ActionSense

Interim Presentation – Extracting Action Items from Unstructured & Noisy Meeting Data

Large Language Models and its Applications

Meir Weinberg & Yael Reina

Project Review

The Problem

- State-of-the-Art models excel in formal English but struggle with Code-Switching (Hebrew-English mixing) at Extracting Action Items
- "Israeli Meeting" characteristics: High noise, slang, and rapid language shifts lead to hallucinations and missed tasks

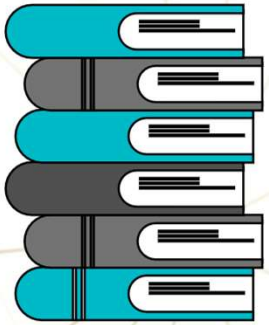
Our Goal

- Develop a robust pipeline for extracting structured Action Items (Who, What, When) from noisy, bi-lingual transcripts

Key Novelty

- Data: Creation of the synthetic dataset for Hebrew-English organizational dialogue
- Method: Fine-tuning open-source LLMs (Llama) specifically for this linguistic niche

Previous work



Title / Year	Task	Methods	Data	Results	Relation to Project
Detecting Actionable Items (2016)	Action Item Detection	**CDSSM** (A convolutional deep structured semantic model) & SVM	**ICSI Meetings:** Human-transcribed, English datasets	**AUC Score:** • **~47%** (Mismatch) • **~69%** (Match)	1. **Historical Baseline**: Establishes the shift from keywords to Deep Learning. 2. **Motivation:** Proves that complex models fail without matched data (47%), justifying our Synthetic Data approach
Action-Item-Driven Summarization (2024)	Generating summaries focused on action items	LLM Prompting & Chain-of-Thought (CoT). Includes Topic-Based Segmentation for long context.	**QMSum / DialogSum:** Standard English meeting datasets	Achieved ROUGE-1 score of **34.2** on AMI dataset, outperforming baselines (BART) by **~12%**	Demonstrates **SOTA LLM capability**, but limited to **English only** (No code-switching)
On LLMs-Driven Synthetic Data Generation (2024)	Survey of synthetic data generation	**Gen & Filter:** Review of techniques to create and clean data (e.g., "LLM-as-a-Judge").	Review of **multiple datasets** across various NLP tasks	Quality over Quantity: Shows that curated synthetic data solves data scarcity effectively.	Provides the **methodological validation** for our Simulator approach (generating Hebrew-English data)

Dataset

Dataset Description

We generated synthetic corpus of Israeli tech meetings ("Heblish")

- Derived from the AMI Meeting Corpus
- **Adaptation:** Extracted a dedicated Task Bank from AMI and performed a two-step transformation:

1. **Translation:** Converting core tasks and technical discussions into Hebrew.
2. injecting high-tech industry slang (e.g., *POC, API Endpoints, UI/UX, Backend*) to reflect authentic Israeli office dynamics

Dataset

Data Generation

- **Technique:** Synthetic Data Generation via **GPT-4o**.
- **Process:** The model was fed seed tasks from the translated AMI bank. it was prompted to generate natural, long-form Israeli office dialogues where tasks are scattered realistically among "Small Talk" and technical syncs.
- **Labeling:** Automated extraction into structured objects containing Assignee, Action, and Deadline.

Exploratory Data Analysis

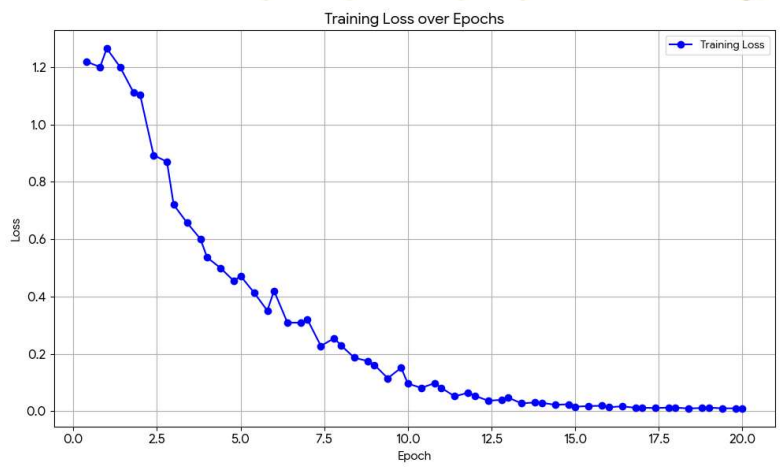
- Realistic imbalance; the majority of utterances are context/discussion, with ~10-15% identified as specific Action Items.
- Variable (15–50 lines per session) to simulate different meeting types (Daily Sync vs. Deep Dive)
- High presence of temporal expressions for deadlines (e.g., "End of the week," "By the 15th")

