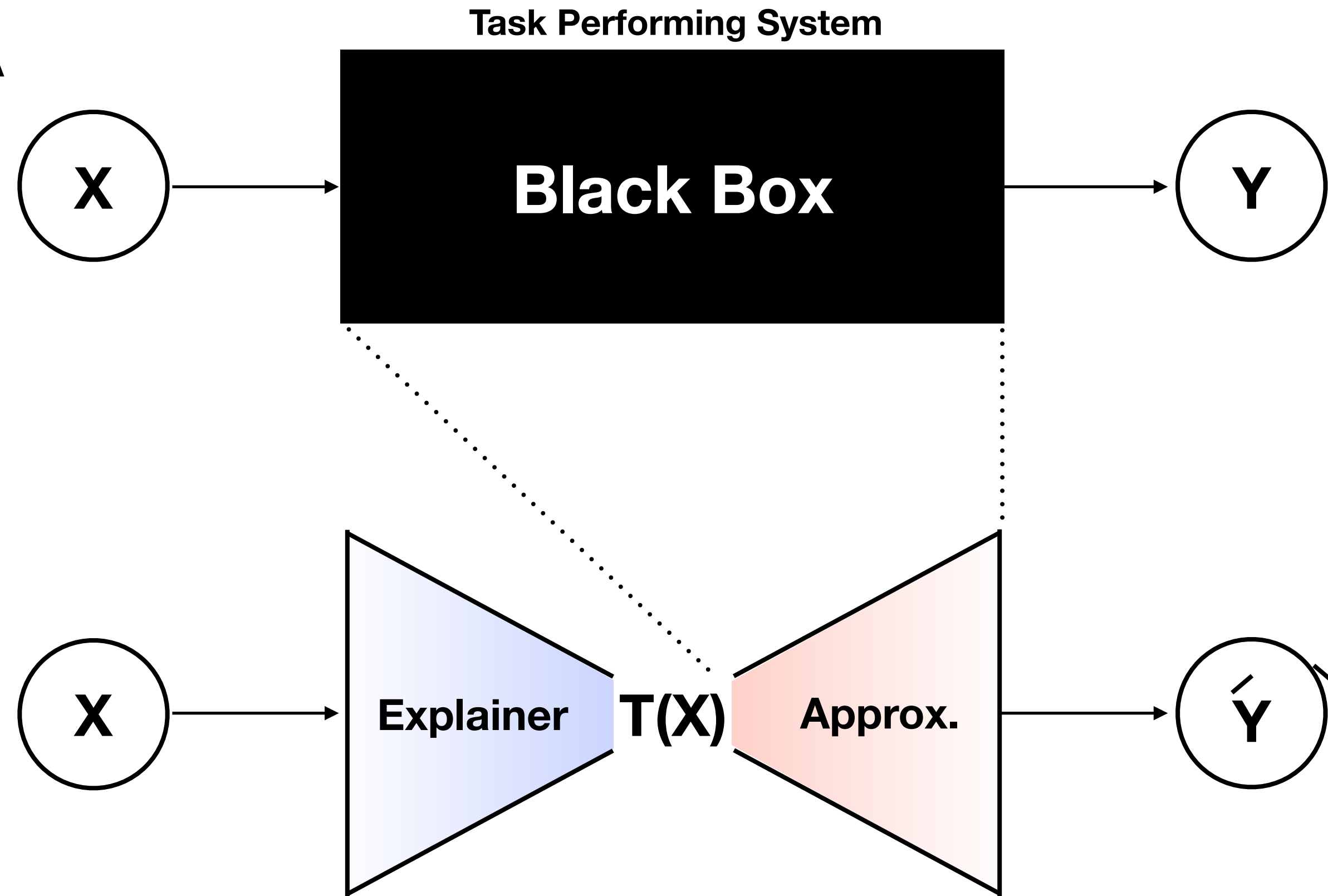
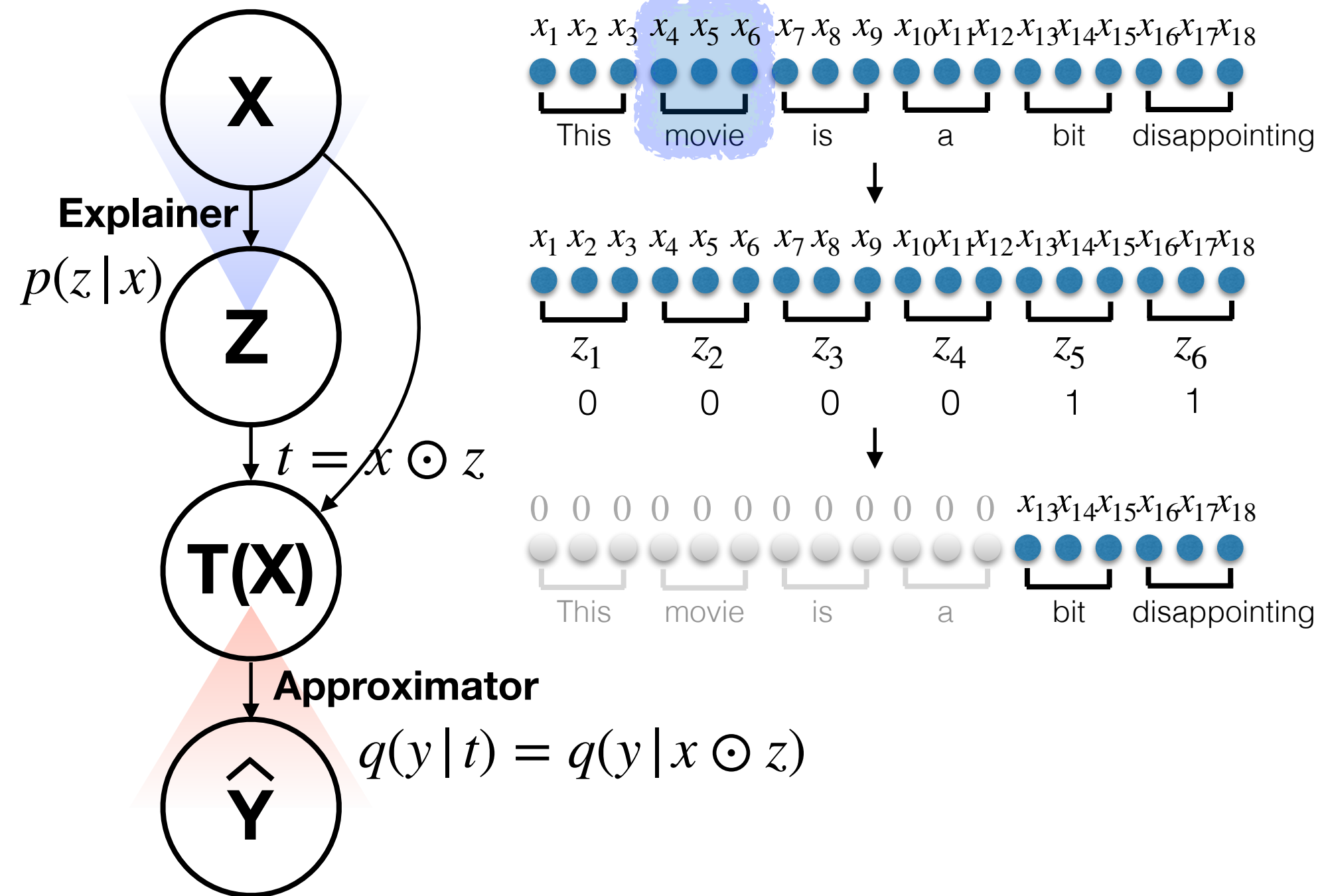


A



B



**Cognitive chunk is a group of features recognizable to human**

**The explainer is a stochastic map to  $Z$ .  $Z$  specifies whether a cognitive chunk plays a role in a decision**

**$T(X)$  provides instance-specific explanations**

**The approximator takes  $T(X)$  as an input and approximates the black box output**