ActionSense

Final Presentation **– Extracting Action Items from Unstructured & Noisy Meeting Data**

Large Language Models and its Applications

Meir Weinberg & Yael Reina

Review and project definition

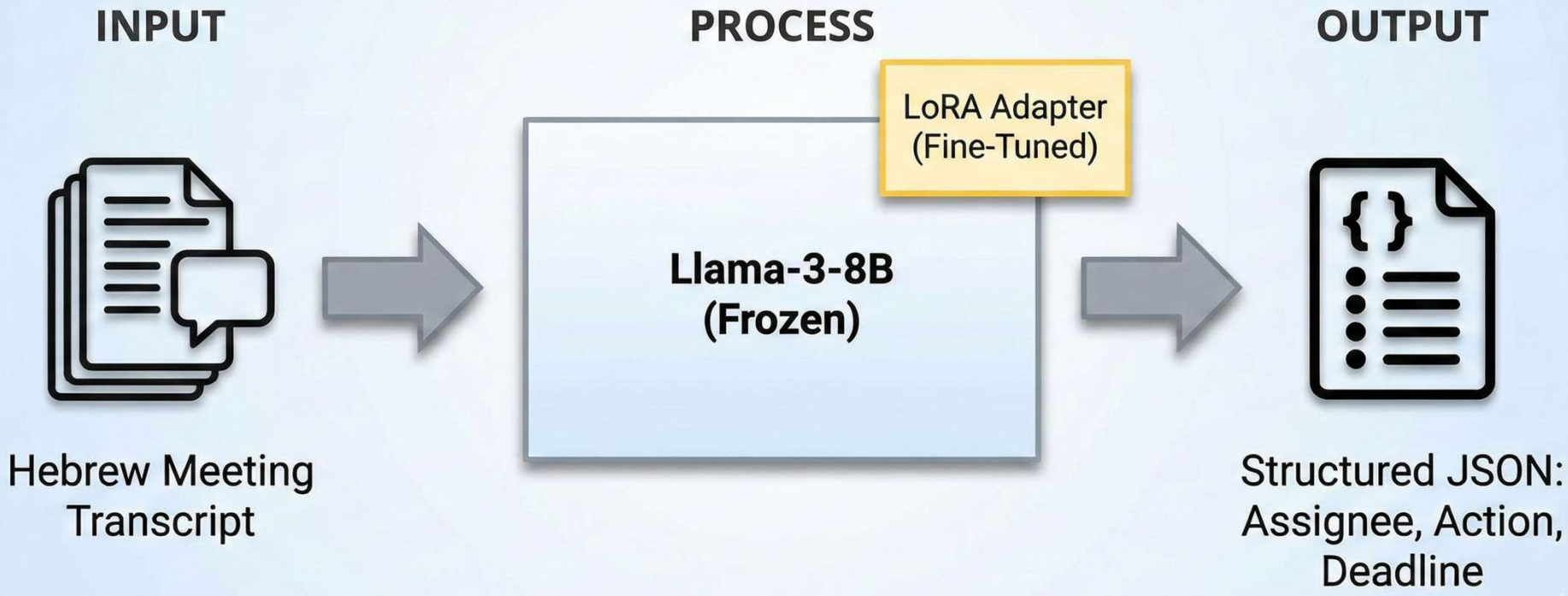
The problem

- **The "Hebrew-English" Gap:** . State-of-the-Art models excel in formal English but cannot accurately extract action items when team members switch languages mid-sentence.
- **Cultural Context:** The informal nature of Israeli high-tech meetings creates "noise" (slang and rapid language shifts) that confuses standard LLMs, resulting in missed tasks and fabricated deadlines.

Our Goal

- **Data:** Creation of the synthetic dataset for Hebrew-English organizational dialogue
- **Automated Action:** Developing a robust pipeline for extracting structured Action Items (Who, What, When) from noisy and bi-lingual Israeli high-tech meetings

FLOWCHART


INPUT
PROCESS
OUTPUT
LoRA Adapter
(Fine-Tuned)
Llama-3-8B
(Frozen)
Hebrew Meeting
Transcript
Structured JSON:
Assignee, Action,
Deadline

Project achievements and novelty

What we achieved?

