

## BẢNG QUY TẮC GÁN NHÃN CHI TIẾT

### 1. LOC (Location - Địa điểm)

**Định nghĩa:** Bao gồm tên Đường, Phường, Quận, Thành phố, Tỉnh, Khu vực, hoặc các địa danh nổi tiếng (Landmark) dùng để định vị.

#### Quy tắc biên (Boundary):

- **NÊN:** Bôi đen cả từ loại đi kèm nếu nó là một phần của tên riêng không thể tách rời hoặc giúp định danh rõ hơn. Ví dụ: [Quận 7], [TP.HCM], [KCN Phước Đông].
- **KHÔNG NÊN:** Bôi đen các từ chỉ vị trí chung chung như "gần", "cạnh", "đối diện", "mặt tiền".
- **Ví dụ từ data của bạn:**
  - o "đường [Số 12]" (Hoặc [đường Số 12] đều được, nhưng phải thống nhất). *Khuyến dùng: Bôi cả từ loại -> [đường Số 12].*
  - o "phường [Hiệp Bình Phước]", "Thành phố [Hồ Chí Minh]".
  - o "Quốc Lộ [13]".
  - o "gần [bệnh viện Đa Khoa]".
  - o "Khu dân cư [Mỹ Hạnh Bắc]".

### 2. PRICE (Giá cả)

**Định nghĩa:** Giá trị tiền tệ của bất động sản.

#### Quy tắc biên:

- **NÊN:** Bôi đen Số + Đơn vị tiền tệ.
- **MỞ RỘNG:** Nếu giá là khoảng hoặc thương lượng, có thể bôi luôn cụm từ đó.
- **Ví dụ từ data của bạn:**
  - o "[14 tỷ]".

- "[120 triệu/m2]".
- "[1 tỷ 290 triệu]".
- "[3,8 tỷ]".
- "[899tr]".

### 3. AREA (Diện tích)

- **Định nghĩa:** Diện tích đất, diện tích sàn, hoặc kích thước Ngang x Dài.
- **Quy tắc biên:** Bôi đen Số + Đơn vị đo (m2, ha, m).
- **Ví dụ từ data của bạn:**
  - "[230m2]".
  - "[11,5x20m]" (Kích thước ngang dọc cũng tính là AREA).
  - "[1000m2]".
  - "[54m2]".
  - "[12x84m]".

### 4. TYPE (Loại hình Bất động sản)

- **Định nghĩa:** Loại tài sản đang giao dịch.
- **Quy tắc biên:** Chỉ bôi đen danh từ chỉ loại hình.
- **Ví dụ từ data của bạn:**
  - "[lô đất]".
  - "[Căn hộ]".
  - "[Officetel]".
  - "[đất thổ cư]".
  - "[nhà vườn]".

### 5. ATTR (Attribute - Thuộc tính/Tiện ích/Pháp lý)

- **Định nghĩa:** Đây là nhãn "siêu to khổng lồ", bao gồm mọi đặc điểm mô tả về BĐS:
  - **Pháp lý:** Sổ đỏ, sổ hồng, pháp lý rõ ràng.

- **Nội thất:** Full nội thất, giao thô.
- **Cấu trúc:** 3 phòng ngủ, view sông, mặt tiền, nở hậu.
- **Tiện ích:** Hồ bơi, công viên, gần chợ.
- **Hướng:** Hướng Bắc, Đông Nam.
- **Ví dụ từ data của bạn:**
  - *Pháp lý:* "[Sổ hồng riêng]", "[pháp lý rõ ràng]", "[giấy tờ pháp lí đầy đủ]".
  - *View/Hướng:* "[view sông]", "[view Landmark 81]", "[Hướng Bắc]", "[mặt tiền]".
  - *Cấu trúc:* "[3 phòng ngủ]", "[2 nhà vệ sinh]", "[đất nở hậu nhẹ]".
  - *Hiện trạng:* "[full thổ cư]", "[đường nhựa rộng 14m]", "[xe hơi ra vào thoải mái]".

## 6. PER (Person - Người)

- **Định nghĩa:** Tên riêng của người môi giới hoặc chủ nhà.
- **Lưu ý:** Trong BĐS thường dùng từ "Chính chủ", có thể gán là PER nếu muốn, hoặc coi là O. Nhưng tốt nhất nên gán tên riêng.
- **Ví dụ từ data của bạn:**
  - "[Dũng]" (trong câu "gặp Dũng").
  - "[Duy]".

## 7. ORG (Organization - Tổ chức)

- **Định nghĩa:** Tên công ty, ngân hàng, chủ đầu tư, dự án lớn (nếu coi dự án là tổ chức quản lý).
- **Ví dụ từ data của bạn:**
  - "[VCB]" (Ngân hàng Vietcombank).
  - "[Lavida Plus]" (Tên dự án - Có thể coi là LOC hoặc ORG, nhưng thường tên dự án chung cư gán là ORG hoặc LOC\_PROJECT. Nếu nhóm chốt 8 nhãn thì gán vào **ORG** hoặc **LOC** tùy thống nhất. *Khuyến dùng: Gán tên dự án vào LOC*).

- "[Phú Đông Group]".
- "[Vincom]".

## 8. O (Others)

- Tất cả các từ không thuộc các nhãn trên (từ nối, động từ "bán", "gửi", "tại", dấu câu...). Không cần bôi đen (Label Studio tự hiểu là O).

---

## CÁC TRƯỜNG HỢP "KHÓ" CẦN THỐNG NHẤT (EDGE CASES)

### 1. Dự án (Project Name): "*Căn hộ Lavida Plus*"

- Lavida Plus là LOC (địa điểm) hay ORG (tên brand)?
- **Quyết định:** Gán là **LOC** (vì người dùng tìm nhà *tại* Lavida Plus).

### 2. Địa chỉ ghép: "*Đường Số 12 phường Hiệp Bình Phước*"

- Cách 1 (Gộp): [LOC đường Số 12 phường Hiệp Bình Phước]
- Cách 2 (Tách - **Khuyến dùng**): [LOC đường Số 12] [LOC phường Hiệp Bình Phước]
- *Lý do:* Tách ra giúp mô hình hiểu cấu trúc phân cấp (Đường -> Phường).

### 3. Diện tích phức tạp: "*100m2 (5x20)*"

- Gán: [AREA 100m2] và [AREA 5x20].

### 4. Giá khoảng/Thương lượng: "*14 tỷ có thương lượng*"

- Cách 1: [PRICE 14 tỷ] [ATTR có thương lượng]
- Cách 2: [PRICE 14 tỷ có thương lượng]
- **Khuyến dùng:** Cách 1 (Giá là số, thương lượng là thuộc tính).

Ví dụ Gán mẫu trên 1 câu hoàn chỉnh

**Câu gốc:**

"Chính chủ gửi bán lô đất mặt tiền đường Số 12 phường Hiệp Bình Phước giá 14 tỷ sổ hồng riêng."

**Gán nhãn:**

[PER Chính chủ] gửi bán [TYPE lô đất] [ATTR mặt tiền] [LOC đường Số 12] [LOC phường Hiệp Bình Phước] giá [PRICE 14 tỷ] [ATTR sổ hồng riêng].