# Brief – IsraPolitics Project

## Project Name:

IsraPolitics

## Group members:

1. Avraham Sommer [Avraham.Sommer@mail.huji.ac.il](mailto:Avraham.Sommer@mail.huji.ac.il)
2. Betzalel Yanir [Betzalel.Yanir@mail.huji.ac.il](mailto:Betzalel.Yanir@mail.huji.ac.il)
3. Oryan Yosef Hassidim [Oryan.Hassidim@mail.huji.ac.il](mailto:Oryan.Hassidim@mail.huji.ac.il)

## Advisors:

1. Effi Levi [EfLe@cs.huji.ac.il](mailto:EfLe@cs.huji.ac.il)
2. Shaul Shenhav [Shaul.Shenhav@mail.huji.ac.il](mailto:Shaul.Shenhav@mail.huji.ac.il)

# Problem Definition

In politics, voters and generally all citizens, want to choose their politicians wisely. The data and amount of information are vast, causing voters not to know what the politicians real opinion.

At the end of this project, we want to make it easy for anyone to access the relevant information, and in ordered shape, so they can make their own choices, according to their opinions.

# Background and Related Work

## Background

Because our project is interdisciplinary, of both CS, and Political science, we will need background knowledge of both subjects.

Regarding CS, we need to learn about NLP tools in general, and specific tools for Hebrew NLP. Generally, the primary tools we assume we will use are BERT, and NLI.

Secondly, we need to analyze key issues in Israel's politics and diagnose the different connotations of those issues on each side of the political map.

## Related Work

1. Our project is based on and continues of the work of Guy Mor-Lan, Effi Levi, Tamir Sheafer, and Shaul R. Shenhav. 2024. [IsraParlTweet: The Israeli Parliamentary and Twitter Resource](https://aclanthology.org/2024.lrec-main.819).

They took and indexed all the data and started a first analyzation of the data, and we want to take it few steps further.

* Paper: [IsraParlTweet: The Israeli Parliamentary and Twitter Resource - ACL Anthology](https://aclanthology.org/2024.lrec-main.819/)
* Dataset: [guymorlan/IsraParlTweet at main on GitHub](https://huggingface.co/datasets/guymorlan/IsraParlTweet/tree/main)

1. GitHub page containing existing NLP tools in Hebrew - [GitHub - NNLP-IL/NNLP-IL: A national initiative for the creation of infrastructure, research and development of advanced capabilities for the advancement of the field of NLP in Hebrew and Arabic.](https://github.com/NNLP-IL/NNLP-IL)

Regarding political science, we will use:

1. <https://www.idi.org.il/parliaments/22242/22263>, an article discussing the main conflicts in Israel's politics.

# Proposed Solution

We can approach this problem in the orthodox way – pay for a group of experts and researchers for analyze the data and make a static dataset with the properties of each parliament member, and above that make statistics.

Our direction is to automate this and use the tools and capabilities of ML for achieve this goal, and to open the door for taking it few steps further and make an option for finding insights from the data on the fly, or even while chatting with a human.

## Solution proposal:

we chose to use NLP tools to organize and process the information, because the data is vast, and it will be more efficient. Also, in the future, using NLP is more changeable and updateable.

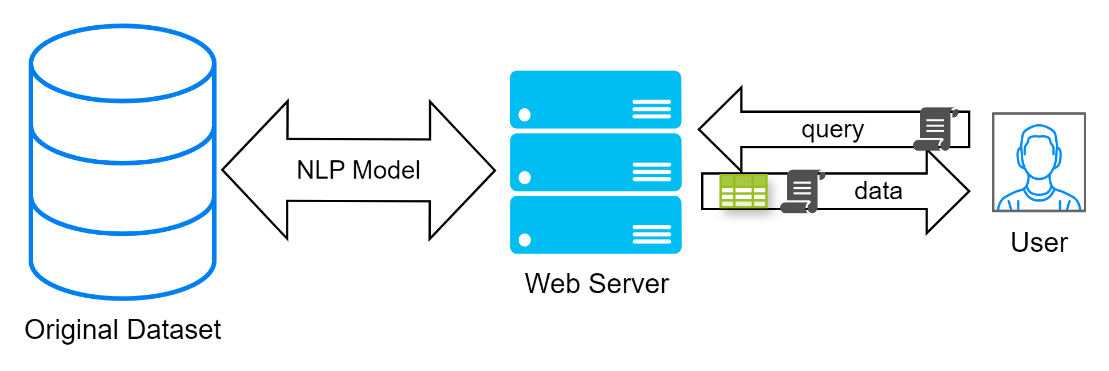
## Key components:

