

על מה ChatGPT חולם? טעימה מהפילוסופיה של בינה מלאכותית

אוניברסיטת בן-גוריון, תשפ"ד, סמסטר א' (סתיו)
יום ב', 10:00-12:00, בניין 35, חדר 115

ד"ר אורי בק, becko@bgu.ac.il

שעות קבלה: יום ב', שעה 14:00-15:00, בניין 74 (דילר), המחלקה לפילוסופיה, חדר 330

בודק תרגילים: מר אלון כהן

1 הקדמה, לאור המצב...

א. מה המצב? מאז ה-7 באוקטובר כל הקהילות בישראל נפגעו ונפגעות מאוד. קהילת אוניברסיטת בן-גוריון היא לא חריג לכלל זה. קהילת אוניברסיטת בן-גוריון בעצמה מורכבת מהרבה קבוצות שונות של אנשים, וחברי כל הקבוצות האלה סובלים או לפחות מדוכדכים עכשיו. חלקנו איבד קרובים וחברים, חלקנו נפצעו, חלקנו בסכנה או שיש לנו קרובים וחברים בסכנה, חלקנו מרגישים חסרי ביטחון, חרדים או מנוכרים, חלקנו מיואש. למרות זאת, כדי לשמר את הקהילה המגוונת שלנו, אנחנו רוצים לסייע לכולם.

ב. למה אנחנו בכיתה? אנחנו בכיתה כדי ללמוד, להתפתח אינטלקטואלית, לרכוש מיומנויות, להתקדם בחיים, לבסס בית, משפחה, קריירה. זו מטרה שמשותפת לכל חברי אוניברסיטת בן-גוריון, לכל הסטודנטים, ולכל המרצים. לכן כדאי לכולנו לעשות ככל יכולתנו כדי לסייע אחד לשני. לאוניברסיטה עצמה, ולי בפרט, חשוב מאוד לסייע לכם.

ג. אבל מי רוצה ללמוד בתנאים כאלה? לימודים יכולים להרגיש כמו פעילות תלושה, אסקפיטיבית, קשה, ומרתיעה. ההרגשה הזו מטעה. האמת היא שעכשיו זה הזמן הטוב ביותר להתמסר ללימודים, ולעיסוקים אינטלקטואליים באופן כללי. בפעילויות אינטלקטואליות כמו מחקר ולמידה אנחנו מחלצים את עצמנו מהכאן והעכשיו, ובעזרת כוח הרצון שלנו בלבד אנו מטיילים במחשבות למחוזות שאנו בוחרים בהם בחופשיות. בימים כאלה והכאן והעכשיו שלנו מלאים כאב או לפחות אי נחת, ולעומת זאת, החופש שלנו להתנער במחשבה מהכאן ומהעכשיו זה בדיוק החופש שמייחד אותנו כבני אנוש, ובדיוק החופש שבעזרתו נתפתח, נתקדם ונשגשג. אני מעודד אתכם להתמסר לחופש הזה, דווקא עכשיו.

ד. ואיך אפשר ללמוד בתנאים שכאלה? אנחנו צריכים בעיקר לגייס את הרצון לעשות את זה, ולא להפריע זולתנו לעשות את אותו הדבר. לשם כך נחוץ לנו תקנון:

תקנון

- שנת הלימודים הקרובה יוצרת אתגר לשיח שלנו בכיתה. הלמידה שלכם חשובה לנו, וכך גם המוגנות שלכם. כדי שנוכל להצליח במשימה הלימודית נגדיר כמה כללים לשנה הקרובה.
- למרצים יש זכות להביע את דעתם המקצועית בתחומי התמחותם, אך גם הם כפופים לכללי התקנון. מרצים גם נדרשים להוגנות כלפי כל הסטודנטים שלהם, ללא הטיות.
- בתקופה הנוכחית כל קהילת האוניברסיטה לוקחת על עצמה מחויבות להגן על התנאים ללמידה בכיתה. לכן, בתקופה זו נקפיד כולנו הקפדה יתרה על תקנות ביטוי מכבד.
- תקנות הנוגעות לתוכן הביטוי:
- חופש הביטוי הוא אחד הערכים המרכזיים במשטר דמוקרטי ליברלי, במיוחד באקדמיה. בסיסו: אנו רוצים בגילוי האמת. כדי לגלות את האמת, עלינו להבטיח שכל הרעיונות יכולים להתבטא ולהתחרות עם רעיונות אחרים. זה כולל רעיונות שנויים במחלוקת, קשים לעיכול, ומכעיסים.
- אף על פי כן, חופש הביטוי הוא לא העקרון החשוב היחיד, והחוק הישראלי מגביל אותו. בימים אלה חשובות לזכור במיוחד שהחוק אוסר על
 - רדיפה, השפלה, ביזוי, גילוי איבה, עוינות או גרימת מדנים כלפי ציבור או חלק אוכלוסייה בשל צבע, גזע, או מוצא אתני-לאומי.
 - הסתה לאלימות.
 - גילוי הזדהות עם ארגון טרור באמצעות מעשה, לרבות דרך פרסום דברי שבח תמיכה או אהדה, הנפת דגל, הצגת סמל או פרסום סיסמא או המנון.
- תקנות הנוגעות לאופן הביטוי:
 - בכיתה זו נפגין כבוד לכלל התלמידים בכך שכולנו נדבר בשם עצמנו, ובסוגיות שעלולות להיות רגישות נמנע מלדבר בשם הקבוצה שלנו או בשם קבוצה אחרת.
 - בכיתה זו נקשיב אחד לשנייה על מנת להבין ולא על מנת להגיב. נימנע משיפוט וביקורת בזמן ההקשבה.

