Performance vs token usage

Combined performance score (avg across datasets) vs Tokens used (avg across datasets)

- KC-agent
- KC-fast
- ToT
- KC-slow
- ReAct
- Reflexion
- Self-Discovery
- Plan-Execute
- Standard

Legend:
- High performance, low token usage
- High performance, high token usage
- Low performance, low token usage
- Low performance, high token usage