1. **Изображения**:
   1. **CV**:
      1. **Классификация**:
         1. На базе CNN (обучаются на изображениях):
            1. **LeNet** (1998)
            2. **AlexNet** (2012);
            3. **GoogLeNet** (2012)
            4. **Inception** (2014)
            5. **VGG** (2014);
            6. **ResNet** (2015);
            7. **ResNeXt** (2017);
            8. **MobileNet** (2017);
            9. **EfficientNet** (2019).
         2. На базе трансформера (обучаются на изображениях):
            1. ViT (2020).
      2. **Детекция** = классификация + локализация:
         1. **R-CNN** (2013);
         2. **SSD** (Single Shot Detector, 2015);
         3. **Faster R-CNN** (2015);
         4. **YOLO** (2015 - 2023);
         5. **GroundingDINO** (zero-shot, 2023).
      3. **Трекинг** - задача получения начального набора обнаруженных объектов, создания уникального идентификатора для каждого из начальных обнаружений, а затем отслеживания каждого из объектов по мере их перемещения по кадрам видео, сохраняя присвоение идентификатора:
         1. **DeepSORT** (2017);
         2. **ByteTrack** (2021);
         3. **StrongSORT** (2022).
      4. **Оценка позы человека** - задача компьютерного зрения, цель которой — определить положение и ориентацию человека или объекта. Обычно это делается путем прогнозирования местоположения конкретных ключевых точек, таких как руки, голова, локти и т. д. в случае оценки позы человека:
         1. **OpenPose** (2018);
         2. **MediaPipe** (2019);
         3. **AlphaPose** (2022);
         4. **ViTPose** (2022);
         5. **YOLOv8** (2023).
      5. **Сегментация**:
         1. **Семантическая** – классификация каждого пикселя входного изображения:
            1. **SegNet** (2015);
            2. **UNet** (2015);
            3. **DeepLab** (2016);
            4. **UNet++** (2018);
            5. **SegFormer** (2021);
            6. **OneFormer** (2022).
         2. **Экземпляров** = **детекция** + **семантическая** сегментация:
            1. **Mask R-CNN** (2017);
            2. **OneFormer** (2022);
            3. **YOLOv7** (2022);
            4. **YOLOv8** (2023).
         3. **Паноптическая** = **семантическая** для аморфных фоновых областей + **экземпляров** для уникальных объектов четкой формы:
            1. **OneFormer** (2022);
            2. **Mask DINO** (2022);
   2. **Генерация изображений:**
      1. **Stable Diffusion** (2021);
      2. **StyleGAN** (2022);
      3. **Kandinsky** (2022).
   3. **Генерация текстовых описаний к изображениям:**
      1. **mPLUG** (2022);
      2. **BLIP** (2022).
2. **NLP**:
   1. Классификация;
   2. Вопрос-ответ;
   3. Последовательность в последовательность;
   4. Распознавание именованных сущностей;
   5. Тегирование;
   6. Поиск зависимостей;
   7. Лемматизация.
3. **Распознавание речи:**
   1. **VOSK** (2020);
   2. **Whisper** (2022);
4. **Синтез речи:**
   1. **Silero** (2020).