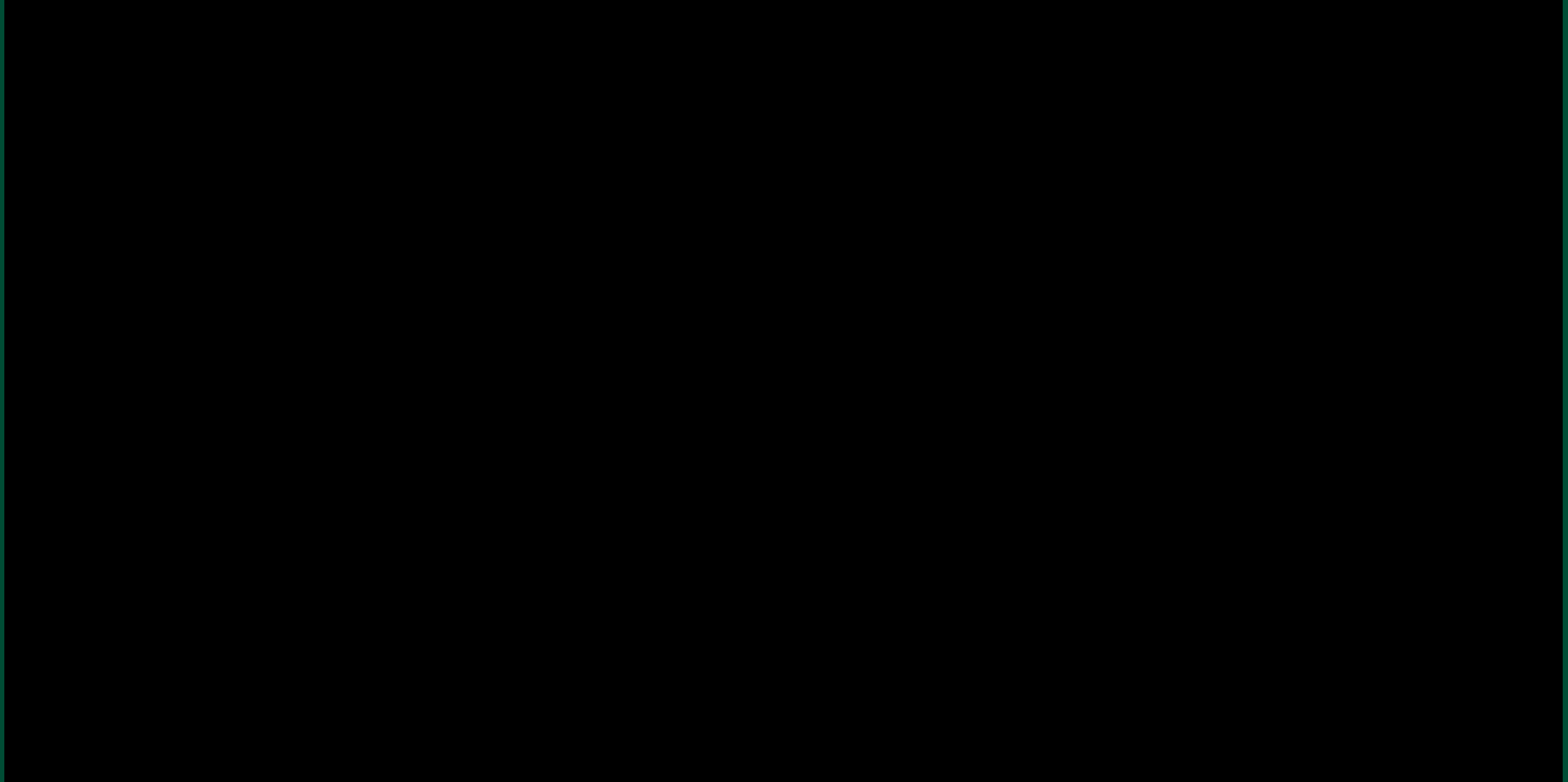


Demo



Example: Dogs are offensive

User post



Suggested edit



Example: Bikes are offensive

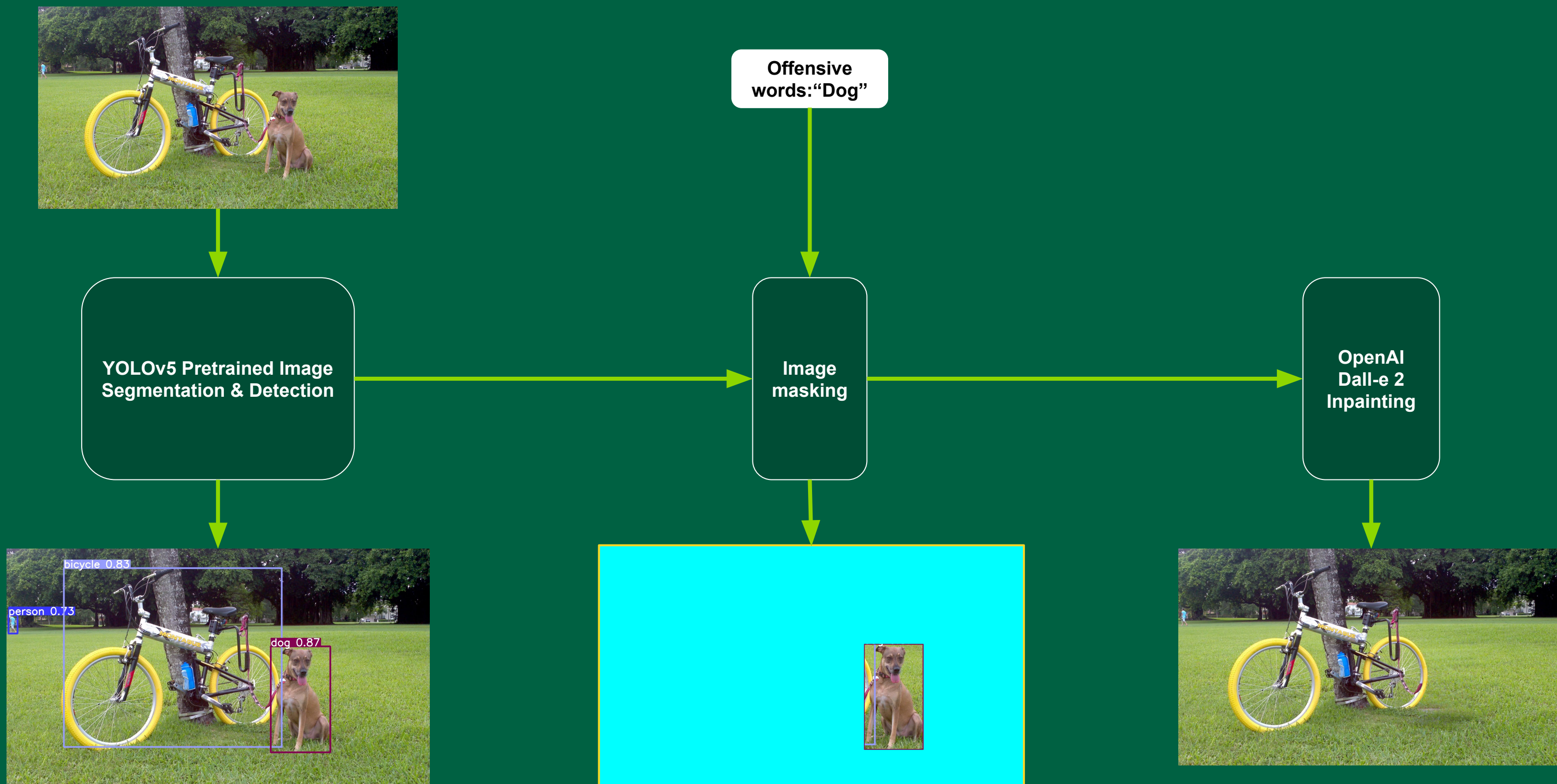
User post



Suggested edit



Architecture



Benefits to existing products

- **Customizable filters and definitions:** Google safe search API uses their own definition of harmful
- **Preserve image information:** No longer have to remove the entire image
- **Visual understanding:** No longer limited to text

Double use case

- **Customizable feeds for users:** Users can choose to filter out certain things they don't want to see in images on their feed



***Growth +
engagement
up***



Limitations + Future work

- **Limitation:** Types of detectable objects limited by pretrained model classes
- **Solution:** Finetune model on custom classes
 1. Ex. Don't want pictures of "iPhones"
 2. Generate pictures of iPhones with text-to-image (Dall-e, CLIP decoder)
 3. Finetune detector on augmented data
 4. 