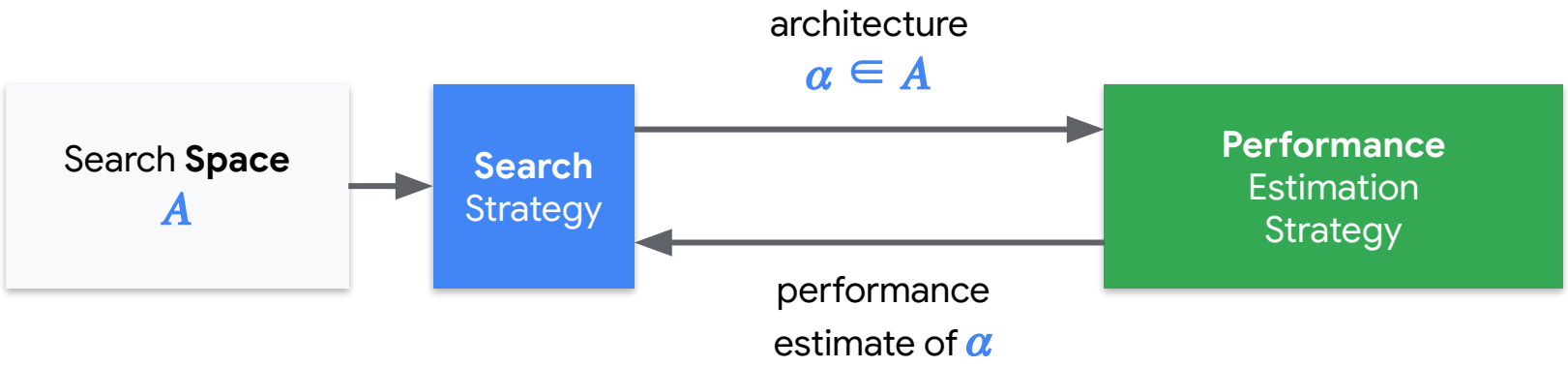
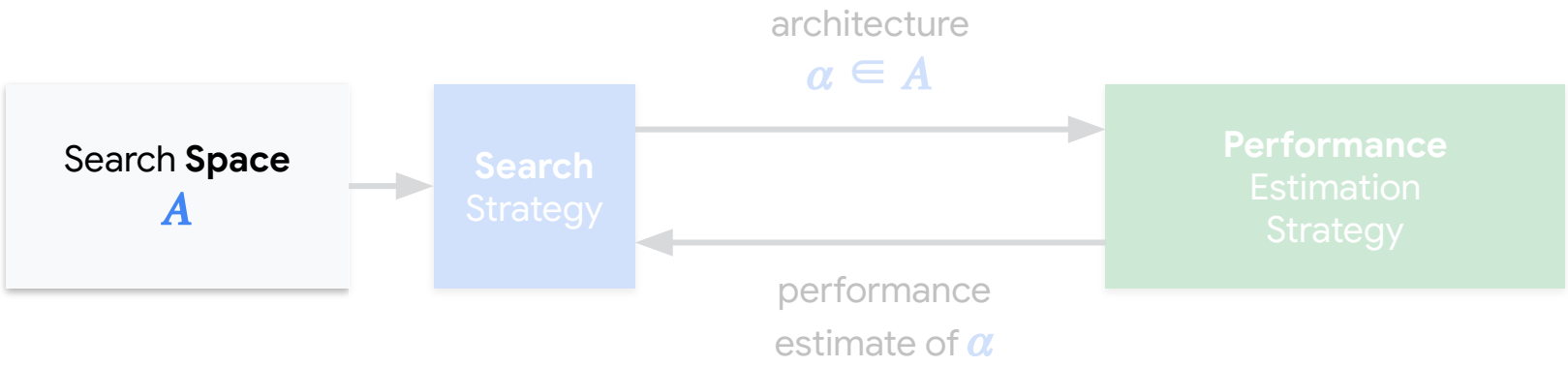
Neural Architecture Search Explained