1. Dataset, we already have it, raw information.
2. Main algorithm – we will need to choose the most suitable NLP tool, and analyze all our dataset with it.
3. Properties dataset – made by the main algorithm and contain static analyzed information about parliament members, parties, main political issues.
4. User interface that will allow people to use our project, and datasets.

תמונה שמכילה טקסט, צילום מסך, גופן, עיגול

התיאור נוצר באופן אוטומטי

First level



Second level

## Solution feasibility:

During our research on this subject, we found a similar project in other languages. So, the idea is possible, and the main challenge will be the choosing of NLP tool that will work in the Hebrew language.

## Technologies:

For our model, we need to check few directions and look for the best solution – we can use one of them or some combination of them. Each one will be need a fine tuning for Hebrew and for our field. On this moment our ideas are:

1. BERT – some NLP engine for transformation of a text into a matching vector, that represents the text such that two input texts with semantic proximity will be represented by the engine as relatively close vectors.
2. NLI models – gets *hypothesis* and *text* and returns one of the options: *entailment*, *contradiction* or *neutral*.
3. LLM engines (Large Language Model, like GPT) with fine tuning for our needs.

For the GUI, the first level will be MS Excel that brings awesome tools for analyzing and displaying our data and insights. In the second level, we will use some static webapp hosted on GitHub pages for interacting with the data. In the third level we will update our app for interact and conclude new insights from the original dataset by the model.

## Challenges and uncertainties:

Our main challenge, as said before, is creating suitable NLP algorithm, that will be good for Hebrew. In the case will not find perfect one, we will use the best we will find.

## Required knowledge:

1. Data processing and visualization tools.
2. SQL.
3. Natural Language Processing Huji course, 67658.
4. NLP theoretics.
5. NLP practices and tools.
6. Specific NLP tools for Hebrew.
7. Web development.
8. Main conflicts in Israel.

# Evaluation and Verification Scheme

We divide the evaluation to two parts:

First, evaluation of the algorithm. We will need to create test cases manually (some parliament members, and issues) and test our algorithm accordingly. We will need to create a distance function that is relevant.

Secondly, we will need to evaluate the accessibility of our project to the user. If the user can't easily use it for his decision, we actually failed in our goal. So we will need to bring some voters that are not involved so much, let them use our interface, and let them rate us how much it helped them.

# Annual work plan

## First steps:

1. Preprocessing our data and create easy way to interact with it.
2. Write sentences which describes the key stands.
3. Create test case manually so we will see what properties are feasible.
4. Research with our advisor NLP tools and choose the most relevant to our properties.
5. Start playing with the algorithms and look the results.

