Legal NLP Explainability Analysis
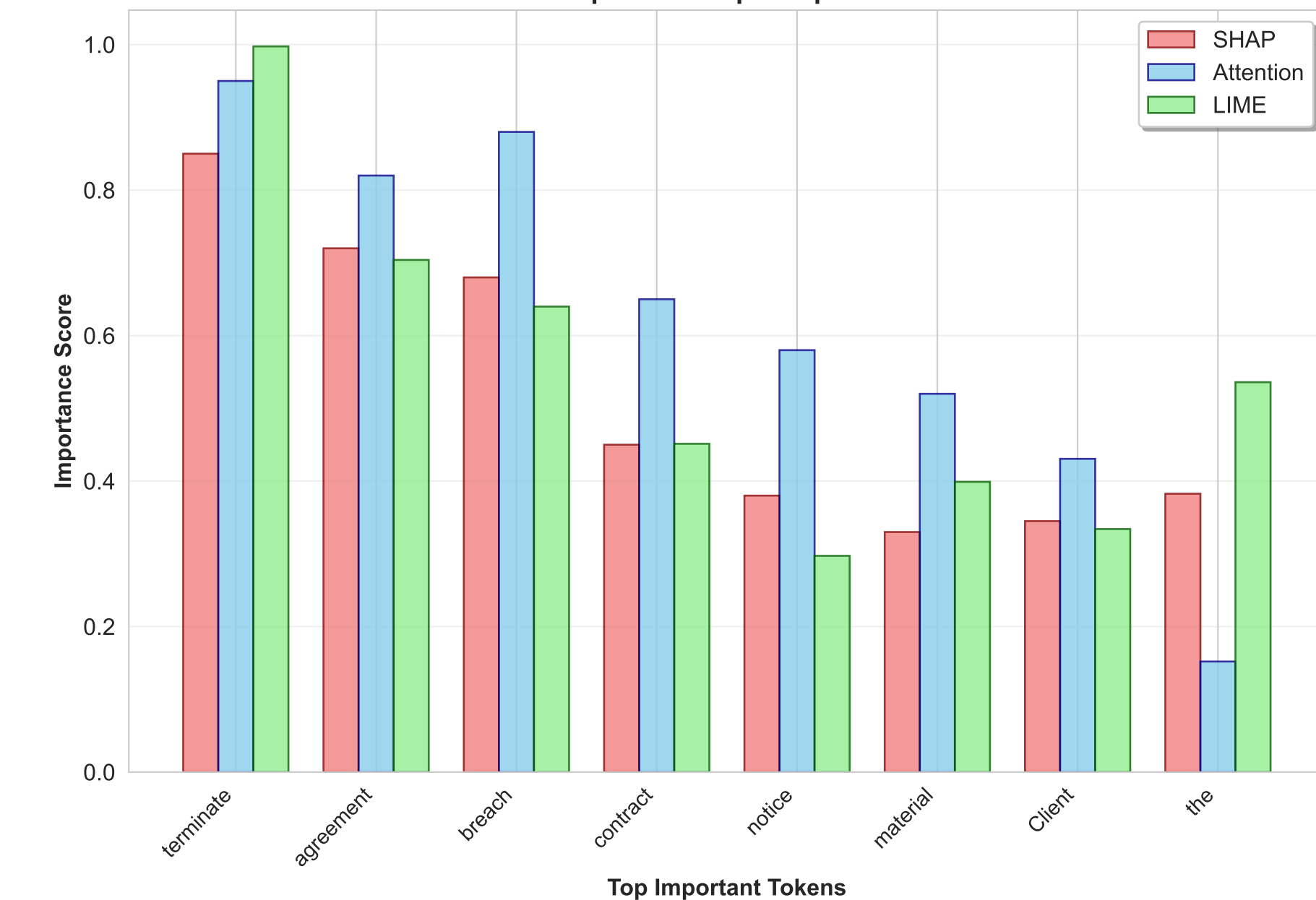
SHAP Feature Importance Analysis

Token	SHAP Value
conditions	-0.05
terms	0.29
contract	0.45
of	-0.18
breach	0.68
material	0.33
for	-0.11
Client	-0.34
the	-0.38
to	-0.03
notice	0.38
written	-0.09
days	0.11
(30)	-0.09
thirty	0.15
upon	0.32
agreement	0.72
this	-0.05
terminate	0.85
shall	0.13
Company	-0.03
The	-0.05