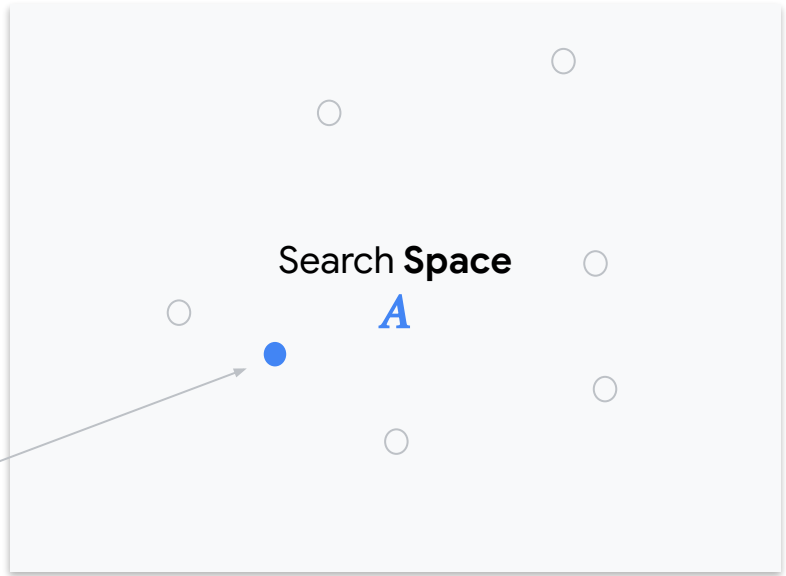
Neural Architecture Search: **Stages**



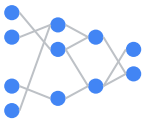
Search **Spaces**



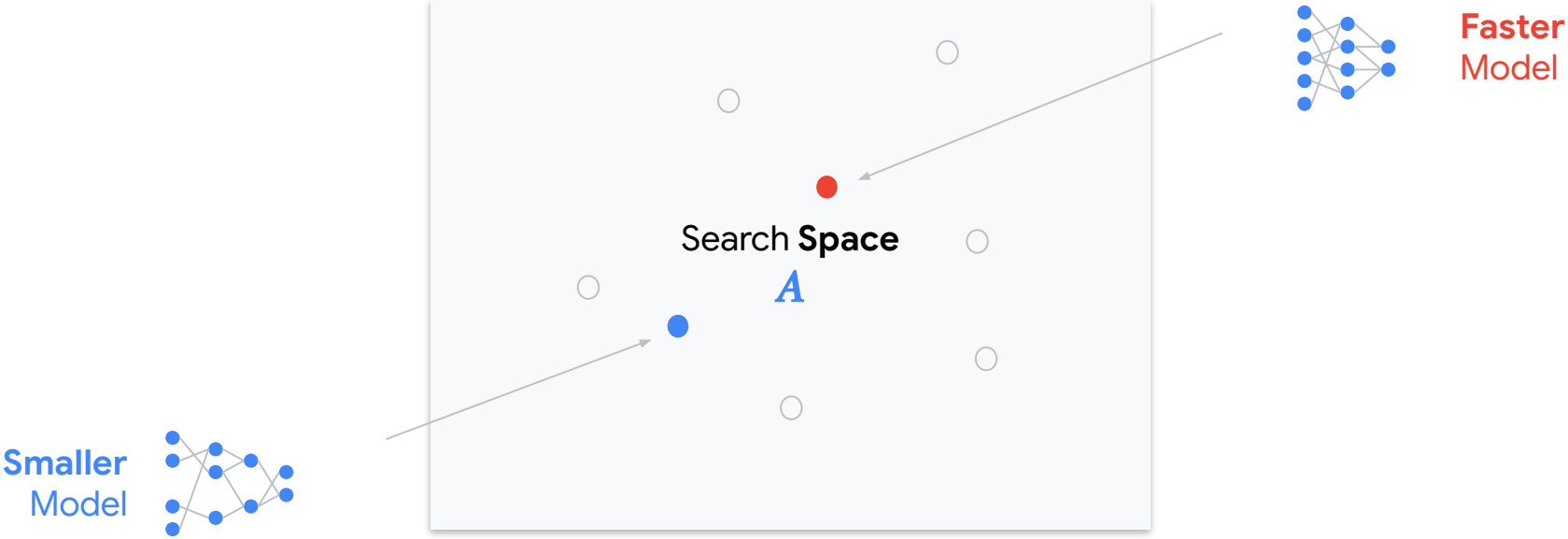
Search Space