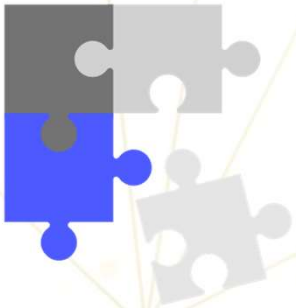
Data:

- **Novel Data Source**: Developed a unique dataset tailored to the linguistic patterns of Israeli tech hubs
- **Data quality**: 230 High-quality synthetic Data Generation using LLMs

Model and Evaluation:

- **From 1% to ~60%**: Successfully fine-tuned llama-3-8b-bnb-4bit model to become a specialized task extractor
- **Robust Evaluation Pipeline**: a custom pipeline using the **Hungarian Algorithm** for list matching and **AlephBERT** for semantic similarity in Hebrew.

methodology



Data

Data preparation:

We created the dataset using the following steps:

1. creating a Task Bank from the AMI dataset
2. translating the Task bank to “Hibrish” using gpt-3.5-turbo
3. Generating synthetic Israeli high-tech meetings using the tasks from the Task bank using gpt-4o

The result:

A synthetic dataset containing 230 high-tech meeting with : 15–50 lines and 1-4 tasks and High presence of temporal expressions for deadlines (e.g., "End of the week," "By the 15th")

Data 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?

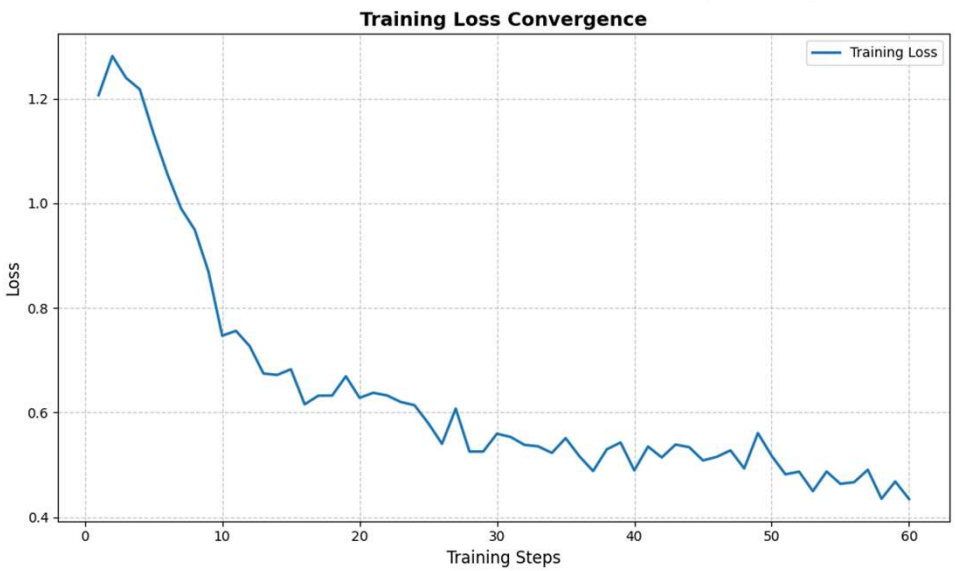

```
{
  "action_items": [
    {
      "assignee": "דנה",
      "action": "לחקור כלים זמינים כמו לוחות לבניה לפעילויות בניית צוות",
      "deadline": "שבוע הבא"
    },
    {
      "assignee": "דנה",
      "action": "לבדוק את ה-API של הכלים השונים",
      "deadline": "שבוע הבא"
    },
    {
      "assignee": "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"
    }
  ]
}
```


The order of the tasks in the JSON files is the same order of the tasks in the meetings

methodology

Training

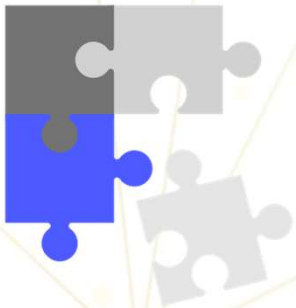
- **Model & Training**
  - Base: llama-3-8b-bnb-4bit (Quantized).
  - Technique: LoRA, dropout, learning rate, "Train-as-you-test" prompt strategy.



methodology

Evaluation

- **Evaluation Metrics (The Challenge):**
  - Since the order of items matter, we used Hungarian Algo to match predicted tasks to ground truth, temperature for determinism
  - **Assignee:** Fuzzy Matching (Ratio > 80).
  - **Action:** Semantic Similarity using AlephBERT embeddings.
  - **Deadline:** Soft string matching (Ratio > 85).



