

## Dataset



White Body ✓  
White Feather ✓  
Black Body ✗

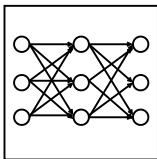


White Body ✗  
White Feather ✗  
Black Body ✓



White Body ✗  
White Feather ✗  
Black Body ✗

## Model



## Concept Vectors



White Body



White Feather

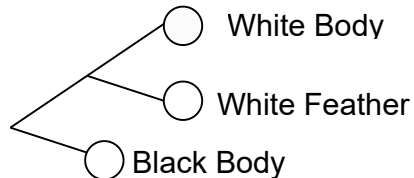


Black Body

● Concept Active ( ✓ )

● Concept Inactive ( ✗ )

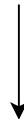
## Concept Hierarchy



## Downstream Tasks



White Body ✓  
White Feather ?  
Black Body ?



White Body ✓  
White Feather ?  
Black Body ✗