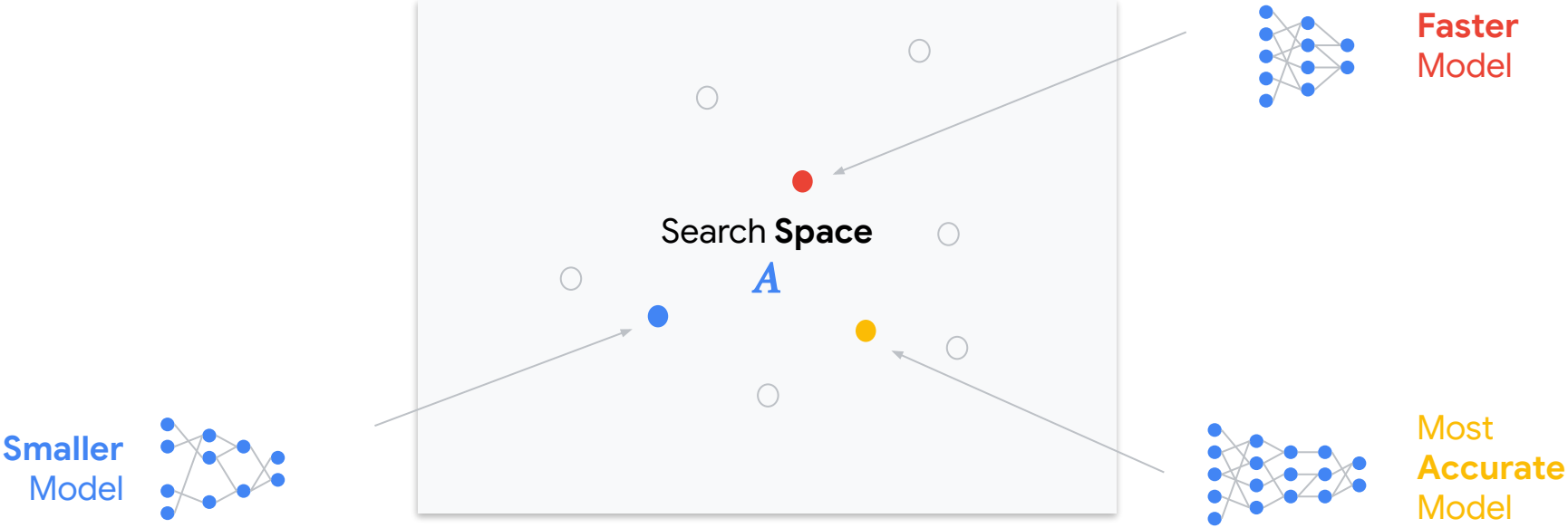
Smaller
Model



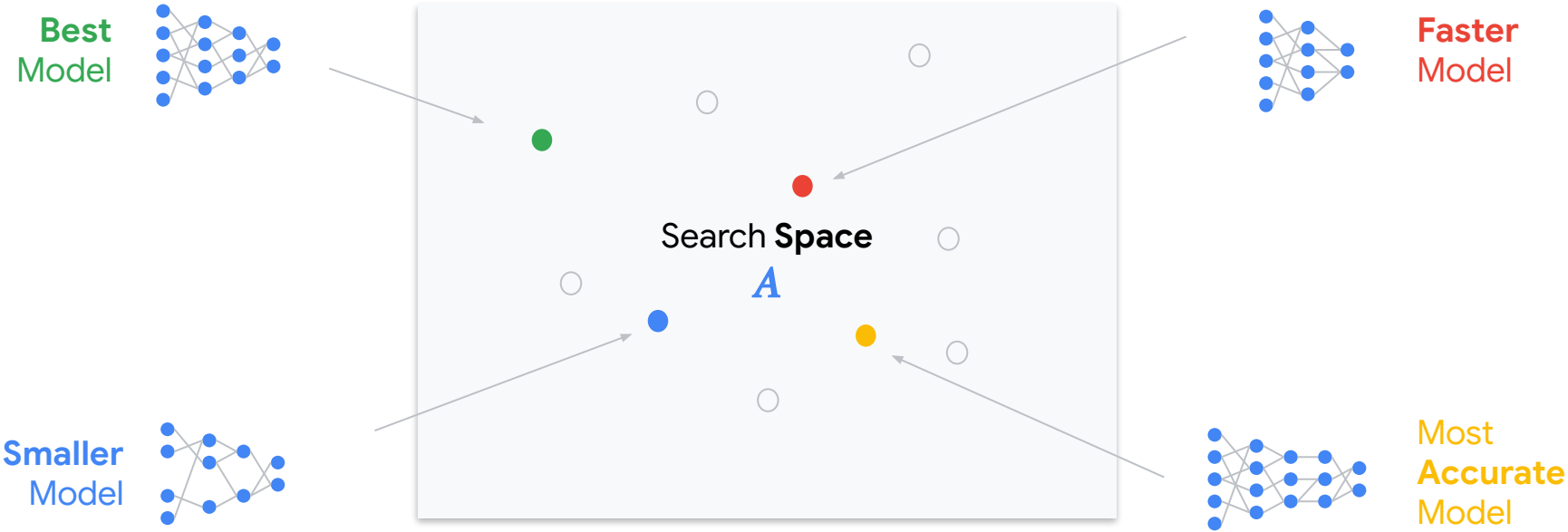
Search Space



Search Space

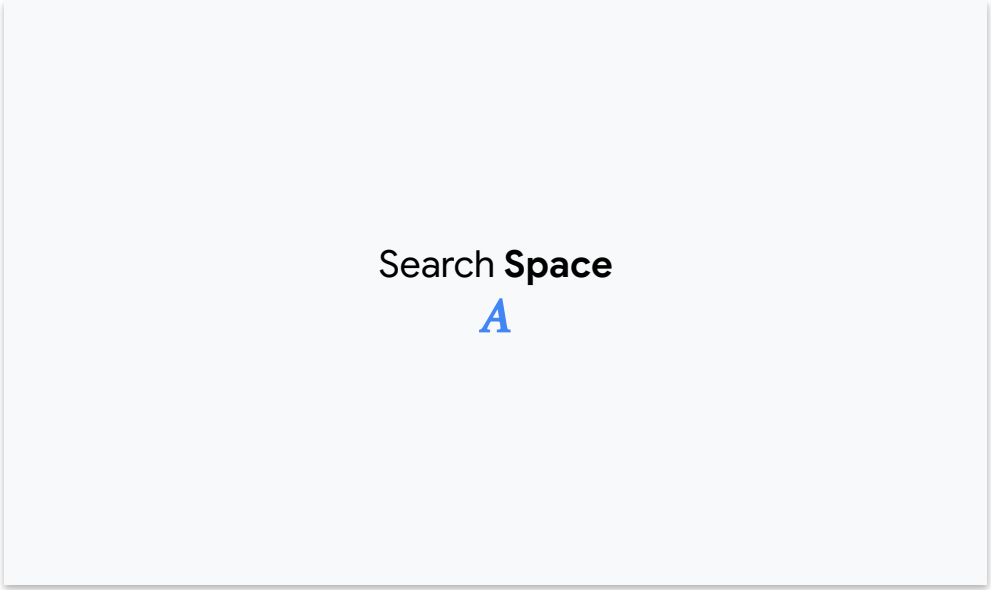


Search Space



Search Space **Size**

Too big!