סכום הציונים של השידוכים

מקסימום(כמות חזויה, כמות אמיתית)

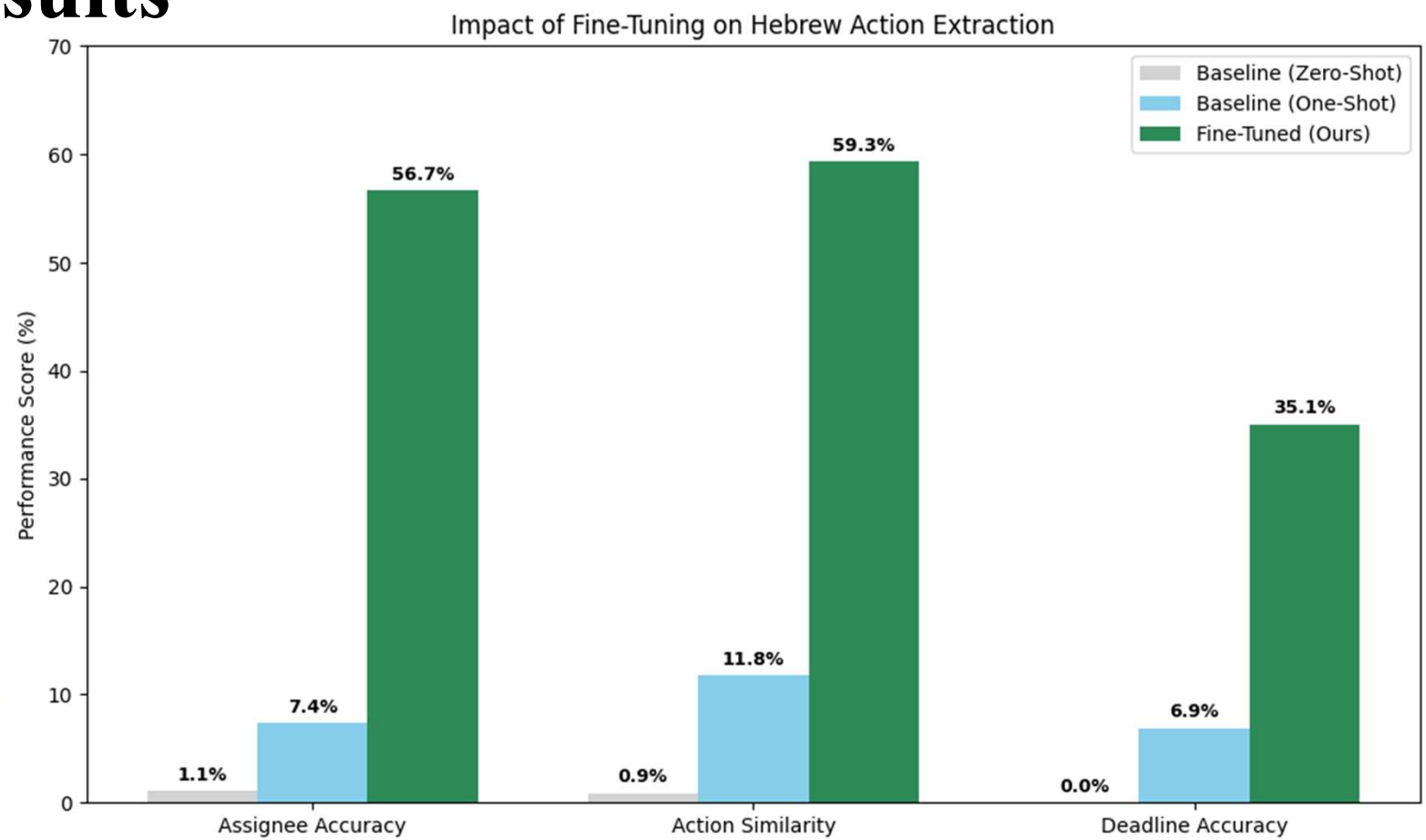
Results

Results: dramatic Improvement over Baseline

Quantitative Results:

- **Baseline (Zero-Shot):** FAILED (~1% Accuracy). Output was unstructured text/English.
- **Baseline (One-Shot):** ~11.5%. Improved structure but struggled with Hebrew content.
- **Fine-Tuned:**
  - **Assignee: 56.7%** (High precision in entity extraction).
  - **Action: 59.3%** (Strong semantic understanding).
  - **Deadline: 35.1%** (Challenging due to relative dates like "next week").

