**Human Activity Recognition**

[1] **SensorLM** (<https://arxiv.org/pdf/2506.09108v1>)

* Обучили новый энкодер для модальности сенсоров активности.  
    
  “The input to the sensor encoder is a matrix of shape [26 features × 1440 minutes], representing one-day windows of sensor data at minutely resolution. The sensor encoder is built using a ViT-2D backbone with a 2D patch size of (2, 10), producing 1872 tokens.”
* Извлекают 3 вида фичей из данных с сенсоров: Statistical, Structural, Semantic captions.
* Также обучили тект энкодер для извлеченных фичей (Statistical, Structural, Semantic captions).
* 3 Вида обучения CLIP, Cap, CoCa. CoCa Показал наилучший результат (contrastive loss + captioning loss).

[2] **ZARA: Zero-shot Motion Time-Series Analysis via Knowledge and Retrieval Driven LLM Agents** (<https://arxiv.org/pdf/2508.04038>)

* Комплексный пайплайн с RAG.
* Есть код.
* 2 Вида описания данных с сенсора (embedding + captions).
* Также как в [1] используют описания для данных с сенсоров (captions). Далее каким то образом (cross-validation, но непонятно на чём) вычисляется важность каждого кэпшена.
* Использует 3 LLM агента: Feature selector, Evidence pruning, Decision insight. (Одна и та же LLM).

[3] **Sensor2Text: Enabling Natural Language Interactions for Daily**

**Activity Tracking Using Wearable Sensors** (<https://arxiv.org/pdf/2410.20034>)

* Подход похож на [1].
* Немного устарела.
* Самый короткий related work)
* Video encoder в качестве учителя для обучаемой модели sensor encoder.
* Q-Former для Q&A на основе запроса и полученных sensor embeddings.

[4] **SensorLLM: Aligning Large Language Models with Motion Sensors for Human Activity Recognition** (<https://arxiv.org/pdf/2410.10624>)

* Есть код.
* Классический стетап с одним time series embedder и MLP, который проектирует его эмбеддинг в LLM.
* 2 Стадии обучения:  
  MLP projector (Sensor Embedding -> LLM tokens)
* Голова классификатора (Отдельный классификатор для определения активности поверх LLM)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Tabular LLMs**

[5] **TableGPT2: A Large Multimodal Model with Tabular Data Integration** (<https://arxiv.org/pdf/2411.02059>)

* Есть код
* Обучен table энкодер   
  (3 вида фичей: semantic table encoder, table language adapter, backbone llm)
* Свой претрейн и SFT
* Q-former для алайна эмбедингов таблицы
* Применяют sentence transformer к каждой ячейки таблицы
* Претрейн table encoder с помощью contrastive loss
* Описан пайплайн подготовки данных

[6] **TableGPT: Towards Unifying Tables, Nature Language and Commands into One GPT** (<https://arxiv.org/pdf/2307.08674>)

* Не тоже самое что [5], одинаковое название
* Свой Cascaded Table Encoder
* Пытаются сделать chain of thought

[7] **BRIDGING THE SEMANTIC GAP BETWEEN TEXT AND TABLE: A CASE STUDY ON NL2SQL** (<https://openreview.net/pdf?id=qmsX2R19p9>)

* Свой table encoder (получает эмбеддинги столбцов, а затем комбинирует их с текстовым описанием таблицы) (prompt template с эмбеддингами)
* Есть код
* Также как [5] кодирует каждую ячейку по отдельности
* Обучение в 3 стадии: Encoder Pre-training (претрейн table encoder со своим contrastive loss), Feature Alignment (table encoder + adapter), Instruction Tuning (adapter + llm)
* Cвежая ICLR 2025
* Описан пайплайн претрейна table encoder

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Обзоры**

[8] **Towards LLM-Centric Multimodal Fusion: A Survey on Integration Strategies and Techniques** (<https://arxiv.org/pdf/2506.04788v1>) не открывается без впн

* Свежий обзор базовых fusion методов