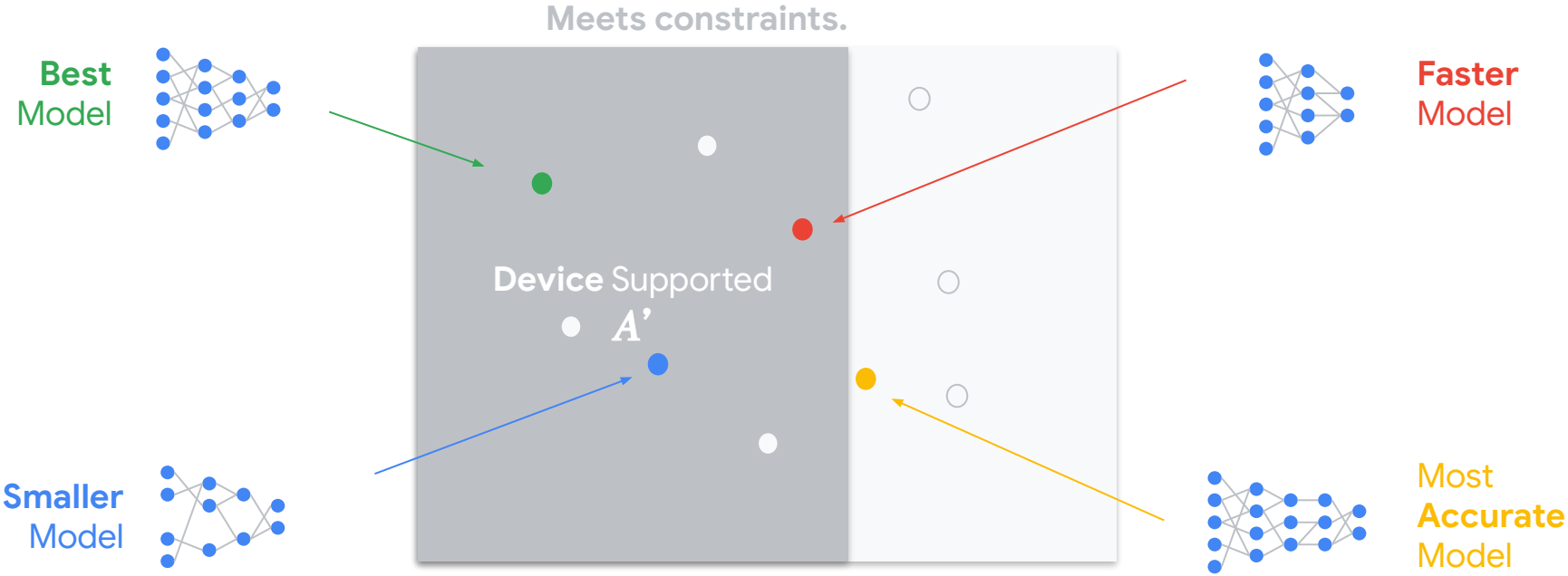
Search Space **Size**

Too small!

