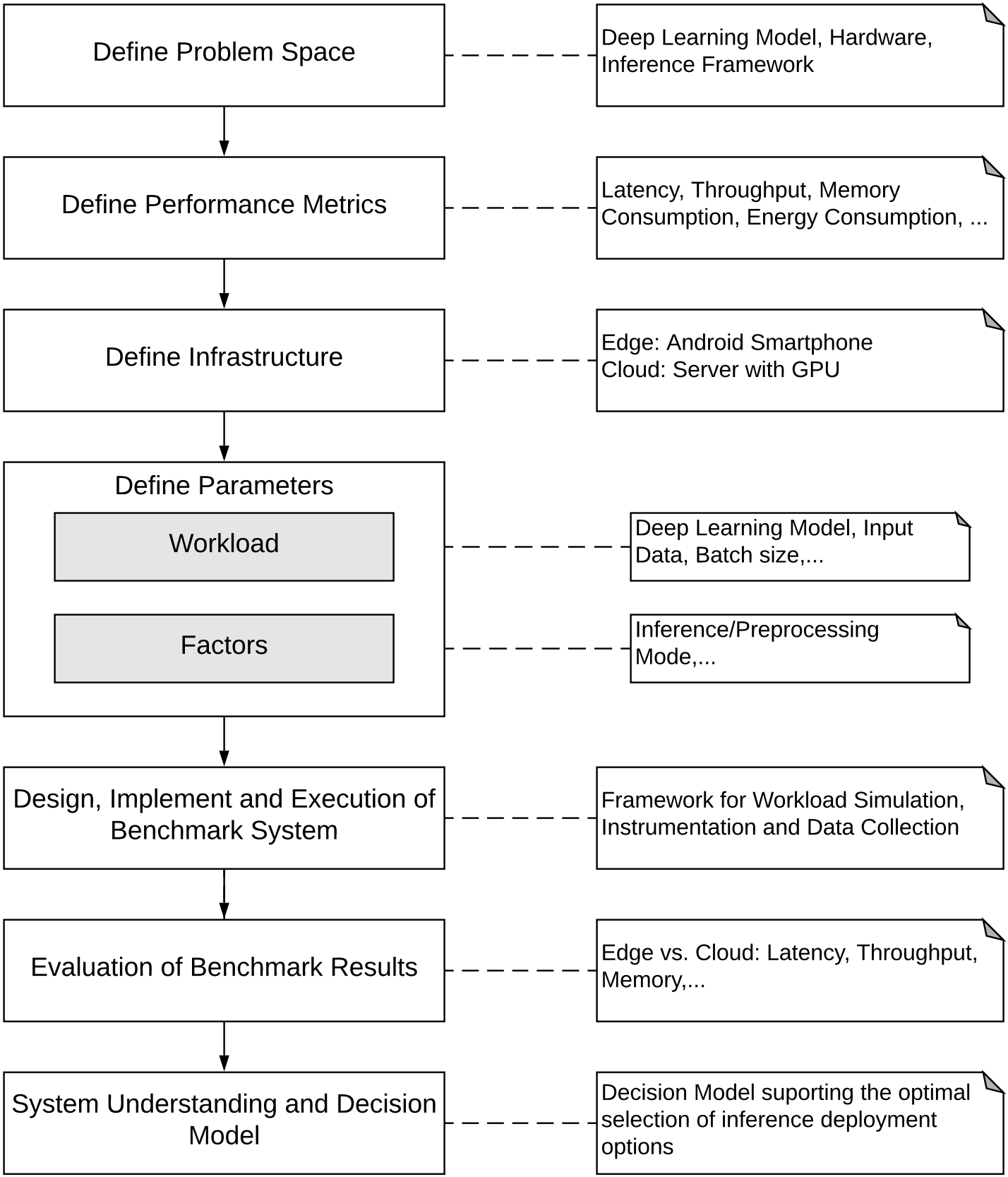
Define Problem Space
Deep Learning Model, Hardware, Inference Framework
Define Performance Metrics
Latency, Throughput, Memory Consumption, Energy Consumption, ...
Define Infrastructure
Edge: Android Smartphone
Cloud: Server with GPU
Define Parameters
Workload
Deep Learning Model, Input Data, Batch size,...
Factors
Inference/Preprocessing Mode,...
Design, Implement and Execution of Benchmark System
Framework for Workload Simulation, Instrumentation and Data Collection
Evaluation of Benchmark Results
Edge vs. Cloud: Latency, Throughput, Memory,...
System Understanding and Decision Model
Decision Model suporting the optimal selection of inference deployment options