



Common Reasoning Failures in Legal LLMs

These aren't hallucinations in the usual sense.

They're systematic reasoning errors that occur even when the model is trained on high-quality legal data.

Failure Mode	What It Is	Example / Signal	Mitigation Strategy
Overconfident Application of Precedent	The model applies a case without addressing key differences in fact pattern, procedural posture, or jurisdictional scope.	"This violates Plyler v. Doe" — but the facts involve adult immigrants in an economic benefit case.	Prompt for precedent comparison: <i>"Which facts differ? Would a court still apply Plyler here?"</i>
Doctrinal Overgeneralization	The model applies a legal test too broadly, missing exceptions or nuances in application across contexts.	Applies strict scrutiny to all fundamental rights cases, ignoring tiers in substantive due process.	Prompt for layered reasoning: <i>"Walk through each element of the test. Are there recognized exceptions?"</i>
Unexamined Ambiguity	The model presents unsettled or split legal questions as if they're resolved.	"This would be unconstitutional" — in an area where circuits are split or doctrine is evolving.	Ask for confidence ranges or alternate views: <i>"What would opposing counsel argue?"</i>
Jurisdictional Shortcutting	The model assumes a federal rule or doctrine applies across states without checking for divergence.	Cites federal free speech precedent in a California state constitutional analysis.	Prompt for jurisdictional awareness: <i>"Does California treat this right differently from the federal rule?"</i>
Factual Flattening	The model overlooks critical facts that affect legal outcome or test application.	Applies intermediate scrutiny without noting age, immigration status, or procedural context of the plaintiff.	Prompt for fact sensitivity: <i>"Which facts matter most to the level of scrutiny here?"</i>
Analogy Without Justification	The model compares to a past case without explaining why the analogy holds (or breaks).	"Like in Brown v. Board" — without showing how racial segregation is relevant in the new scenario.	Prompt for analogy breakdown: <i>"Which elements of Brown are truly analogous here?"</i>
Citation Anchoring Without Reasoning	The model gives a correct case name or test but skips the logical bridge to the conclusion.	"Chevron applies here" — but doesn't show why the agency interpretation would survive Step Two scrutiny.	Ask for test walkthrough: <i>"Apply each step of Chevron to this regulation. What's the weak link?"</i>