0.70%
Layer 25	**9** 22.2%	Both 8.2%	both 5.7%	Both 2.7%	neither 2.5%	both 2.1%	∅ 1.9%	BOTH 1.0%
Layer 28	**neither** 32.9%	∅ 9.3%	Neither 6.8%	9 6.4%	Both 5.7%	Neither 1.5%	both 1.4%	Both 1.2%
Layer 30	**9** 58.5%	∅ 24.7%	neither 4.6%	Neither 2.3%	They 1.1%	Both 1.0%	depends 0.40%	Neither 0.15%
Layer 31	∅ 87.4%	The 1.3%	They 1.3%	Neither 1.2%	Both 1.00%	It 0.66%	neither 0.56%	This 0.47%
Layer 32	∅ 87.4%	The 1.3%	They 1.3%	Neither 1.2%	Both 1.00%	It 0.66%	neither 0.56%	This 0.47%

Layer 25 (highlighted in purple) is the critical divergence point where paths separate

🟥 Token "9"
🟦 Hedge tokens (Both, 11)
🟩 High probability (>50%)
🟩 Medium probability (20-50%)
🟨 Low probability (10-20%)