Search **Space**

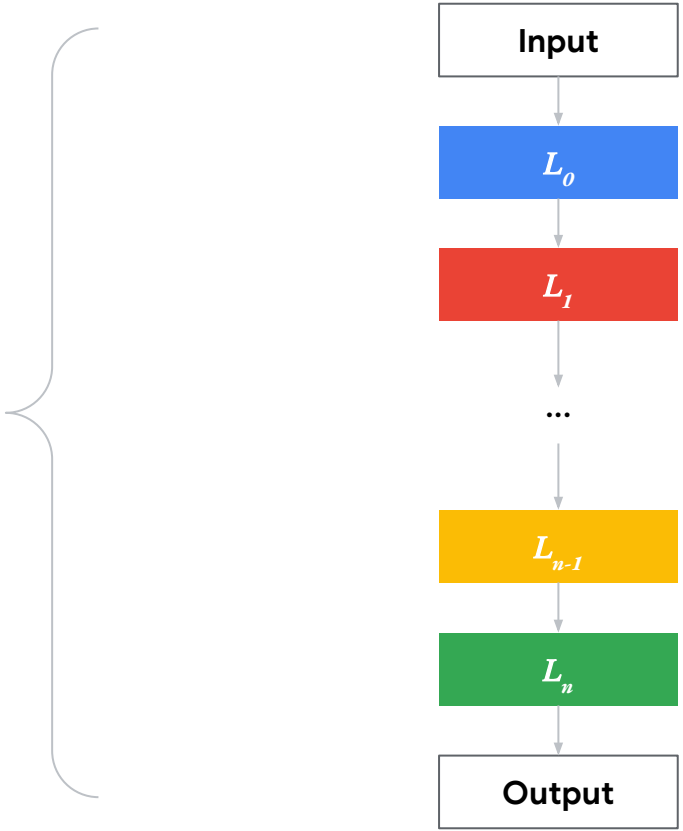
A

Prior Knowledge



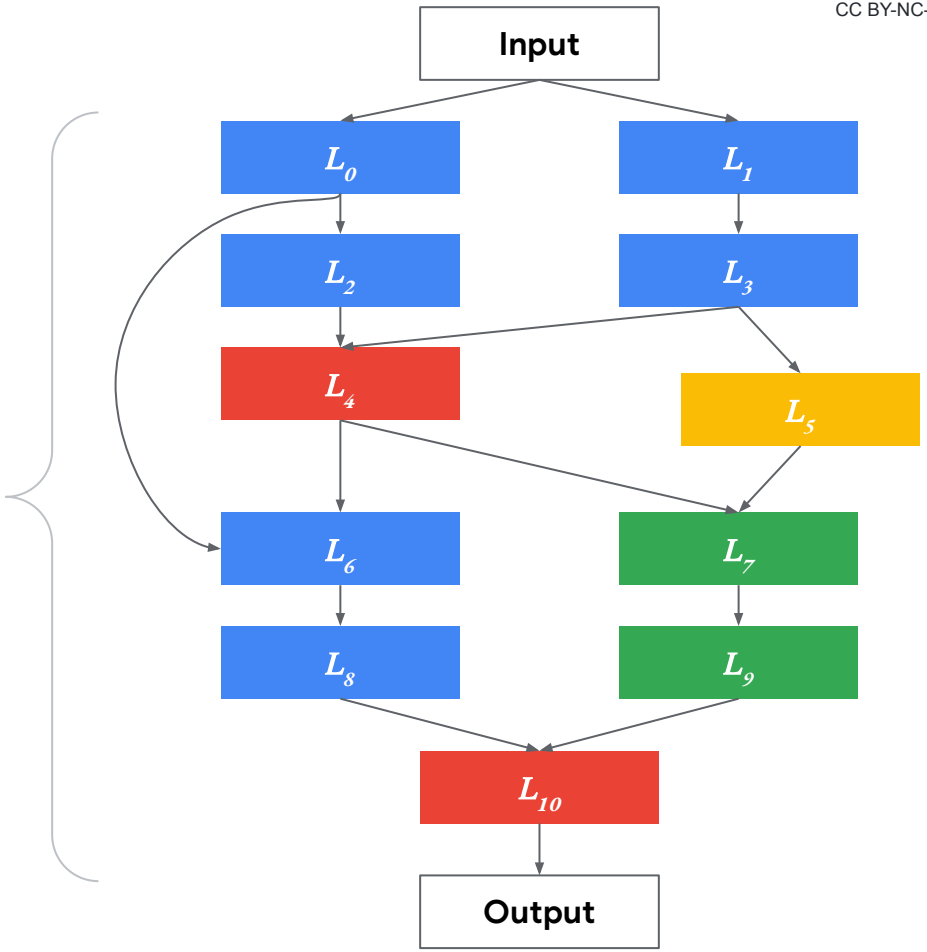
Search Space: **Types**

Chain structured



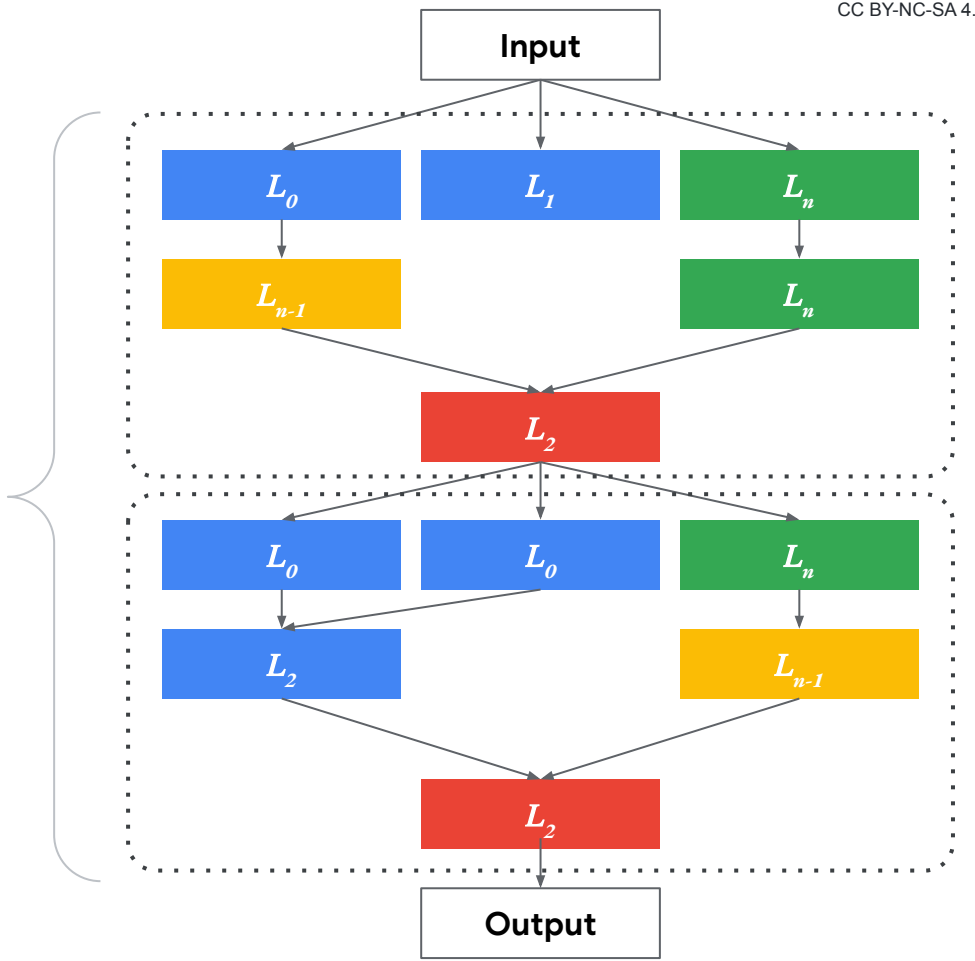
Search Space: **Types**

Multi branch

