With Guidelines and examples (301 tokens)

You will be given a question, a set of reference answers and the answer given by an LLM. Your task is to judge if the answer given by the LLM is correct, as if you were the LLMs teacher grading their exam. An answer should be counted as correct if it is semantically equivalent to (one of the) reference answers. In doing so, please follow the following guidelines (1-3):
1. Underspecified answers (e.g. "December" instead of "December 20") should be marked incorrect
2. Answers that have more information than requested (e.g. "December 20, in Paris" instead of "December 20") should be marked correct, provided the extra information is not incorrect or contrasting the rest of the answer
3. Answers that have correct references but are unnecessarily verbose should be marked correct (E.g. "Thanks for asking this question! The correct answer is: ...)
This is part of an automated evaluation process, therefore you must only output a single word: "correct" or "incorrect".
If you have trouble judging whether the answer is correct, for instance because you feel you are lacking knowledge required to judge so, please indicate so by marking the answer "maybe correct" or "maybe incorrect", so that we can further review it.

Question:
Which Australian did Roger Federer defeat to win his first Wimbledon Men's Singles title in 2003?

References:
MARK PHILIPPOUSSIS
MARK PHILIPPOUSSIS

Model Response:
Mark Philippoussis

Evaluation (correct/incorrect):