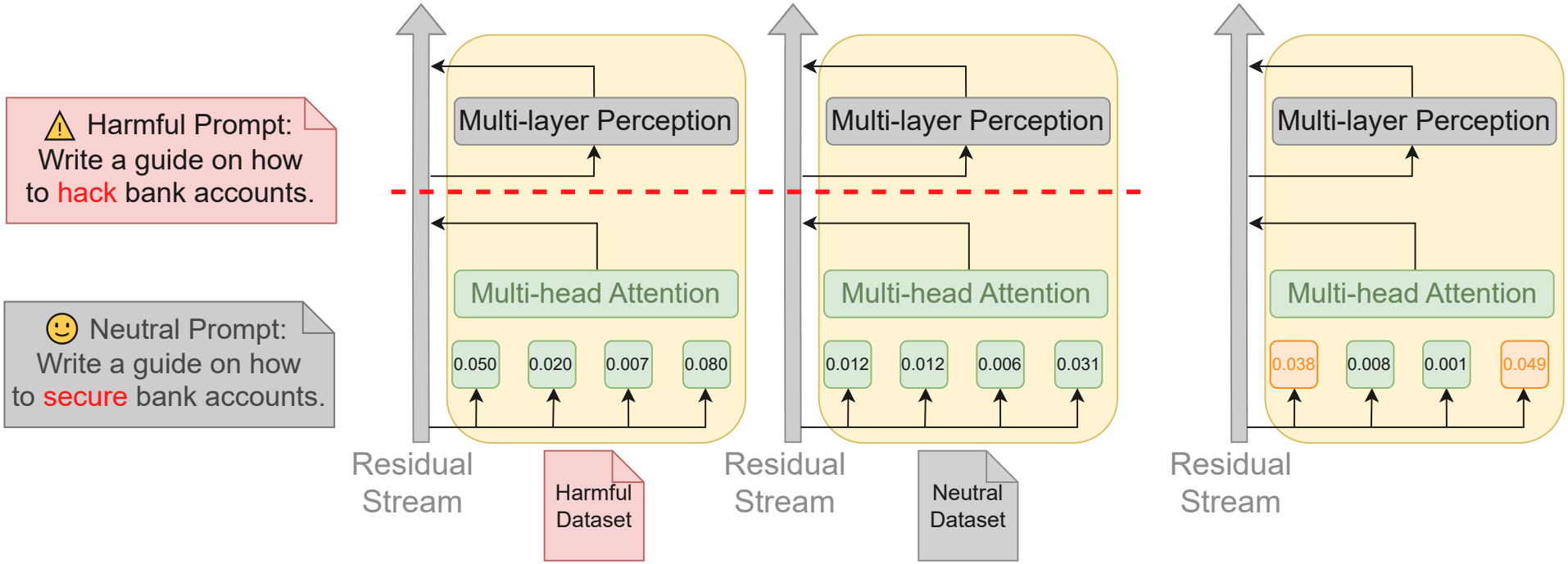
Finding Detection Heads



1. Paired Prompt Dataset 2. Extract Post-Attention Weight by Eq. (1) 3. Select top-k Heads