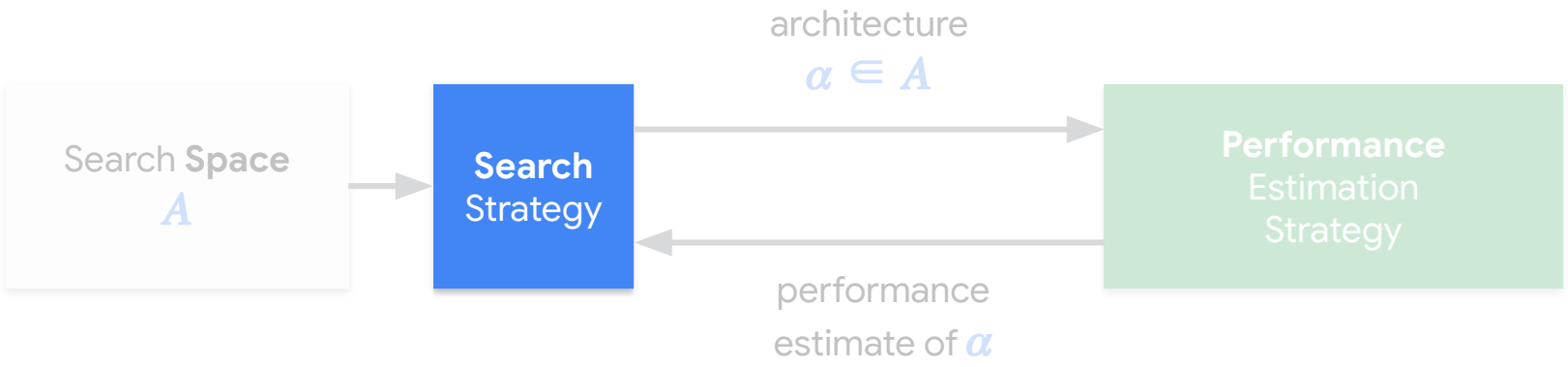


Search Space: **Types**

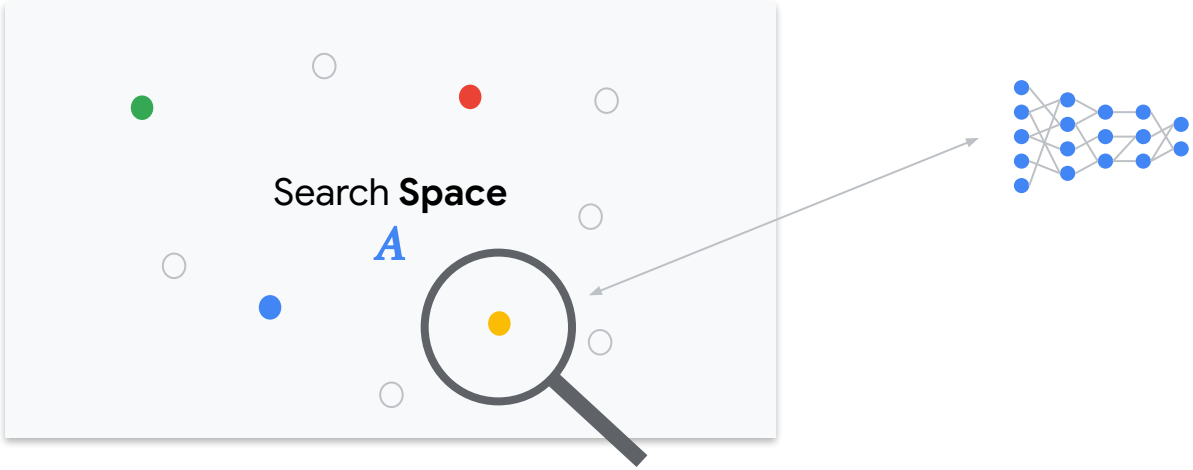
Cell-based



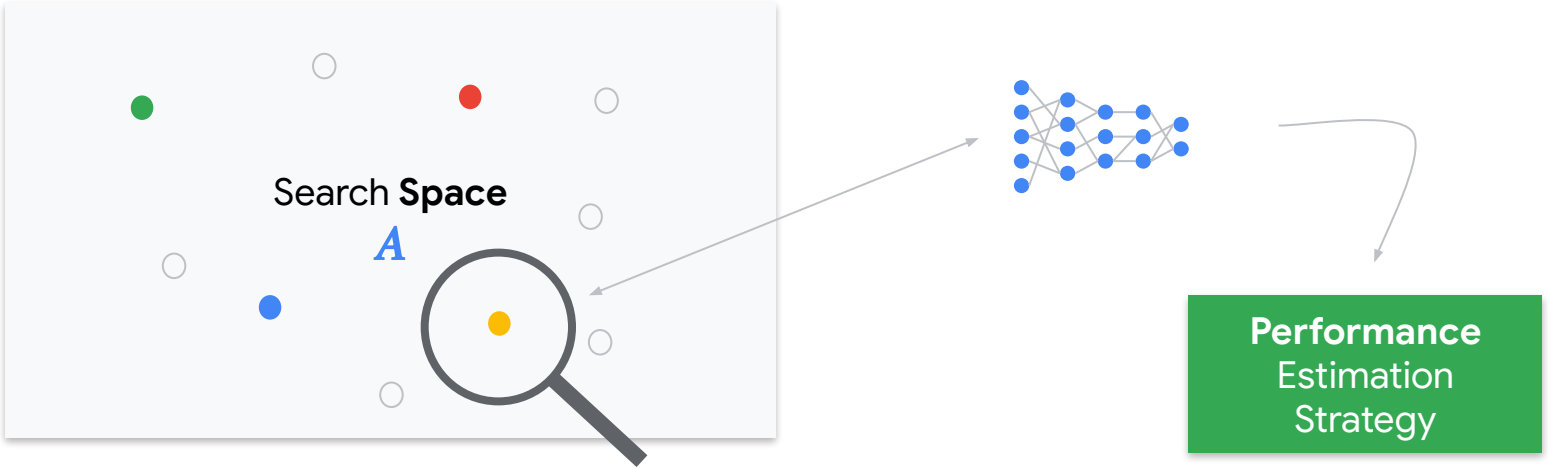
Search **Strategy**



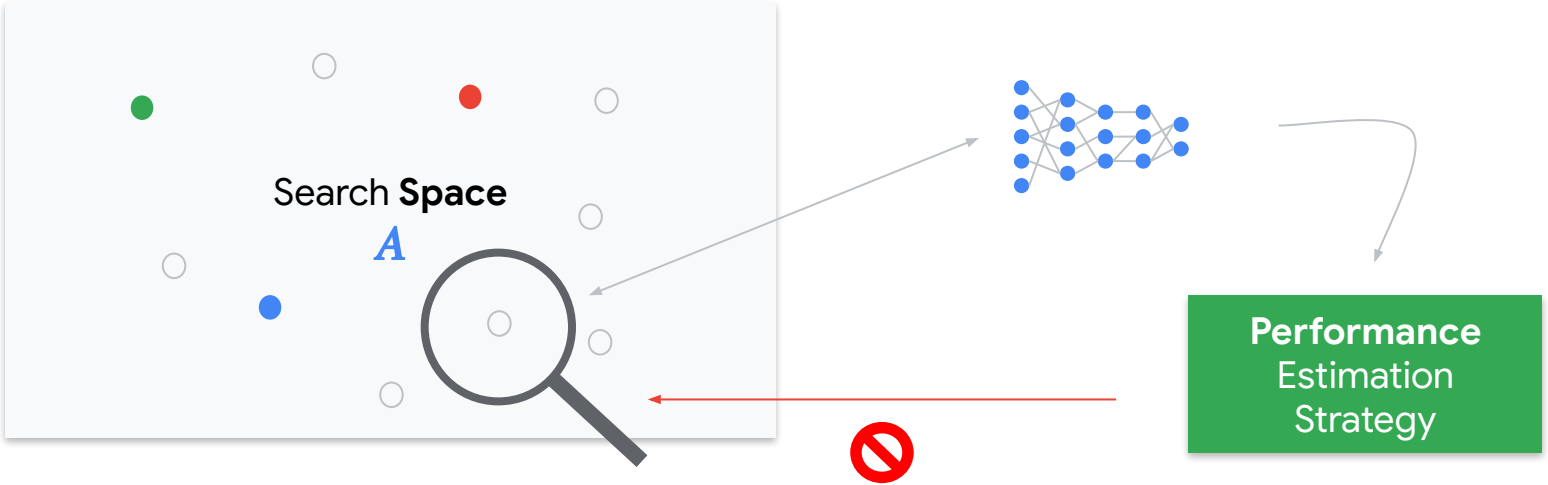
What is a Search **Strategy**?



