

ברוכים הבאים לסמינר על יעילות בעיבוד שפה טבעית!








הסמינר הזה יעבור בצורה שונה מסמינרים רגילים, אז בבקשה **קראו את ההסבר שלהלן** לפני ההרשמה לקורס.

## **מבוא**

אחד החסרונות הבולטים בסמינרים הם שקשה ליצור דיון משמעותי ורוב המשתתפים מוצאים את עצמם משקיעים הרבה מאוד בשבוע אחד (השבוע של ההצגה) ומעט מאוד בשאר הזמן. זה הופך את הלמידה מהסמינר לחוויה חד-גונית וממוקדת. הסמינר הזה שואב השראה מיוזמות של קבוצות **באוניברסיטת צפון קרוליינה ואוניברסיטת טורונטו**, שהציעו פורמט של **משחק תפקידים** בסמינרים. הרעיון הוא שכל שבוע במקום שסטודנט/ית אחד/ת יציגו מאמר, קבוצה של סטודנטים יציגו אספקטים שלו. למשל, סטודנט אחד ייתן סיכום של הרעיונות המרכזיים במאמר; סטודנטית אחרת תיתן סקירה היסטורית של הרקע שהוביל לכתיבת המאמר, ושל החידושים המרכזיים שיצאו בעקבותיו; סטודנטים אחרים יכתבו ביקורת (peer review) על המאמר, וכו'.

השאיפה היא שחווית הלמידה תהיה מוצלחת ומהנה, ושכל תלמיד ותלמידה יצאו מהסמינר הזה גם עם הבנה טובה של הנושאים המרכזיים בתחום של ייעול עיבוד שפה טבעית, וגם יתנסו בסט כישורים רלוונטיים למחקר.

## **התפקידים השונים**

-  **מבקרים:** המבקרים יכתבו ביקורת (peer review) על המאמר, משל הוא היה מועמד לקבלה בכנס. המבקרים יכתבו ביקורת מנומקת על החוזקות, החולשות, הסיבות לקבלה ולדחייה של המאמר, ולבסוף המלצה האם לקבל את המאמר. הביקורת צריכה להיות פורמט של ביקורת ב-ACL, הכנס המוביל ב-NLP, לפי **התבנית הבאה**. כל אחת מהמבקרות תציג בקצרה את הביקורת לכיתה, ויתקיים דיון בין המבקרים השונים (בהשתתפות שאר התלמידים).
-  **מסכם:** המסכם יכתוב סיכום של עמוד של המאמר, ויציג אותו לכיתה בכמה שקפים.
-  **ארכיאולוגית:** הארכיאולוגית תיתן לנו קונטקסט היסטורי ועתידי על המאמר. היא תספר לנו קצת מה היה לפני המאמר, איפה המאמר הזה יושב בהקשר של העבודה הקודמת, ותסקור בקצרה מאמר אחד שעליו מתבסס המאמר הנוכחי. היא גם תציג לנו מה קרה מאז שהמאמר פורסם, מה החידושים העיקריים שיצאו בעקבותיו, ותציג בקצרה אחד את המאמרים המרכזיים שמסתמכים על המאמר.
-  **חוקר אקדמי:** החוקר האקדמי יציע רעיון לעבודת המשך המבוססת על המאמר הזה. החוקר יכתוב תקציר של עד חצי עמוד על הרעיון, לרבות סקירה של ניסויים ותוצאות היפותטיים שאולי היו מתקבלים בעבודה על הרעיון הזה, ויציג את הרעיון לכיתה.
-  **חוקרת פרטית:** החוקרת הפרטית תנסה לברר את הרקע של אחד הכותבים של המאמרים (בעדיפות לכותבת הראשונה). היא תספר לנו בקצרה על העבודות הקודמות של הכותבת, ותנסה קצת לברר מה הסיבות שהובילו אותה לכתוב את המאמר הזה. החוקרת יכולה (ואף מומלץ לה!) לפנות לכותבת עם כמה שאלות (בצורה מכובדת ומנומסת כמובן!).
-  **פצחן (אופציונלי):** הפצחן ינסה לממש גירסא פרוטוטיופית של הרעיון המרכזי במאמר. התוצאה תהיה מחברת jupyter שמציגה את המימוש, והצגה שלה בכיתה, ביחד עם פירוט של האתגרים המרכזיים במימוש, ודברים מעניינים אחרים שעלו. הפצחן יכול להשתמש בקוד שזמין באינטרנט לצורך השראה, ולהשתמש בחבילות תשתית קיימות, אבל עליו לכתוב את הליבה של הקוד בעצמו. אין צורך להשקיע יותר מידי זמן בדביבוג ובהוספת פונקציונליות מיוחדת, הרעיון הוא לקבל משהו ראשוני ש(בערך) עובד.
-  **העורכים הלשוניים:** העורכים הלשוניים יציעו כותרת אלטרנטיבית למאמר.

