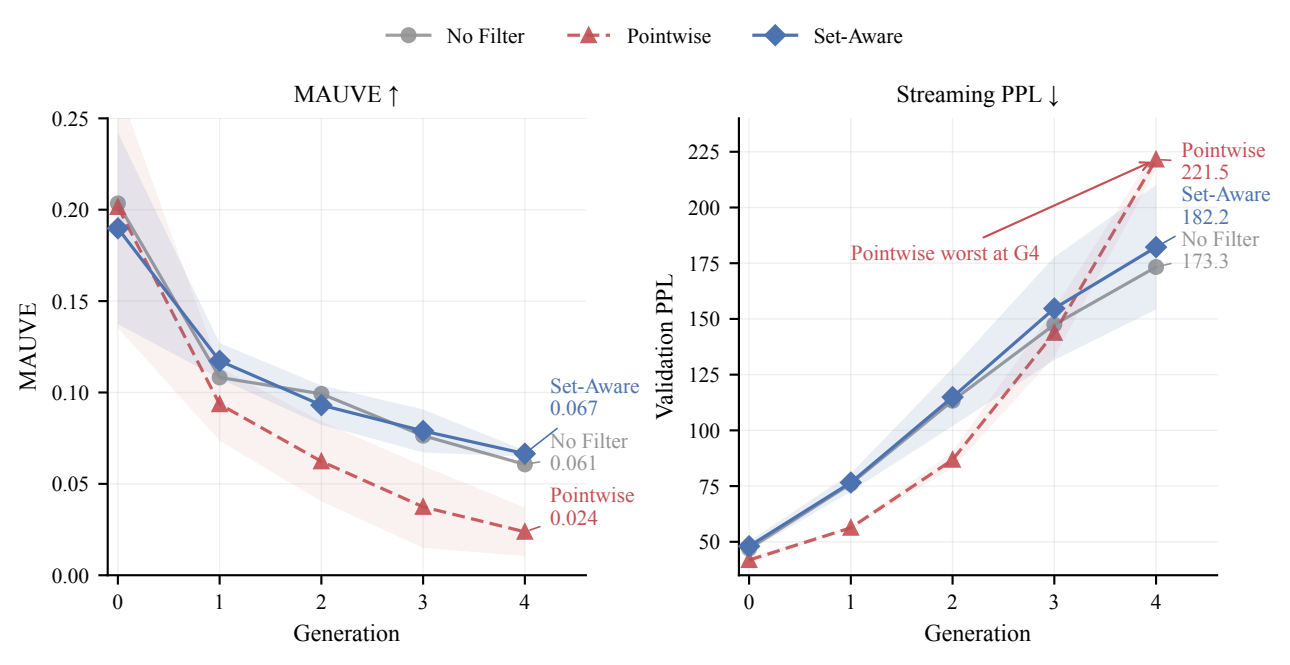
No Filter
Pointwise
Set-Aware
MAUVE ↑
MAUVE
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
Generation
Set-Aware
0.067
No Filter
0.061
Pointwise
0.024
Streaming PPL ↓
Validation PPL
225
200
175
150
125
100
75
50
Pointwise worst at G4
Pointwise
221.5
Set-Aware
182.2
No Filter
173.3