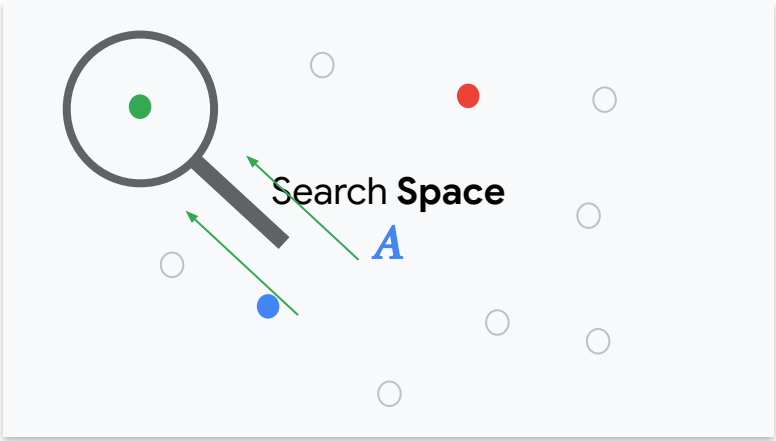
What is a Search **Strategy**?



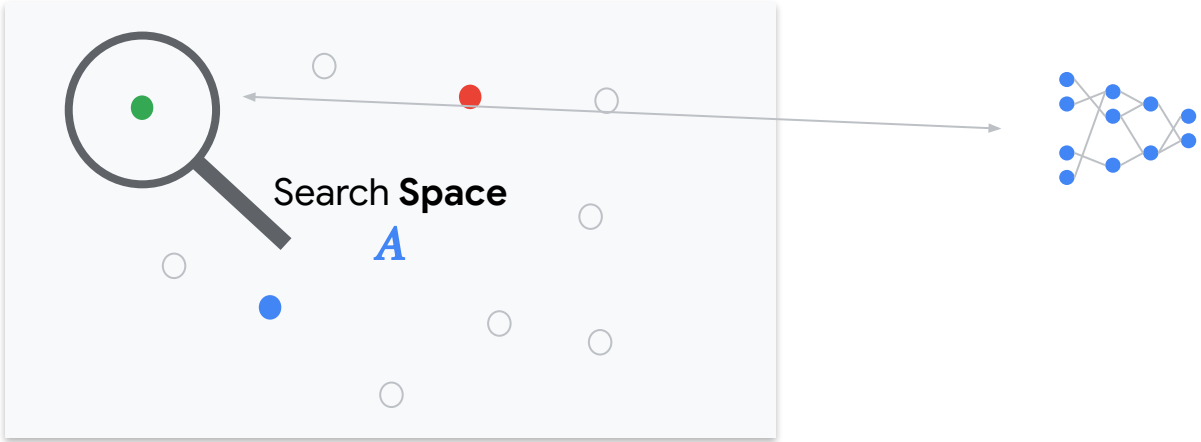
What is a Search **Strategy**?



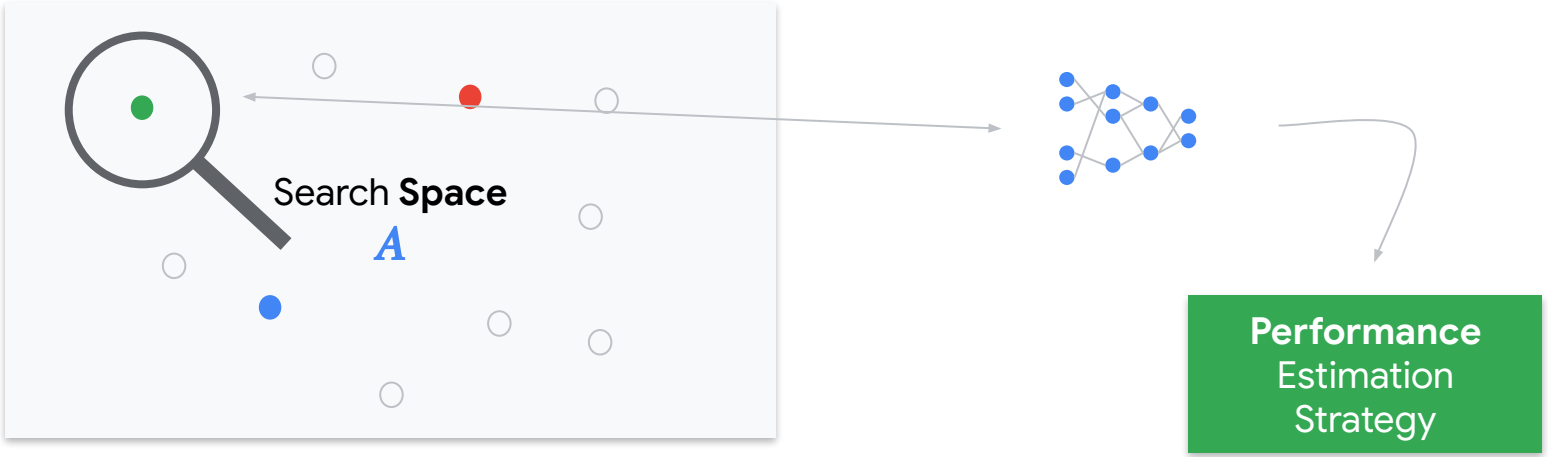
What is a Search **Strategy**?



What is a Search **Strategy**?



What is a Search **Strategy**?