## פרטים טכניים:

1. בכל שבוע יהיה לפחות סטודנט אחד שמבצע את תפקידים 2-5, ולפחות שלושה סטודנטים שמבצעים את תפקיד 1. תפקיד 6 הוא אופציונלי. תלמידות שלא מבצעות אף אחד מהתפקידים 1-6 באותו השבוע יבצעו את תפקיד 7.
2. כל סטודנט/ית יבצעו במהלך הסמסטר פעם אחת לפחות את תפקידים 2-5. את תפקיד 1 כל סטודנטית תצטרך לבצע לפחות שלוש פעמים. תפקיד 6 (האקר) הוא אופציונלי, ורק מי שרוצה תבצע אותו.
3. ציונים:
  - a. כל הצגה תזכה את המציג/ה בציון שיקבע על פי ההצגה והמסמך בכל שבוע.
  - b. תפקידים 1-5 שווים 10 נקודות בציון הסופי כל אחד.
  - c. תפקיד 6 שווה 20 נקודות
  - d. תפקיד 7 שווה 5 נקודותבנוסחים:
- a. בתחילת כל שיעור, נציג את כל הכותרות החלופיות למאמר, ונערוך הצבעה בכיתה על הכותרת המוצלחת ביותר. הכותרת המנצחת תקבל 2 נקודות בונים לציון בסופי.

## סילבוס:

- הסמינר יהיה מחולק לארבעה חלקים, כל אחד מהם עוסק באספקט אחר של יעילות. להלן פירוט הנושאים והמאמרים לפי שבועות (טנטטיבי, ייתכנו שינויים במהלך הסמסטר):
1. פתיחה, הצגת הפורמט (שבוע 1)
  2. ארכיטקטורות (שבועות 2-4)
    - a. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. [Attention is all you need](#). NeurIPS 2017
    - b. Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer. [Deep contextualized word representations](#). NAACL 2018.
    - c. Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei. [Language Models are Few-Shot Learners](#). NeurIPS 2020.
  3. הפיכת מודל גדול ויקר לקטן וזול (שבועות 5-8)
    - a. Geoffrey Hinton, Oriol Vinyals, Jeff Dean. [Distilling the Knowledge in a Neural Network](#). NIPS 2014 Deep Learning Workshop
    - b. Jonathan Frankle, Michael Carbin. [The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks](#). ICLR 2018

- c. Angela Fan, Edouard Grave, Armand Joulin. [Reducing Transformer Depth on Demand with Structured Dropout](#). ICLR 2020
- d. Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, Dmitry Kalenichenko. [Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference](#). CVPR 2018.

4. יעילות בדאטא

- a. Yoshua Bengio, J. Louradour, Ronan Collobert, J. Weston. [Curriculum learning](#). ICML 2009
- b. Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, Yizhong Wang, Hannaneh Hajishirzi, Noah A. Smith, Yejin Choi. [Dataset Cartography: Mapping and Diagnosing Datasets with Training Dynamics](#). EMNLP 2020
- c. Roy Schwartz, Gabi Stanovsky, Swabha Swayamdipta, Jesse Dodge, Noah A. Smith. [The Right Tool for the Job: Matching Model and Instance Complexities](#). ACL 2020

5. אלטרנטיבות לטרנספורמר

- a. Iz Beltagy, Matthew E. Peters, Arman Cohan. [Longformer: The Long-Document Transformer](#). 2020
- b. Hao Peng, Nikolaos Pappas, Dani Yogatama, Roy Schwartz, Noah A. Smith, Lingpeng Kong. [Random Feature Attention](#). ICLR 2021
- c. Tao Lei. [When Attention Meets Fast Recurrence: Training Language Models with Reduced Compute](#). 2021

### שאלות נפוצות:

1. **שאלה:** למי כדאי להירשם לסמינר?

**תשובה:** מי שמתעניין בעיבוד שפה ושרוצה ללמוד על שיטות שונות לייעול הכלים הקיימים, ועל הדרך להתנסות בכישורי מחקר רלוונטיים. הסמינר הוא למעשה מיני קורס בנושא, ודרך טובה להיכנס לנושא של יעילות.

2. **שאלה:** האם יש מגבלה על הרישום לסמינר?

**תשובה:** הסמינר מיועד לתלמידי מתקדמים או לתלמידי שנה ג'-ד' באישור המרצה. דרישות הקדם הן עיבוד שפה טבעית (67658), מבוא למערכות לומדות (67577), ורקע כלשהו בלמידה עמוקה (למשל 67912, 67822, 67750). למקרים מיוחדים נא לפנות לרועי. על מנת ליצור אווירה שמעוררת דיון של כולם, ההשתתפות בסמינר תוגבל ל-20 תלמידים.

3. **שאלה:** למה כדאי לצפות מבחינת עומס העבודה?

**תשובה:** הקורס ידרוש השקעה של בין שעה-שעתיים ליום בשבוע, תלוי בתפקיד של אותו השבוע. הציפייה היא שהקורס לא ידרוש יותר זמן במצטבר מסמינר רגיל, שבו צריך לקרוא את המאמר כל שבוע ולהתכונן במשך שבוע-שבועיים להצגה של המאמר, אבל החלוקה בין השבועות תהיה שונה.

4. **שאלה:** למה הכאב ראש הזה? לא עדיף לי סמינר רגיל?

**תשובה:** יכול להיות שכן, תלוי בכך. אם השאיפה היא למינימום השקעה וראש קטן, כנראה שזה לא המקום המתאים. אם השאיפה היא ללמוד בצורה די רצינית את הנושא, להתנסות בחוויות מעולם המחקר, ולהיות באווירה של דיון אקדמי מעניין (👉), שווה לך לנסות!

5. **שאלה:** למה הכאב ראש הזה? לא עדיף לי סמינר רגיל?

**תשובה:** יכול להיות שכן, תלוי בכך. אם השאיפה היא למינימום השקעה וראש קטן, כנראה שזה לא המקום