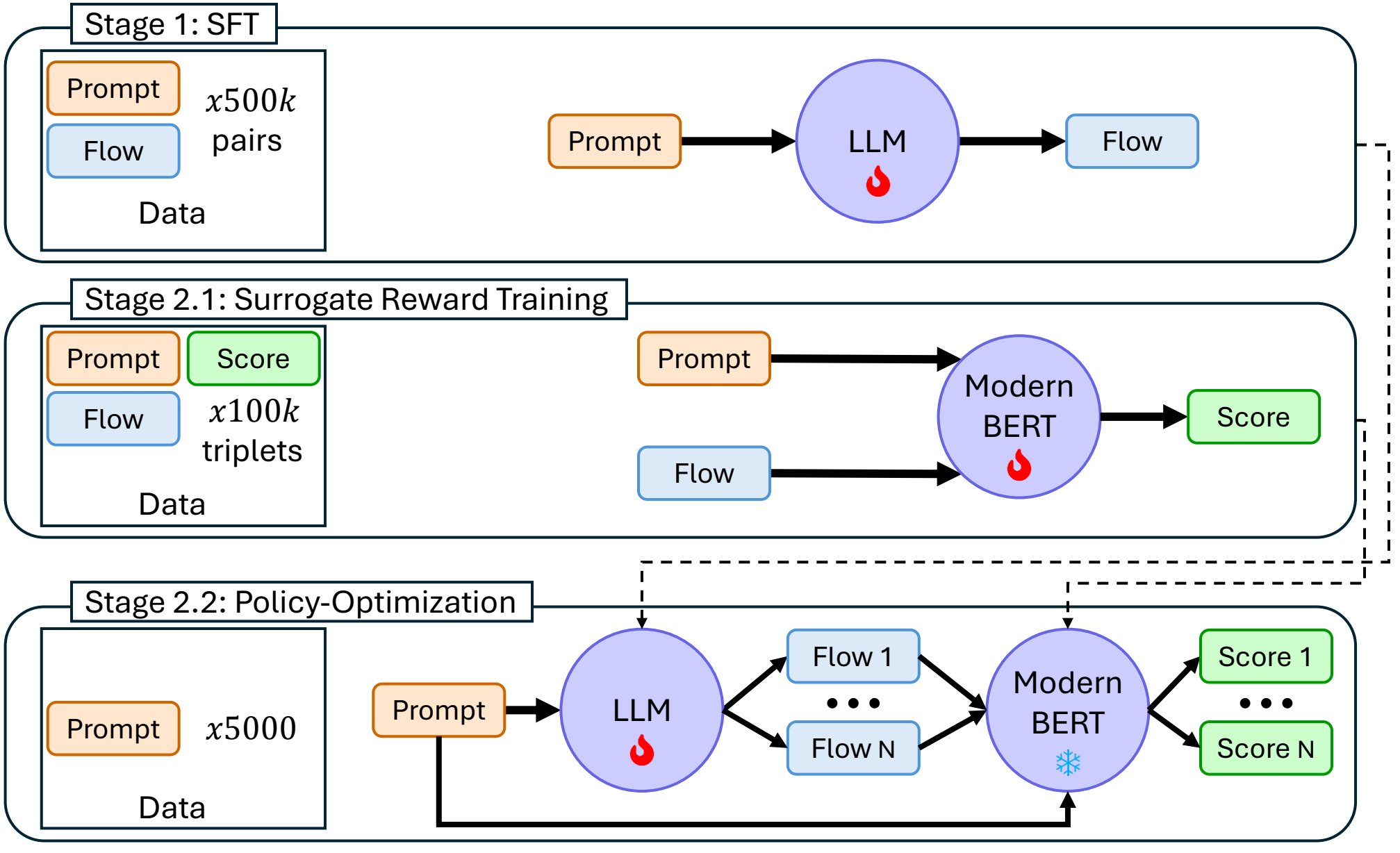
Stage 1: SFT
Prompt
Flow
$x500k$ pairs
Data
Prompt
LLM
Flow
Stage 2.1: Surrogate Reward Training
Prompt
Score
Flow
$x100k$ triplets
Data
Prompt
Flow
Modern BERT
Score
Stage 2.2: Policy-Optimization
Prompt
$x5000$
Data
Prompt
LLM
Flow 1
Flow N
Modern BERT
Score 1
Score N