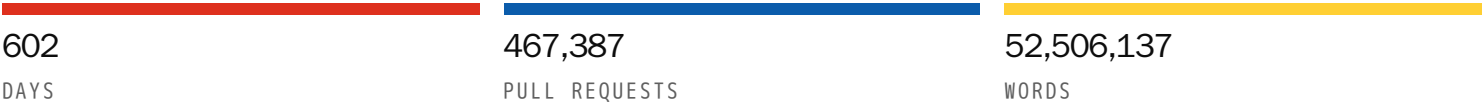


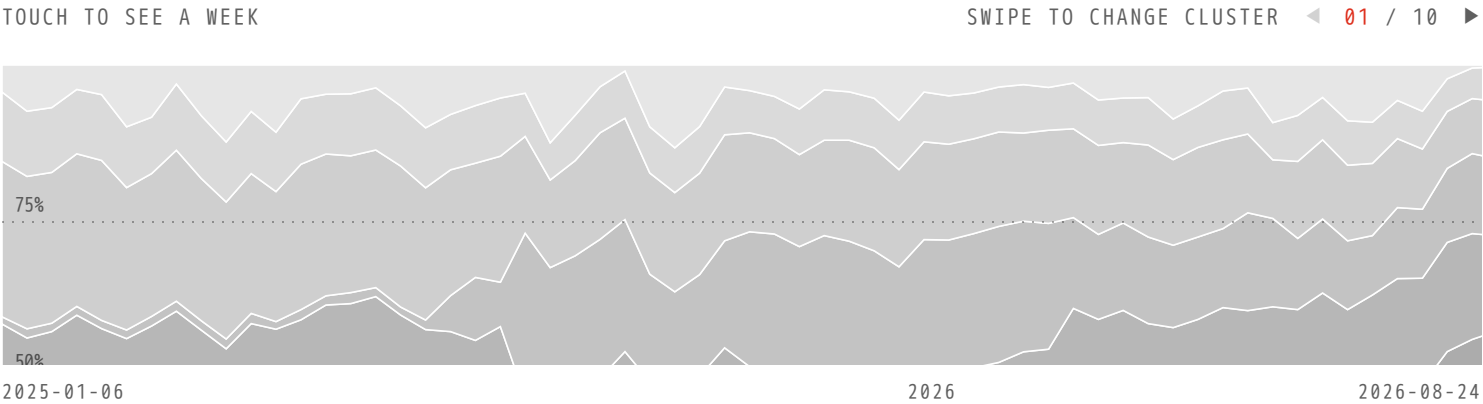
THE LOAD-BEARING VOCABULARY OF CLAUDE

We scrape 1,000 GitHub Pull Requests daily to analyse trends in vocabulary.
So far we have analysed:



We grouped GitHub PRs since 2025 into 10 clusters of vocabulary using KL-divergence k-means.
One of the clusters appeared in 2026 and represented 40% of all human-attributed pull requests last month.
Its most representative words should look familiar to anyone who uses coding agents.

Full code and methodology write-up.
Try our detector tool.



plainly

34.26× MORE FREQUENT IN THIS CLUSTER

TOUCH TO SEE A WEEK

54 per million, at the peak
7 per million words across the whole corpus



SEARCH FOR A WORD...

SCROLL · MOST REPRESENTATIVE FIRST



plainly

quietly

nobody

halves

genuinely

survived

