30.5.2024

שלום לאבנר, נועם ואנשי האיגוד הישראלי לטכנולוגיות שפת אנוש,

כ-CTO בחברת מלינגו, השותפה באיגוד, רציתי להביע את הערכתי לעבודת האיגוד ולחומרים המשובחים שיוצאים בחסות האיגוד לטובת קידום נושא טכנולוגיית השפה בעברית וכן בערבית.

מלינגו עושה שימוש בחומרים אלה לדיוק טכנולגיית השפה שלה. לחומרים מתויגים היטב ע"י אדם ומאומתים חשיבות קריטית הן באימון והן במדידה של מודלי שפה.

בעידן שבו הבינה המלאכותית מתקדמת בצעדים גדולים גם בתחום הבנה ויצירת שפה טבעית, חשיבות היכולת למדוד ולדייק את התוצאות עולה עוד יותר.

בעידן זה יכולת ההבנה והיצירה מודלי שפה הגדולים הרב-לשוניים (LLM), הולכת ועולה, כולל בשפות כמו עברית. יש לשים דגש על כך שיצירת מודלים מדויקים ומבימנים דורשת שילוב של אימון unsupervised – על כמות גדולה מאוד של טקסטים, ביחד עם כמויות גדולות ככל האפשר של טקסט מתויג, הן לצורך fine tuning והן למדידה ובדיקה. ללא מדידה, המפתחים מגששים באפילה – וכאן בולטת חשיבות החומר המתויג.

כמו כן, נושאים כמו זיהוי ישויות, מבנים תחביריים, קשרים ביו ישויות, דיוקי ניתוח מורפולוגי ולמטיזציה, הם רכיבים קריטיים בפתרונות שדרושים ללקוחות ששואפים לייעל ולחדש את ממשקי השפה הטבעית שלהם עם המשתמשים, בין אם בטקסט ובין אם בקול, וחומרי האיגוד מאפשרים, וימשיכו לאפשר עוד יותר כאשר ימשיכו להתרחב, שיפור ודיוק ממשקי שפה אלה, ויתרמו להבאתם לרמת המוצרים המקבילים בשפות המובילות בעולם, אתגר שהוא חשוב הן לאור המיעוט היחסי של דוברים וחומרים קיימים בעברית והן לאור המורכבות המובנית של השפה העברית.

כמו כן לגבי ערבית בכלל וערבית שבשימוש בישראל בפרט, חומרי האיגוד חשובים ויעזרו ללא ספק בקידום פתרונות גם בשפה זו. הערבית המדוברת, ובפרט הערבית המדוברת בישראל היא שפה אשר יש עבורה משאבי NLP מעטים מדיי באופן בולט, ועבודת האיגוד בתחום זה היא חלוצית וחשובה ביותר.

בברכה,

יוני נאמן

מייסד ו-CTO

מלינגו בע"מ

03-6070400 www.melingo.com
16 Tozeret Ha'aretz st. Tel Aviv 67891, Israel