Waterfall Graph for Probe-15k

Multi-Step Data Filtering (y-axis)

- Initial Corpus: 151,000
- Length: 117,000
- Uniqueness: 95,000
- Technicality: 54,000
- Granularity: 15,500

Number of Questions (x-axis): 0K, 20K, 40K, 60K, 80K, 100K, 120K, 140K

Dataset Statistics:
Initial: 151,000
Final: 15,500
Filtered: 135,500
Data Reduction: 9.7x