## Minimum viable product:

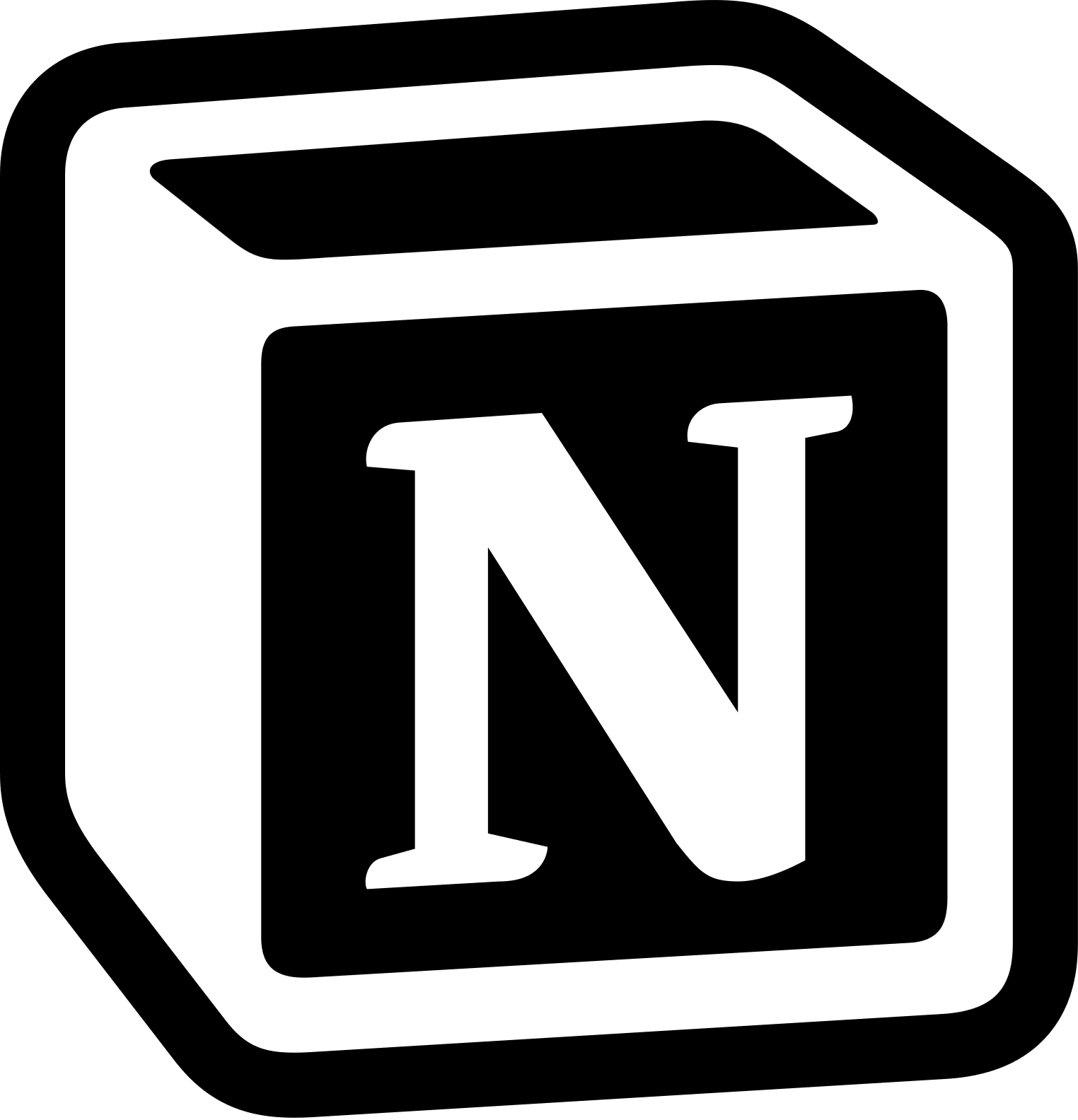
There is an interactive and accessible Excel file. The user can choose a parliament member and see his opinions on some key fields in Israel's politics. He can also choose subject from a list, and see each member's or parties' opinion on the matter.

In the excel file the user fills some fields of form and see the most relevant members and parties for him.

## Future steps:

1. Creating webapp interface.
2. The user can choose a subject of his own choice (which doesn’t exist yet), and ask for opinions about it. The system now accessing the source database, and creating new property in the properties database about this subject.

# Project Management

 Notion: [IsraPolitics Tasks at Notion](https://accidental-aardwolf-e38.notion.site/1b9dc50d0c264b4ca003b96a9f233634?v=8cfb5787dcde42d28fb9d416995b94fa&pvs=4)

 GitHub: [IsraPolitics at GitHub](https://github.com/Oryan-Hassidim/IsraPolitics)