# LLaVA-NeXTによる
# 商業施設向けAI案内技術


1. 場所の種類:
- 現在の写真は、公共施設の周りです。具体的な場所名は、見えませんが、学校や大
2. 写っている物体や構造物の詳細:

# LLaVA-NeXT（バージョン1.6）

- **視覚エンコーダ**と**大規模言語モデル**を組み合わせた**マルチモーダルモデル**
- 対応解像度：672×672、336×1344、1344×336の3つ
- 視覚的推論とOCR能力の向上

# 仕組み

## 画像と、プロンプトをAIに与える

```
'station': '駅構内の画像です。以下の項目を日本語で箇条書きで分析してください：\n・プラットフォーム番号、路線名、行
'airport': '空港内の画像です。以下の項目を日本語で箇条書きで分析してください：\n・ゲート番号、搭乗口、チェックイン
'mall': 'ショッピングモール内の画像です。以下の項目を日本語で箇条書きで分析してください：\n・店舗名、ブランド名、営
'general': 'この画像について以下の項目を日本語で箇条書きで詳しく分析してください：\n・場所の種類（屋内/屋外、建物の
```

## AIからの返答

```
1 1. 場所の種類: - 屋外 - 建物の種類: 学校 - 具体的な場所名: 大学キャンス 2. 写っている物体や構造物の詳細: - 人々が走り、カメラを持ち、動作を記録するような活動が進行中です。 3...
2 1. 場所の種類: - 現在の写真は、公共施設の近隣です。 - ビルとしても見えませんが、周りには学校や大学の建物が存在します。 2. 写っている物体や構造物の詳細: - 中央には、...
3 1. 場所の種類: - 現在の写真は、公共施設の近邊です。 - 建物の種類: 学校のビルドグラウンとして使われています。 - 具体的な場所名: "キャンス"という名前のビルドグラウンです。 2. 写っ...
4 1. 場所の種類: - 現在の写真は、公共施設の近隣です。 - ビルとしても見られますが、具体的な建物の種類や名称は不明です。 2. 写っている物体や構造物の詳細: - 中央には、人々が散りゆくような...
```

# AnyRes機構（動的高解像度画像処理）

- AnyRes機構（動的高解像度画像処理）により、画像を動的にタイル分割して処理
- バージョン 1.5 の336×336ピクセルから最大1344×336ピクセルまでの高解像度画像を扱える
- 利用シーン
  - 商業施設内の動画映像から、看板、標識、案内表示、商品情報を読み取る
  - 駅、空港、ショッピングモールなどの施設で撮影された映像を解析

# プロンプトの工夫例

[ステップ１] **初回フレーム**解析時に**画像内容から施設の種類を自動判定**

**画像内で施設に特徴的な文字を探す**

```
# 駅の特徴的なキーワード
if any(keyword in text for keyword in ['プラットフォーム', '路線', '時刻表', '改札', '乗り場',
    return 'station'
# 空港の特徴的なキーワード
elif any(keyword in text for keyword in ['ゲート', '搭乗', 'フライト', '出発', '到着', 'gate',
    return 'airport'
# ショッピングモールの特徴的なキーワード
elif any(keyword in text for keyword in ['店舗', 'ショップ', 'フロア', 'セール', 'レストラン',
```

[ステップ２] **施設タイプに応じて、専用の解析プロンプトを適用**する

```
'station': '駅構内の画像です。以下の項目を日本語で箇条書きで分析してください：\n・プラットフォーム番号、路線名、行
'airport': '空港内の画像です。以下の項目を日本語で箇条書きで分析してください：\n・ゲート番号、搭乗口、チェックイン
'mall': 'ショッピングモール内の画像です。以下の項目を日本語で箇条書きで分析してください：\n・店舗名、ブランド名、営
'general': 'この画像について以下の項目を日本語で箇条書きで詳しく分析してください：\n・場所の種類（屋内/屋外、建物の
```