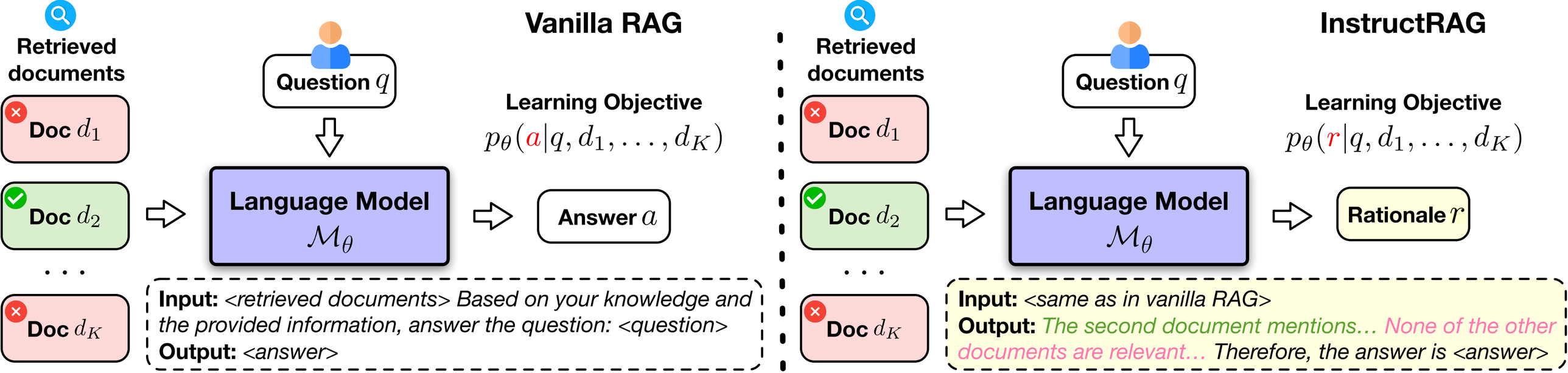
Vanilla RAG
Retrieved documents
Doc d_1
Doc d_2
Doc d_K
Question q
Language Model $\mathcal{M}_\theta$
Learning Objective
$p_\theta(a|q, d_1, \ldots, d_K)$
Answer a
Input: <retrieved documents> Based on your knowledge and the provided information, answer the question: <question>
Output: <answer>
InstructRAG
Retrieved documents
Doc d_1
Doc d_2
Doc d_K
Question q
Language Model $\mathcal{M}_\theta$
Learning Objective
$p_\theta(r|q, d_1, \ldots, d_K)$
Rationale r
Input: <same as in vanilla RAG>
Output: The second document mentions… None of the other documents are relevant… Therefore, the answer is <answer>