Results


Impact of Fine-Tuning on Hebrew Action Extraction
Performance Score (%)
0
10
20
30
40
50
60
70
Baseline (Zero-Shot)
Baseline (One-Shot)
Fine-Tuned (Ours)
1.1%
7.4%
56.7%
0.9%
11.8%
59.3%
0.0%
6.9%
35.1%
Assignee Accuracy
Action Similarity
Deadline Accuracy

speaker	utterance
דני	חבר'ה, בוקר טוב. אנחנו צריכים להעריך את הביקוש השוקי לשלט הרחוק החדש שלנו.
נועה	נכון, השלט צריך להיות ידידותי למשתמש ומשובח מבחינת ה-UI וה-UX.
יוסי	אני יכול להתחיל לנתח את הנתונים הקיימים ולראות איך הם משתלבים בשוק הנוכחי.
דני	מעולה, יוסי. תוכל לסיים את זה עד סוף השבוע?
יוסי	כן, אני אעבוד על זה ואעדכן אתכם ביום חמישי.
נועה	אנחנו גם צריכים לתאם את הגעתה של פאולה בשבוע הבא. צריך להקצות לה חדר.
דני	נכון, אני אבדוק עם מחלקת הלוגיסטיקה אם יש חדר פנוי.
יוסי	פאולה מגיעה ביום שלישי, נכון? כדאי שזה יהיה מוכן מראש.
דני	כן, אני אעשה את זה עד יום שני.
נועה	טוב, אז יש לנו את כל המשימות לחלק הזה. יש עוד משהו שצריך לדון בו?
יוסי	לא כרגע, אבל אני אעדכן אתכם כשיהיו תובנות מהניתוח.
נועה	בסדר גמור. תודה לכולם, ניפגש שוב מחר.
דני	תודה נועה, יוסי, תעדכן אותי על כל התקדמות.
יוסי	בהחלט, אני אשלח עדכון במייל.
נועה	יוסי, אם אתה צריך עזרה בניתוח, אני כאן.
יוסי	תודה נועה, אני אעדכן אם אצטרך עזרה.
דני	טוב, אז אנחנו מסיימים להיום. שיהיה לכם יום נפלא!
נועה	יום טוב לכולם!
יוסי	נתראה מחר!
דני	להתראות.
נועה	ביי!

Expected	[{"assignee": "יוסי", "action": "לנתח את הנתונים הקיימים להערכת הביקוש השוקי לשלט", "deadline": "חמישי"}, {"assignee": "דני", "action": "לתאם עם מחלקת הלוגיסטיקה להקצאת חדר לפאולה", "deadline": "שני"}]
Baseline zero-shot	[]
Baseline one-shot	[{"assignee": "יוסי", "action": "להעריך את הביקוש השוקי לשלט הרחוק החדש שלנו", "deadline": "סוף השבוע"}, {"assignee": "יוסי", "action": "לסיים את זה עד סוף השבוע", "deadline": "יום חמישי"}, {"assignee": "דני", "action": "לבדוק עם מחלקת הלוגיסטיקה אם יש חדר פנוי", "deadline": "יום שני"}, {"assignee": "יוסי", "action": "לעשות את זה עד יום שני", "deadline": "יום שני"}, {"assignee": "יוסי", "action": "לעדכן את כולם על כל התקדמות", "deadline": "יום שני"}, {"assignee": "יוסי", "action": "לשלח עדכון במייל", "deadline": "יום שני"}, {"assignee": "יוסי", "action": "לעדכן אם צריך עזרה", "deadline": "יום שני"}]
After Fine-tuning	[{"assignee": "יוסי", "action": "נתח את הנתונים הקיימים וראה איך הם משתלבים בשוק הנוכחי", "deadline": "יום חמישי"}, {"assignee": "דני", "action": "תאם עם מחלקת הלוגיסטיקה אם יש חדר פנוי לפאולה", "deadline": "יום שני"}]

Conclusion & Future Work

Did we achieve the goal?

- Yes. We transformed a model that couldn't handle the task into a reliable extractor for Hebrew action items.

Lessons Learned:

- **"Train as you test":** Consistency between training prompts and inference prompts is critical.
- **Metric Design:** Evaluating generative lists requires semantic matching algorithms, not just string comparison.

- **Future Work:**
  - **Dataset Expansion:** 230 samples is small. Increasing to 1000+ would likely solve the Deadline accuracy issues.
  - **Date Normalization:** Integrating a dedicated date-parser to convert "next Sunday" to YYYY-MM-DD.