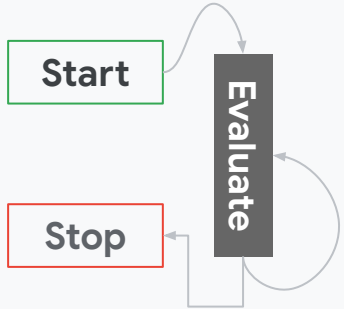


Types of Search Strategies

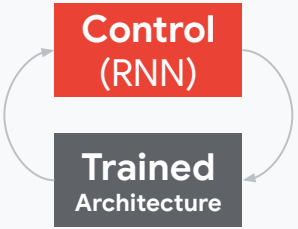
Bayesian Optimization

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

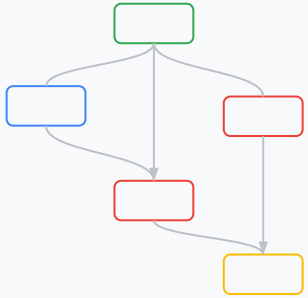
Evolutionary Methods



Reinforcement Learning



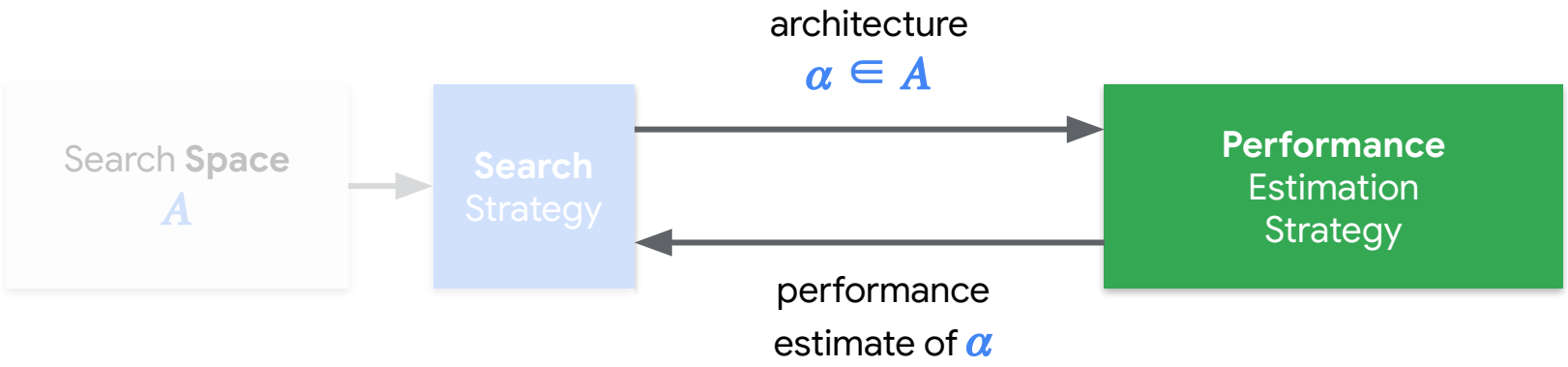
Gradient-based Methods



Selecting a Search Strategy

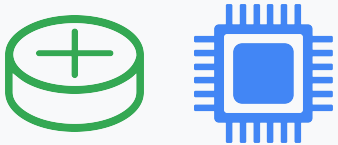
- The “Best” Search Strategy is Task dependent
- **DNAS** is direct and fast but requires *continuous* functions
- Blackbox methods are often *slower* but can be more **diverse/flexible**

Performance Estimation

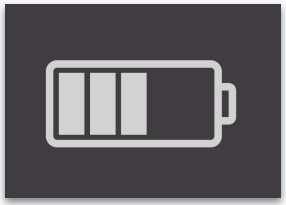


Profiling Metrics

Latency

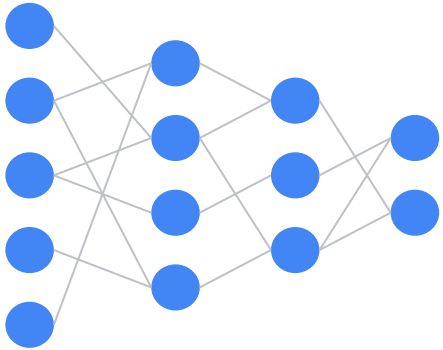


Memory



Energy

Direct Measurement or **Proxy** Metrics

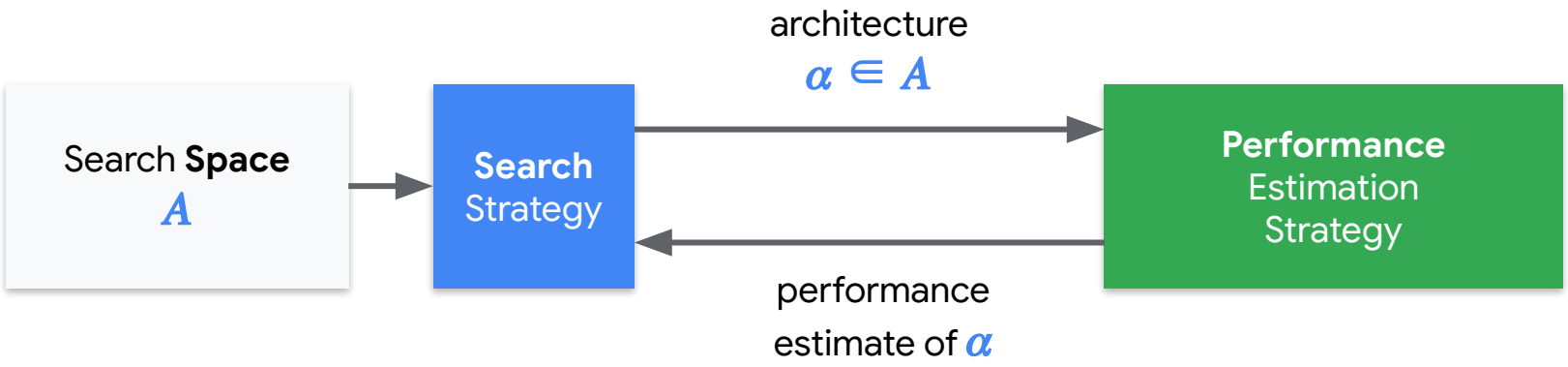


Math OPs

~

Latency

Neural Architecture Search



What is Neural Architecture Search?