O YOLO(você olha apenas uma vez, do inglês, You Only Look Once) é um framework de modelo de deep learning amplamente utilizado para tarefas relacionada a computação visual, assim como deteção de objetos, classificação, segmentação...

Comparando os modelos YOLOv11 temos algumas diferenças e característica específica de cada um como:  
**1. YOLO11 (Detecção Padrão):** Tem como objetivo detectar objetos em imagens e vídeos, possui desempenho rápido e eficiente em tarefas gerais de detecção, porém possui dificuldade em lidar com objetos pequenos ou parcialmente oclusos. Bastante confiável em ambientes claros e pouco congestionados.

**2. YOLO11-seg (Segmentação):** Além da detecção, produz máscaras pixel-a-pixel que delineiam os objetos, essas máscaras mais detalhas habilita a utilização em cenários mais complexos e com múltiplos objetos, porém requer mais recurso computacional.

**3. YOLO11-pose (Detecção de Pose):** Específico para detectar articulações e postura dos seres humanos, retornando pontos chaves da pose. Sendo capaz de lidar com diversas pessoas simultaneamente e possui desempenho bom em ambiente com pessoas bem visíveis.

**4. YOLO11-obb (Caixas Delimitadoras Orientadas – Oriented Bounding Boxes):** detecta objeto e retorna eles em caixas delimitadoras inclinadas, bem aplicável em objetos que não estão alinhados horizontalmente, ideal para cenários com objetos em diversas orientações, porém requer maior complexibilidade computacional.

**5. YOLO11-cls (Classificação):** Optimização para imagens simples, com preferencialmente apenas um objeto, classificando as em uma ou mais categorias, necessitando de uma menor complexidade computacional. Ideal para o nosso caso de estudo