Example

speaker	utterance
דנה	היי רועי, אנחנו צריכים לדבר על הפרויקט החדש ולתכנן את לוחות הזמנים.
רועי	נכון, יש לנו כמה דברים לסגור. מה לגבי הכלים? חשבת על לוחות לבניה לפעילויות בניית צוות?
דנה	עדיין לא, אבל אני יכולה לחקור את הכלים הזמינים ולראות מה מתאים לנו.
רועי	מעולה, אולי כדאי לבדוק גם את ה-API של הכלים השונים כדי להבין את ההתאמה.
דנה	צודק, אני אבדוק את זה.
רועי	ומה עם ה-UX של המערכת? דיברנו על לעשות A/B Testing.
דנה	נכון, צריך לקבוע תאריך לבדיקה.
רועי	אני יכול לקחת את זה על עצמי.
דנה	אוקיי, אז בוא נקבע תאריך לשבוע הבא.
רועי	מצוין, אני אשלח לך זימון ליומן.
דנה	ומה עם הדאטה אנליטיקס? צריך לוודא שאנחנו עוקבים אחר כל המדדים.
רועי	נכון, אני אעבור על זה עם המחלקה ואראה מה חסר לנו.
דנה	יש לך עדכון לגבי ה-Backend?
רועי	כן, אני עובד על ה-API Endpoints, זה ייקח עוד יומיים.
דנה	בסדר גמור, תעדכן אותי כשזה מוכן.
רועי	בהחלט. האם יש עוד משהו שצריך לסגור?


```
{
  "action_items": [
    {
      "assignee": "דנה",
      "action": "לחקור כלים זמינים כמו לוחות לבניה לפעילויות בניית צוות",
      "deadline": "שבוע הבא"
    },
    {
      "assignee": "דנה",
      "action": "לבדוק את ה-API של הכלים השונים",
      "deadline": "שבוע הבא"
    },
    {
      "assignee": "רועי",
      "action": "לקבוע תאריך ל-A/B Testing",
      "deadline": "שבוע הבא"
    },
    {
      "assignee": "רועי",
      "action": "לעבור על הדאטה אנליטיקס עם המחלקה",
      "deadline": "שבוע הבא"
    },
    {
      "assignee": "רועי",
      "action": "לסיים את העבודה על ה-API Endpoints",
      "deadline": "בעוד יומיים"
    }
  ]
}
```

Baseline Solution & optimization



The Baseline Approach

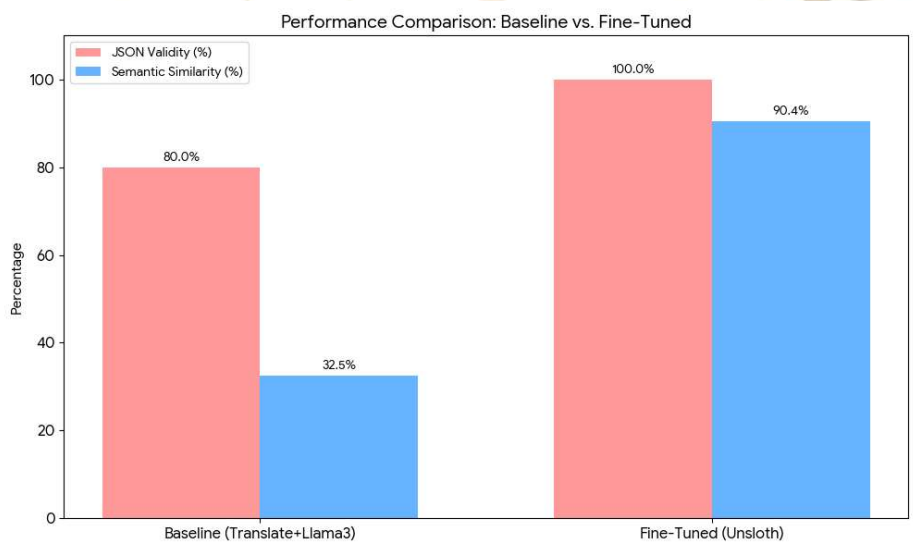
- **Method**: Zero-shot llama-3-8b-bnb-4bit + Hebrew-to-English Translation
- **Outcome:** Failed to capture linguistic nuances
- **Semantic Similarity:** Only 32.5% (Vectors were misaligned due to translation errors)
- **JSON Validity**: 80% (Model frequently hallucinated keys or broke syntax)