SHAP Value (Impact on Prediction)
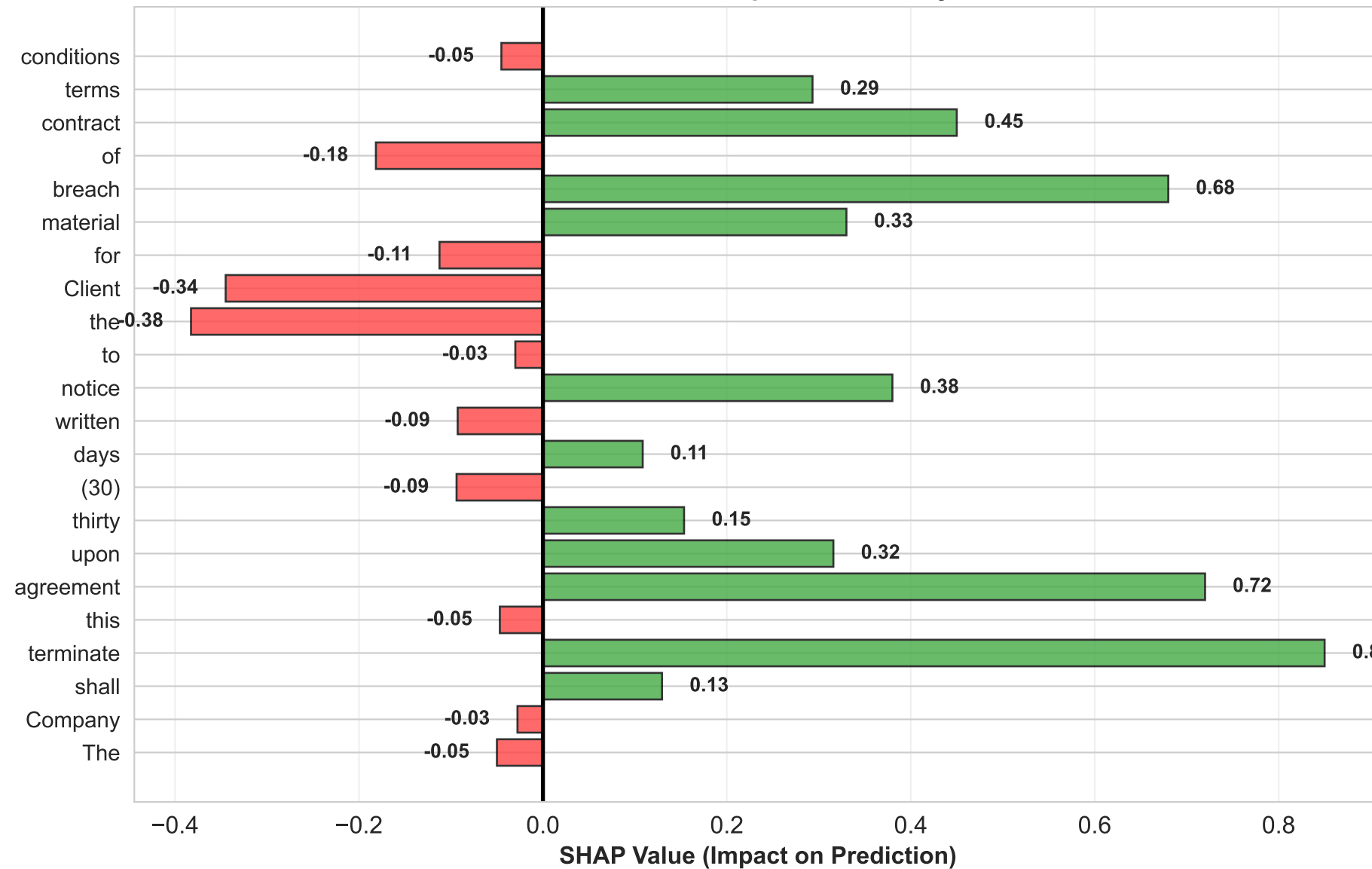
Multi-Head Attention Weights

Attention Head: Head 1, Head 2, Head 3, Head 4, Head 5, Head 6, Head 7, Head 8

Tokens: The, Company, shall, terminate, this, agreement, upon, thirty, (30), days, written, notice, to, the, Client, for, material, breach, of, contract, terms, conditions