- נכבד את נקודת המבט והדעה של הזולת. לכולנו יש זכות לדעה אישית.
- בכיתה זו נקשיב אחד לשנייה על מנת להבין ולא על מנת להגיב. נימנע משיפוט וביקורת בזמן ההקשבה.
- אם הערה מסוימת פוגעת או מעליבה, אנא אמרו שהערה (ולא הדובר) פוגעת או מעליבה, והסבירו למה.
- לכולנו יש זכות שלא להשתתף בדיון.
- זה בסדר גמור לחוש מבוכה, או לא לדעת את התשובה לכל דבר.
- על מנת לאפשר דיון פורה, כולנו נשקול את השפה שאנו משתמשים בה, ונברור את המילים שלנו בזהירות.

2 תאור הקורס

חקור את העולם המרתק של בינה מלאכותית (AI) והשלכותיה הפילוסופיות. בשיעור זה, נצא למסע פילוסופי לחשיפת המסתורין של בינה מלאכותית, ונבחן את הגבולות שבין מחשבה אנושית ומכאנית. נחקור את היסודות הפילוסופיים של בינה מלאכותית, מגישות קלאסיות לרשתות עצביות מודרניות, ונדון באתגרים ובאפשרויות שהן מציגות. ייתרה מכך, השיעור יצלול למימדים המוסריים של בינה מלאכותית, ויעסוק בדיונים מעוררי מחשבה על ההשלכות של בינה מלאכותית על החברה. ידע פילוסופי קודם לא נדרש. (תמצית זו נכתבה ע"י ChatGPT.)

Explore the fascinating realm of artificial intelligence (AI) and its philosophical implications. In this course, we will embark on a philosophical journey to unravel the mysteries of AI, examining the boundaries between human and machine cognition. We will explore the philosophical foundations of AI, from classical AI to modern neural networks, and discuss the challenges and possibilities they present. Moreover, the course will delve into the ethical dimensions of AI, and engage in thought-provoking discussions surrounding the impact of AI on society. No prior background in philosophy is required. (This abstract was written by ChatGPT.)

מילות מפתח: בינה מלאכותית, מחשבה, קוגניציה, פסיכולוגיה, נפש, מוסר, פילוסופיה, ChatGPT, רשתות עצביות, רשתות נוירונים

3 מטרת למידה

בתום קורס זה, את/ה תוכל/י

- להסביר משגים, בעיות ועמדות בפילוסופיה של בינה מלאכותית.
- לבקר טיעונים בעל פה לאור טעמים.

4 מבנה הקורס

ככהנה לקראת כל נושא/שיעור, עליכם לקרוא את טקסטי החובה להלן:

תאריך	נושא	קריאת חובה
1/1	הקדמה	M: Prologue Bostrom: How Long Before Superintelligence? (https://nickbostrom.com/superintelligence)
8/1	רקע	M: 1-2 Turing: Computing Machinery and Intelligence
15/1	רקע	M: 3-4 Buckner: Deep Learning: A Philosophical Introduction
22/1	רקע	M: 5-6 Hassabis, Kumaran, Summerfield, Botvinick: Neuroscience-Inspired Artificial Intelligence (https://doi.org/10.1016/j.neuron.2017.06.011)
29/1	מחשבים חושבים?	M: 14 Putnam: Minds and Machines
5/2	מחשבים חושבים?	Lucas: Minds, Machines and Gödel
12/2	מחשבים חושבים?	Searle: Minds, Brains, and Programs.
19/2	מחשבים חושבים?	Chalmers: A Computational Foundation for the Study of Cognition
26/2	מחשבים מרגישים?	Butlin et al: Consciousness n Artificial Intelligence: Insights from the Science of Consciousness
4/3	רובוטים טובים?	WA: 1-2
11/3	רובוטים טובים?	WA: 6-8

שימו לב: (1) כל חומרי הקריאה נמצאים באתר הקורס. (2) הנחיות קריאה מהצורה "M: #" הן לקריאת פרק מספר # מתוך הספר של Melanie Mitchell, Artificial Intelligence: a guide for thinking humans. (3) הנחיות קריאה מהצורה "WA: #" הן לקריאת פרק מספר # מתוך הספר של Wendell Wallach ו-Colin Allen, Moral Machines: Teaching Robots Right from Wrong.