The Optimization (Our Solution)

- **Method:** Fine-Tuning with Unsloth (QLoRA) on synthetic Hebrew data
- **Challenge:** The model suffered from "Catastrophic Forgetting" (lost JSON formatting skills)
- **Fix:** Applied One-Shot Schema Forcing during inference

Result

Achieved SOTA performance on our dataset

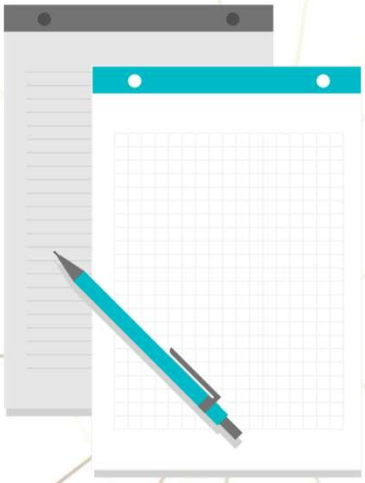


Significant improvement in both stability and understanding after Fine-Tuning

- **JSON Validity: 100%** (Perfect structural stability)
- **Semantic Similarity: 90.4%** (High accuracy in extracting Action Items)

Expected	{"action_items": [{"assignee": "יוסי", "action": "לעבור על תיקיית מסמכי הפרויקט ולתמצת", "deadline": "מחר בצהריים"}, {"assignee": "רונית", "action": "לקבוע עדיפויות להגשות אבסטרקט", "deadline": "סוף השבוע"}, {"assignee": "דני", "action": "להכין סקיצה ראשונית של ה-UI", "deadline": "יום ראשון"}, {"assignee": "יוסי", "action": "להעריך את האפשרות ליישום מערכת חיישנים לאיתור השלט", "deadline": "יום שני"}]}
Before Fine-tuning	{"action_items": [{"assignee": "ו", "action": "לעיין במסמכי הפרויקט"}, {"assignee": "רונית", "action": "לתת עדיפות להגשות מופשטות"}]}
After Fine-tuning	{"action_items": [{ "assignee": "יוסי", "action": "לעבור על מסמכים ולתרגם לנוח למשתמש", "deadline": "מחר בצהריים" }, { "assignee": "רונית", "action": "להגש עבודה על הגדרת אינטראקציות ומחקר על מחירים", "deadline": "סוף השבוע" }, { "assignee": "דני", "action": "ליצור סקיצה ראשונית לעיצוב המראה וההסתכים", "deadline": "יום ראשון הבא" }, { "assignee": "יוסי", "action": "לבדוק את האפשרות ליישום מערכת חיישנים לאיתור שלט", "deadline": "יום שני הבא" }]}

Plan



Phase	Task	Expected Outcome
1. Optimization	**Hyperparameter Tuning & Regularization**	Experiment with different LoRA ranks and dropout rates to ensure model stability (avoiding overfitting).
2. Scale	**Data Augmentation**	Generate ~100 additional synthetic samples to validate robustness on a larger test set.
3. Validation	**Qualitative Error Analysis**	Inspect edge cases (e.g., extreme slang, empty lists) and refine the prompt strategies accordingly.
4. Submission	**Final Repo & Documentation**	Update repository with clean code, datasets, Visual Abstract, and a detailed README.
5. Closing	**Final Presentation**	Prepare the final presentation and demo for the end-of-semester class.