Attention Weight
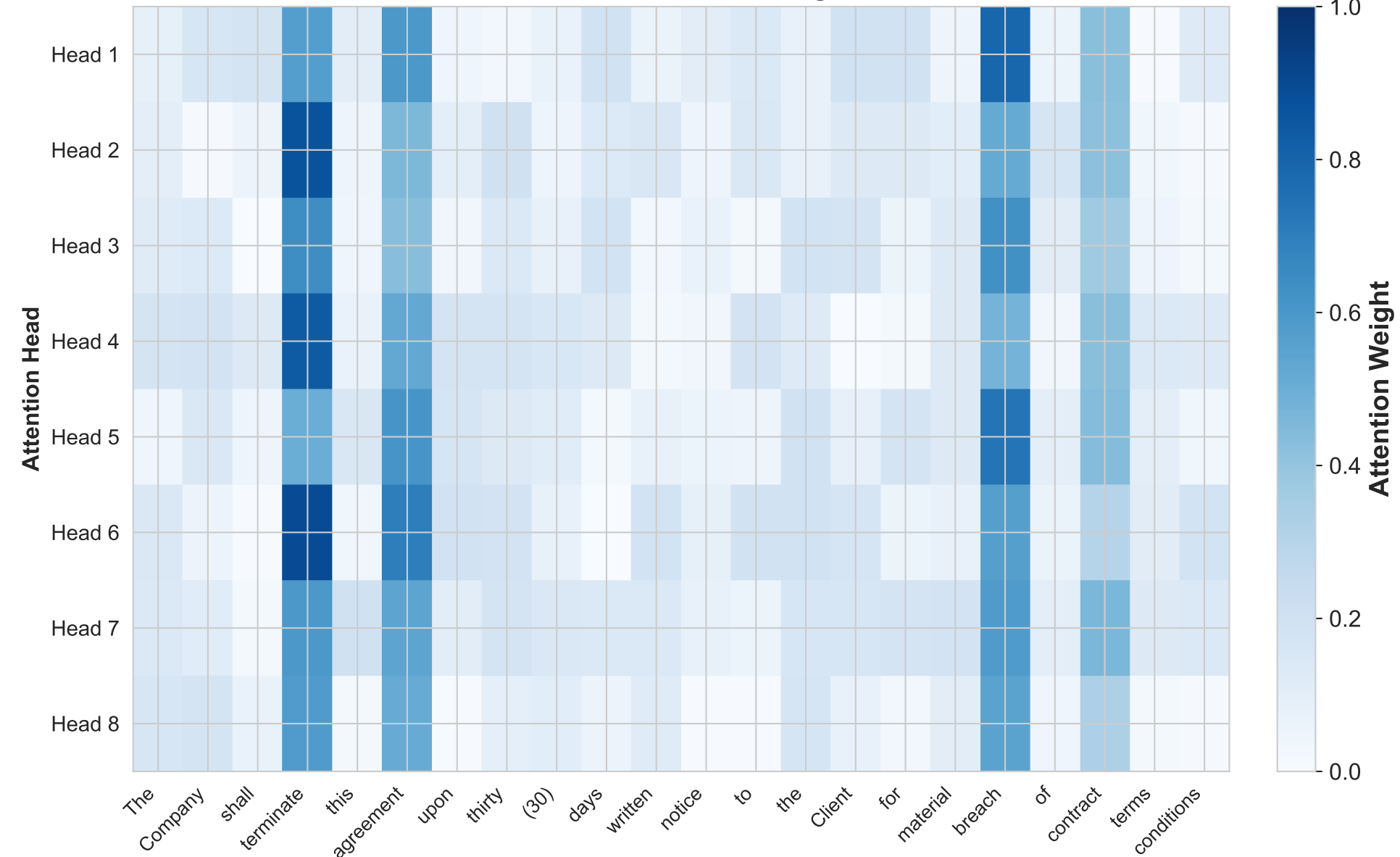
Method Comparison: Top 8 Important Features

Legend: SHAP, Attention, LIME

Top Important Tokens: terminate, agreement, breach, contract, notice, material, Client, the

Importance Score

Inter-Method Agreement Analysis

Comparison	Correlation Coefficient
SHAP vs Attention	0.764 (Strong)
SHAP vs LIME	0.971 (Strong)
Attention vs LIME	0.709 (Strong)