Waterfall Graph for IntelliAlign-15k

Stage	Number of Questions
Initial Corpus	151,000
Length	117,000
Uniqueness	95,000
Technicality	54,000
Granularity	15,500

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

Number of Questions