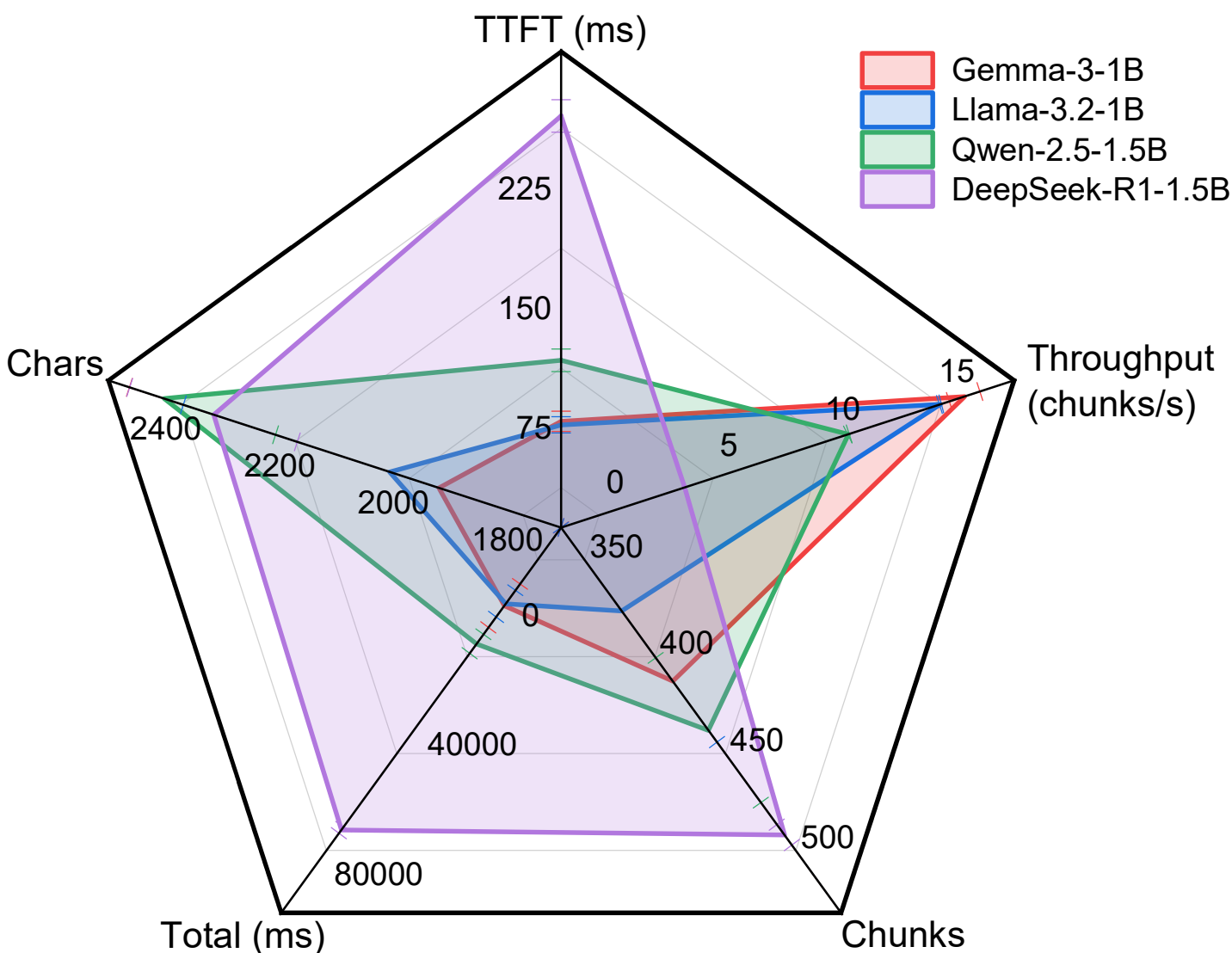


Evaluation on 2B-3B Models



Evaluation on 1B-1.5B Models