Table 1: Comparison of white-box unlearning analysis approaches. All columns are oriented so that ✓ is desirable. UDS is the only method achieving ✓ on all five axes. †See column definitions below.

Work	Score	Causal	Train-Free	Data-Inv	Meta-Eval
Lynch et al. (2024)	✗	✗	✗	△†	✗
Guo et al. (2025)	✗	△†	✗	✗	✗
Hong et al. (2025)	✗	△†	✓	✗	✗
Patil et al. (2024)	✗	✗	✓	✓	✗
Hong et al. (2024a)	✓	✓	✓	✗	✗
UDS (Ours)	✓	✓	✓	✓	✓

Score: proposes a named, bounded retention metric. **Causal**: uses controlled internal intervention (△† = causal methods in pipeline but final assessment is observational). **Train-Free**: no auxiliary model training required. **Data-Inv**: applicable to new forget sets without dataset-specific analysis. **Meta-Eval**: validated via faithfulness (AUC-ROC) and robustness (quantization/relearning stability).