5 מתודולוגיות הוראה ולמידה

למרות שארצה בפני הכיתה, חלקים נכרים מהשיעור יוקדשו לדיון. אני מצפה מכולם להשתתף. אני גם מצפה מכולם לקרוא את ספרות החובה הנדרשת לפני השיעור. הציון יתבסס על נוכחות, ועל בחינה מסכמת. הנחיות לקריאת וכתובת פילוסופיה, ובחינת בית לדוגמה יופיעו באתר הקורס במועד מתאים.

6 הערכה

משקל מרכיבי הציון הסופי הוא:

6.1 נוכחות (חובה)

הדרך הטובה ביותר להבין את החומר וללמוד לחשוב בבהירות היא לשוחח על נושאי השיעור בקול רם, להקשיב למה שאחרים אומרים, ולהשיב להם באופן ענייני. לכן, תלמידים שנעדרו מיותר מ-3 שיעורים ללא צידוק מתועד לא יוכלו להשלים את הקורס.

6.2 בחינה מסכמת (100% מהציון הסופי)

סוף הסמסטר תתקיים בחינה. בחינה לדוגמא תתפרסם באתר הקורס במועד מתאים.

7 בריאות

אם אינך שבע/ת רצון מהצלחתך בקורס, או אם את/ה מודאג/ת שהשתתפותך בקורס מושפעת ממצב גופני או נפשי שאינו מיטבי, פנה/י אלי בהקדם האפשרי. את/ה רשאי/ת גם לבקש מגורם שלישי שמבין את מצבך לפנות אלי בעבורך. דע/י שאת/ה רשאי/ת בנוסף לפנות ליחידה לשירות פסיכולוגי ולהכוון מקצועי ולימודי, in.bgu.ac.il/Dekanat/Pages/about-pserv.aspx.

אם את/ה שייך לקבוצת אוכלוסיה מיוחדת, סובלת/ת מנכות או לקות למידה, או מתמודד עם השלכות שירות מילואים תובעני, ביכולתך לפנות למדור לקידום סטודנטים, <https://in.bgu.ac.il/Dekanat/Pages/promotion.aspx>, ו/או למרכז התמיכה לסובלים מנכויות או לקויי למידה, in.bgu.ac.il/Dekanat/Pages/support_center.aspx.

8 יושר אקדמי

האוניברסיטה ואני מעריכים הוגנות, יושר וכבוד. מעשי פלגיאט פוגעים בערכים אלה ומהווים עבירת מרמה חמורה.

העונש על מעשי פלגיאט מובהר במדריך האוניברסיטה לציטוט ושימוש בידע קודם: "האקדמיה רואה בפלגיאט מעשה של אי יושר אקדמי חמור שאינו מתאים לציפיות לגבי אדם הנושא תואר אקדמי. לפיכך, הפלגיאט נחשב באקדמיה לעבירה חמורה מאין כמוה שזוכה לענישה חמורה בהתאם. לפי התקנון של אוניברסיטת בן גוריון, נקבעו לסטודנטים שיוורשו בעבירה זו עונשים כבדים. דוגמאות לעונשים שהוטלו בעבר: ביטול תואר או תעודה, הרחקה מהאוניברסיטה לתקופה ארוכה, הרחקה לצמיתות."

המדריך המלא, שמסביר מהו פלגיאט וכיצד להימנע ממנו, נמצא בכתובת in.bgu.ac.il/acadsec/DocLib2/plagiarism_heb.pdf.

לידיעתכם: העבודות ובחינת הבית יוגשו לאתר הקורס, שם יעברו בדיקות מקוריות ומניעת העתקות באמצעות תוכנה ייעודית שתשווה בין הטקסטים השונים, בינם לבין טקסטים שהוגשו בשנים עברו, ובינם לבין טקסטים המופיעים ברשת. הזהרו!

9 משוב

אני מעודד אותך לתת לי משוב. משוב יכול לשפר את חווית הלמידה שלך ושל חבריך, ולשפר את מיומנויות ההוראה שלי. אם את/ה מעדיף שלא לתת לי משוב באופן אישי, ביכולת לתת לי משוב אנונימי דרך האתר החיצוני, <https://sites.google.com/view/bgu-anon-feedback>.

10 קשר עם חבריך לכיתה

כולנו בני אדם, ולפעמים אנחנו לא מגיעים לשיעור בזמן או לא מצליחים להתרכז. כשזה קורה, אנו צריכים עזרה בהשלמת החומר. דרך טובה לעשות זאת הוא להיעזר בחברכים לכיתה. זה יועיל גם להם – מחקרים מראים שסטודנטים שמסבירים את החומר לחבריהם מצליחים בלימודים יותר מסטודנטים אחרים. לכן, נצלו את ההזדמנות וצרו לקורס קבוצת קשר, למשל בווטסאפ.