**סיכום פגישה**

**מספר פרויקט: 600 - Text Complexity and Readability**

תאריך פגישה: 30.12.24

משתתפים: תומר סורוז'ון, יפעת גסטפריינד, רויטל יוסופוב, גיא פרבשטיין

(נעדרו: אורי בראל)

מנחה: ד"ר אור חיים אניג'דר

**דיווח ביצוע משימות**

| **#** | **Category** | **Task** | **Assign to** | **Due date** | **Status** |
| --- | --- | --- | --- | --- | --- |
| 1.1 | Data Collection & Preprocessing | Download publicly available versions of the U.S. Tax Code for the years 2017 and 2018 in PDFs | Tomer | 08/12/2024 | DONE |
| 1.2 | Develop scripts for data and features extraction from the Tax Codes (2017 + 2018) | Yifat | 22/12/2024 | DONE |
| 1.3 | Clean and structure the data into CSV format | Yifat | 01/01/2025 | DONE |
| 1.4 | Review and quality control of the extracted data | Tomer | 02/01/2025 | IN PROGRESS |
| 2.1 | Project Definition | Writing a full characterization & definition of the project, according to final project design template | Tomer+ Yifat | 01/01/2025 | DONE |
| 3.1 | Model Building & Integration | Headings and titles analysis - develop scripts to identifying semantic similarity using LaBSE and lexical similarity using Jaccard Score and Levenshtein distance. | Tomer+ Yifat | 22/12/2024 | DONE |
| 3.2 | Model and Algorithm extension - comparative analysis to examine similarity (semantic and lyrical) between smaller units of information, in a sample of cases where the titles were found to be similar. | Tomer+ Yifat | 05/01/2025 | IN PROGRESS |
| 3.3 | Examining common words and phrases, between the code before/after the reform (using N-Grams) | Revital | 05/01/2025 | DONE |
| 3.4 | Examining statistical indices of common words/phrases (such as: median, average, etc.), and Visual presentation (Histogram and other relevant graphs) | Ori | 05/01/2025 | IN PROGRESS |
| 3.5 | Model and Algorithm extension - Expanding semantic similarity and lexical similarity, considering headings and content | Tomer+ Yifat | 19/01/2025 | TO DO |
| 3.6 | Model and Algorithm extension - analyzing structural analysis | Tomer+ Yifat | 19/01/2025 | TO DO |
| 3.7 | Model Building & Integration | Compute number of words, number of headings and subsections and number of paragraphs with similar titles | Revital | 29/01/2025 | TO DO |
| 3.8 | Develop scripts for hierarchical parsing and cross-referencing | Tomer+ Yifat | 29/01/2025 | TO DO |
| 4.1 | Dashboards & Reporting Development | Build initial interactive dashboards for visualization (presented to end users\legal and technical stakeholder) | Ori+ Guy | 29/01/2025 | TO DO |
| 4.2 | Implement customizable reporting features | Ori+ Guy | 29/01/2025 | TO DO |
| 4.3 | Fine-tuning of the dashboards and customizable reporting features according to legal and technical stakeholder's requirements | Tomer+ Yifat | 06/02/2025 | TO DO |
| 5.1 | Project documentation | Writing a detailed report (10 pages) explaining the project process, decisions made and evaluation | Tomer+ Yifat+ Revital | 21/01/2025 | TO DO |
| 5.2 | Review, quality control and proofreading of the project report | Tomer+ Yifat+ Revital | 26/01/2025 | TO DO |
| 5.3 | Writing relevant sections in the article intended for publication, addressing the analysis of tax code complexity using NLP models | Tomer+ Yifat+ Revital | 27/02/2025 | TO DO |
| 5.4 | Review, quality control and proofreading of the article | Tomer+ Yifat+ Revital | 01/03/2025 | TO DO |
| 6.1 | Evaluation and Testing | Validate results and gather feedback from legal and technical stakeholders | Team | Untill 15/03/2025 | TO DO |

דגשים מהמנחה:

* הצגת המשימות תתבצע בטבלה שתפרט את המשימות, המבצע ולוחות הזמנים (כמוצג לעיל).
* הצגה בפגישות תתבצע באמצעות מצגת.