AlphaGo combines a policy network and a value network with Monte Carlo Tree Search.

The policy network was first trained by supervised learning directly with expert human moves (30 million positions from KGS Go Server). This enables the network to predict expert moves with an accuracy of 57.0% using all input features. It had an impressive 55.7% accuracy even using just the raw board position and move history.

The policy network is further trained through reinforcement learning optimizing for winning the game instead of just predicting moves. The network played games between the previous policy network and a randomly selected previous iteration of the network to prevent overfitting. At this current stage, AlphaGo chooses moves that give the highest likelihood of a win. With just this additional reinforcement stage and no searching, the policy network was able to beat Pachi, 2 amateur dan on KGS, 85% of the games.

Next stage is to add a value network which evaluates each move and returns a value representing the score of such move. Previously, this cost / score / evaluation function was designed. AlphaGo's is trained and learned through previous game data. Another interesting point with AlphaGo's approach is that it prevented overfitting with training on complete game data by training on a 30 million distinct positions, each sampled from a separate game. This minimizes the difference on MSEs between the training and test sets. The final evaluation function is a combination of the value network output plus an exploration bonus. The exploration bonus decays with repeated consideration to encourage exploration.

AlphaGo is the stronger computer Go player winning games against all commercial and open source Go programs. In October, 2015, AlphaGo also won 5-0 games against a professional 2 dan and winner of 2013-2015 European Go championships, Fan Hui.