API vs Local Models: Reproducibility Under Greedy Decoding