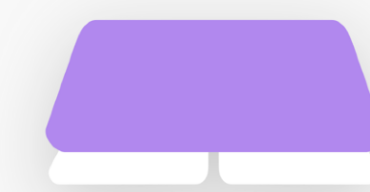


投機的実行

パフォーマンス計測



IMPlane

Convert

日本語IME
(クラウド・東京リージョン)

Selected

キャッシュのヒット率

95.65%

レスポンスタイム

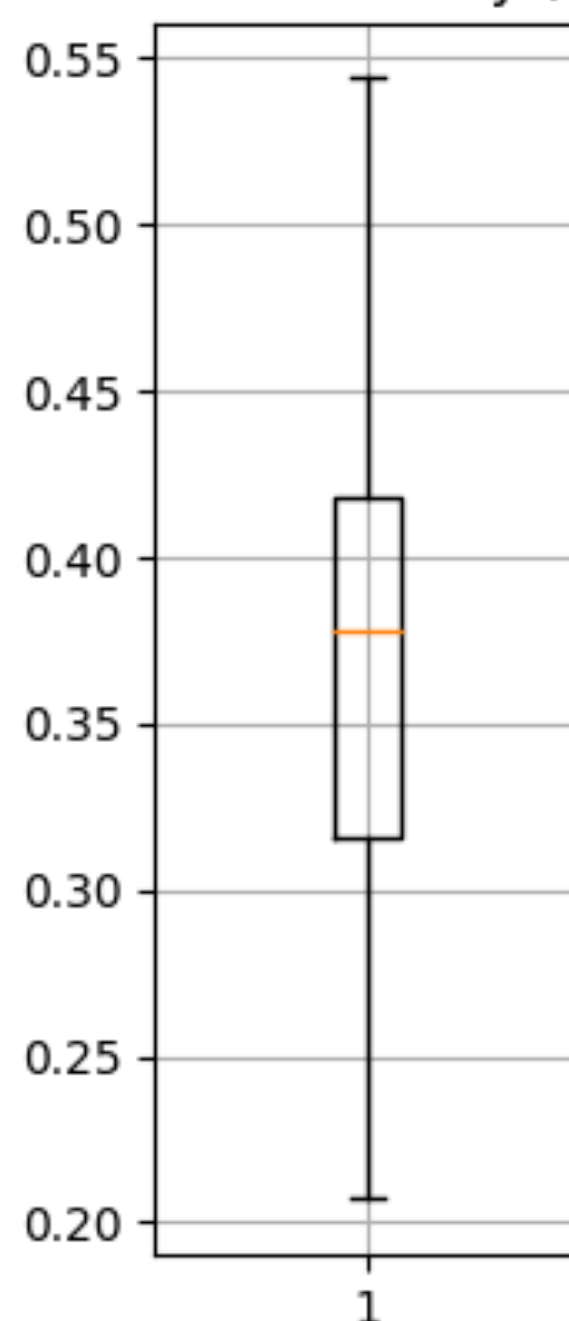
中央値

0.38 [ms]

25~75%

0.32~0.42 [ms]

Convert Latency (ms)



キャッシュのヒット率

52.38%

レスポンスタイム

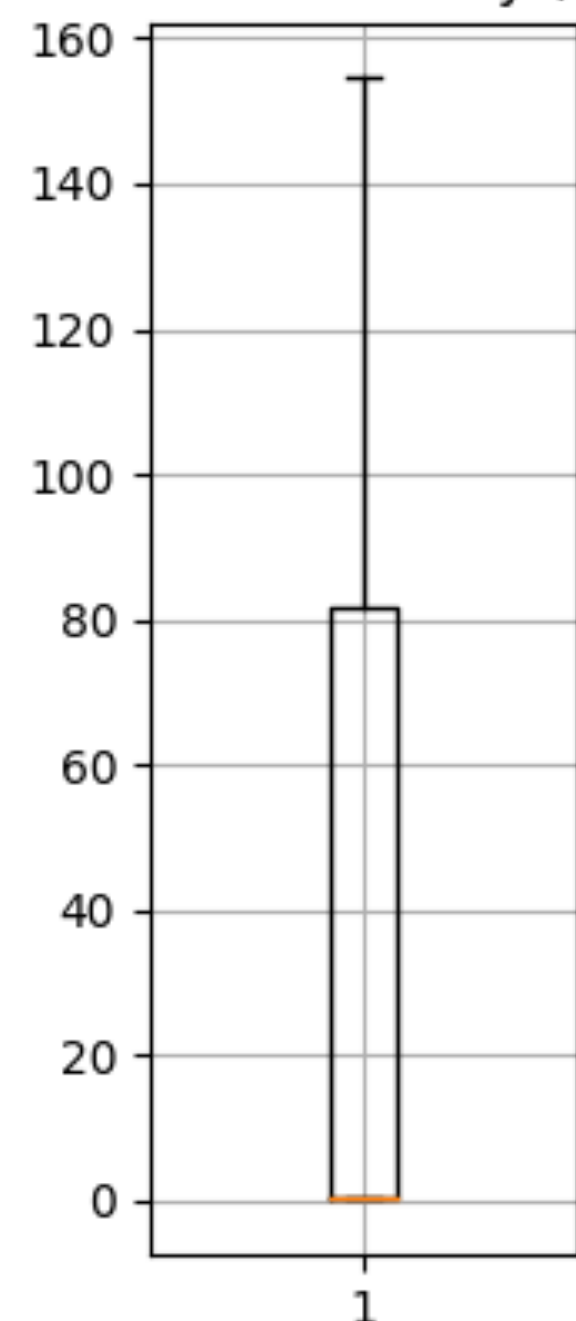
中央値

0.36 [ms]

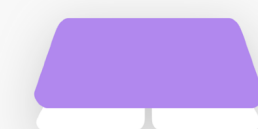
25~75%

0.25~81.55 [ms]

Selected Latency (ms)



GPT-3 IME



IMPlane